The Nature, Prevalence and Effectiveness of Strategies Used to Prepare Pupils for Key Stage 2 Maths Tests
Final report
September 2019

M. Ehren, N. Wollaston, J. Goodwin, J. Star
UCL Institute of Education

Corresponding author:
Professor Melanie Ehren
m.ehren@ucl.ac.uk
m.c.m.ehren@vu.nl

Acknowledgments

We express our gratitude to the Nuffield foundation for funding this study (grant number EDO 41601). The Nuffield Foundation is an independent charitable trust with a mission to advance social well-being. It funds research that informs social policy, primarily in Education, Welfare, and Justice. It also funds student programmes that provide opportunities for young people to develop skills in quantitative and scientific methods. The Nuffield Foundation is the founder and co-funder of the Nuffield Council on Bioethics and the Ada Lovelace Institute. The Foundation has funded this project, but the views expressed are those of the authors and not necessarily the Foundation. Visit www.nuffieldfoundation.org

Our appreciation also goes to the teachers who were willing to be interviewed by us; their time and insights were invaluable for this study.

Table of Contents

Executive summary

Problem statement and research questions

This study investigated year 6 teachers' understanding of (changes in the) key stage 2 mathematics test and their instructional strategies to prepare pupils for the test. The study aimed to enhance our understanding of the practices teachers adopt to prepare pupils for this test; examining the features of the test to which teachers are responsive, and how beneficial or harmful their practices are in promoting student learning in mathematics. We explored the following research questions in a small scale qualitative study:

1. To which features of key stage 2 mathematics tests are teachers responsive?
2. What instructional strategies do teachers use in response to key stage 2 mathematics tests?
3. How beneficial or harmful are these strategies in promoting student learning?

Research methods

Our research included three phases:

1. Expert analysis: a former principal manager of the QCA mapped the KS2 mathematics test, administered between 2010 and 2017 (inclusive) onto the 2016 test framework (STA, 2015). The analysis allowed us to analyse predictable patterns in the test, as well as how these patterns changed after the introduction of the new national curriculum tests in 2016.
2. Interviews with 30 Year 6 teachers from schools scoring below and above the floor target on the KS2 mathematics test; teachers were interviewed before and after the change in the test (round 1 in February and May 2015, round 2 in June/July 2016). In the interviews we asked teachers about their understanding of clusters of frequently occurring test items and how they prepare their pupils for these clusters. We were interested to learn if teachers have a similar understanding of test patterns as identified in our expert analysis.
3. Reviewing the effectiveness of teachers' instructional strategies.

Findings

Patterns in the test

Findings from our expert analyses indicated the following patterns on the 2010-2017 key stage 2 mathematics test:

- The curriculum strand 'calculation' has the most items over the years with 'ratio', 'algebra' and 'position' the fewest.
- Most items on the 2010-2017 tests included one numeric step where the 2016 and 2017 tests see an increase in items with a larger number of simple numeric steps and a decrease in items with no numeric steps.
- Most items require no spatial reasoning across the years, nor any data interpretation, whereas most items require children to select or construct a response. The number of items requiring children to construct a set of responses has decreased over the years.
- Test papers between 2010 and 2017 primarily included constrained questions.

The change in test in 2016 has led to an initial increase in number of items to test 'frequency' and a reduction in 'statistics'. There is no clear change in items for the various sub-strands. Between 2010 and 2015, depth of understanding was mostly tested on level 2 and 3, while the 2016 and 2017 test included more items on level 1 and level 4. A higher level represents a greater difficulty. The increase in level 4 items can be explained by the discontinuation of the separate level 6 test and incorporating these more complex items in the regular test papers. As we only have two years of analysis of the new test, these changes may be spurious.

Teachers' understanding in (general) changes in the test:

Teachers talked about the following three changes in the 2016 key stage 2 mathematics test:

- Removal of levels where the test now has a random order of items according to difficulty level, instead of a gradual increase in difficulty. This change seemed to reduce children's motivation to take tests as they are discouraged when encountering a difficult item at the start of the test. The removal of levels also affected teachers' understanding of the hierarchical nature of mathematics, and change their instruction towards a more 'mastery style' of teaching where all children needed to master the basics before moving on to more complex skills.
- Removal of mental mathematics paper and the introduction of an arithmetic paper with context-free calculations which have to be completed in 30 minutes seemed to lead to an increase in practicing arithmetic questions. As formal methods get higher marks, children are often also no longer allowed to use their preferred, less formal, method.
- The introduction of two papers titled 'reasoning', which include word problems, appears to change teachers' understanding of what problem-solving entails, where this is now (by some teachers) seen as the same skill as reasoning. Some teachers instruct pupils in how to explain their thought processes more clearly and learn to work more systematically in order to do well in answering items in the 'reasoning' papers.

Teachers' description of, and instructional responses to, clusters of similar test items

Teachers described 30 clusters in total, where three clusters (word problems, specifically 2-step money problems, and context-free calculations) were described by most teachers in both rounds of interviews.

1. Word problems (including 2-step money problem) which require children to read a problem and decide which operation(s) to carry out.

Teachers believed that success in word problems and 2-step money problems comes about when 1) pupils are able to read the words in the problem in order to understand the problem context, and once the pupil understands the problem context, 2) he/she must be able to identify the most important words in the problem statement, particularly the important *mathematical* words. Teachers talked about teaching problem-solving heuristics (e.g. reading the problem aloud, identifying important mathematical words), having pupils draw pictures to represent the problem context, and/or providing pupils with practical work in the context to aid in developing familiarity (e.g., cooking).

Our review offered some reflections on whether these heuristics are necessary or sufficient to promote problem solving success, and suggested schema-based instruction as an alternative, potentially more effective, strategy. Such instruction would focus on the strategy to solve problems, offering pupils the opportunity to look for related problems. This would also allow for problems unfamiliar to the children to be considered and strategies established which could be related to problems in the tests.

2. Context-free calculations: requiring pupils to complete a context-free calculation, set out in horizontal format, typically involving three-digit number(s). The calculation may involve multiplication or division which would get pupils two marks for a correct answer, while answers need to be explained in a 'show your working' box. Calculations may also involve addition or subtraction, which are more likely to be worth one mark only and have no designated working out space.

Teachers mentioned a number of methods they would teach pupils to calculate the correct answer on context-free items which vary according to ability level of pupils, the size of the numbers involved, or whether the calculation involved round numbers or decimals. Strategies included practicing times tables with pupils, and going over number facts every week to make sure that pupils can recall these quickly when tested. Methods for more complex calculations included chunking, standard algorithm, the grid method, the bus shelter method, the ladder method, a formal method of long division, the column method. Standard algorithms (e.g. long division) were considered favourable as pupils would get higher marks when using this method.

Instructional strategies to prepare pupils for context-free calculations included practicing times tables and number facts to improve fluency and teaching procedures for more complex calculations. Traditional methods (e.g. column method) were prioritized after the change in test as these would get pupils higher marks, according to teachers.

Studies suggest that a large number of pupils find these traditional methods difficult and fail to understand the conceptual basis for these procedures. Although the primary mathematics literature does not offer a generally accepted best practice, a number of authors suggest student-generated approaches as a more effective alternative; in such an approach a problem is posed, and then pupils spend time discussing how they solved it, using a variety of strategies for solving the problem. The teacher then uses the range of strategies presented to engage pupils in rich mathematical conversations.

Implications for policy and practice

Our study findings indicate that changes in the 2016 Key Stage mathematics test have caused a shift in teaching with an emphasis from conceptual knowledge towards procedural knowledge. We expect this shift to strengthen students' instrumental understanding of mathematics and their use of written methods for calculating and following procedures to solve calculations using for example column methods of division and multiplication. However, the test in its current form suggests a reduced emphasis on a relational and conceptual understanding of Number and Calculation and of 'mental strategies' advocated under the National Numeracy Strategy. Given the important role of the Key Stage 2 test in prioritising what is taught, we suggest that Key Stage 2 mathematics tests need to be scrutinized for items which explicitly support the development of mental and flexible problem solving, while also ensuring teachers are supported in teaching mental strategies for problem-solving.

1. Introduction

For two decades, primary schools in England have been held to account for the effectiveness of their teaching on the basis of results from national curriculum tests in core subjects. Originally, these tests covered English, mathematics and science; although the science tests have been dropped and testing requirements in English continue to evolve. Primary schools face increasingly challenging targets for pupils' performance on test-based accountability measures.
These accountability measures are a double-edged sword: on one edge, they are intended to motivate higher attainment; on the other, however, they risk effort being wasted on strategies that inflate test scores without a corresponding impact on robust, generalizable knowledge, skill and understanding. In recent years, there has been a lot of attention paid to the use of test preparation strategies by teachers in both primary and secondary schools in England. Research has demonstrated a strong influence of accountability testing upon both curriculum organisation and teaching behaviour (e.g. Mansell, 2007).

Official publications and reviews have repeatedly identified, as a major concern, the increased prevalence of the use of dubious test preparation strategies, such as 'teaching to the test', over-rehearsal, and focusing undue effort upon borderline pupils (e.g. Ofsted, 2008; 2012; Bew, 2011). An Ofsted (2012) report for example describes how schools provided revision or booster classes for pupils in years 6, 9 and 11 and focused particularly on those pupils at risk of narrowly missing the key threshold targets. In response to this growing unease at the impact of accountability testing upon teaching and learning, a former senior official at the Department for Education has argued that teachers should have the moral courage to avoid dubious test preparation strategies, and focus instead upon genuine strategies with the potential to facilitate robust, generalizable knowledge, skill and understanding; e.g. authentic and deep subject teaching (Coles, 2012).

Unfortunately, although there is research evidence and descriptive taxonomies concerning the use of *general strategies* for inflating test scores (e.g. 'teaching to the test'), there is very little evidence at all concerning the range and prevalence of more *specific strategies*, particularly those specific to subject domains, like mathematics. Indeed, at the subject level, there is actually limited understanding as to what might constitute a 'dubious strategy' (for inflating test scores) as opposed to a 'genuine strategy' (for raising attainment). Exploratory studies have however provided anecdotal evidence of the existence of such subject-specific strategies and have pointed out how they may compromise the quality of the curriculum and reduce what children learn about a subject (Department for Children, Schools and Families, 2010). Ehren and Star (2012), for example, described how teachers tried to improve pupils' ability to solve mathematical word problems on the high stakes New York state test, by teaching pupils to look for and 'decode' key words and phrases that often appeared in the items of that test to signal addition, subtraction or division. Mansell (2007, p.30) described how a primary school teacher in England tailored her teaching to questions similar to those which were likely to appear in the tests.

Clear empirical evidence on the effectiveness of these strategies is lacking. Mathematics education research provides some guidance, for example in showing that teaching pupils general heuristics to solve mathematical word problems does not have a positive impact on learning and transfer. The evidence for most of the subject-specific test preparation strategies is however mixed at best (Ehren and Star, 2012).

This report aims to enhance our understanding of teachers' test preparation strategies and the appropriateness and efficacy of these strategies. We investigated the key stage 2 test for recurring patterns of similar test items, how teachers understand such patterns and how these patterns potentially influence their instruction. The following research questions informed our study:

1. To which features of key stage 2 mathematics tests are teachers responsive?
2. What instructional strategies do teachers use in response to key stage 2 mathematics tests?

3. How beneficial or harmful are these strategies in promoting student learning?

The section below first presents a description of the key stage 2 test and changes in the test.

2. Context of the key stage 2 test

Key stage 2 tests are administered in English and Mathematics in May at the end of year 6, i.e. the end of primary schooling. The test measured children's mastery of the 2000 curriculum prior to 2016, while the 2016 and 2017 tests measured the national curriculum introduced in 2014 with English and Mathematics coming into force for all year-groups from September 2015[1].

KS2 Mathematics test 'old curriculum' (prior to 2016)

Prior to 2016, the key stage 2 mathematics tests were administered to 11-year olds in primary schools in England during May at the end of year 6. Pupils sat the test over two days, completing a non-calculator paper (paper 1; generally 22 items, 40 marks), a calculator paper (paper 2, generally 23 items, 40 marks), and a mental mathematics paper (20 marks). From 2014 calculators could no longer be used by any children sitting the levels 3–5 mathematics test. The tests contained a mixture of a limited number of open response questions; closed response questions, including multiple choice, true false and matching; and, written explanations and methods. Until the introduction of the new national curriculum in 2014, the aim of the key stage 2 test was to measure pupils' content knowledge of the 2000 national curriculum, their using and applying of mathematics, and their reasoning and communication skills. The areas covered were 'number and algebra', 'shape, space and measures', and 'handling data'.

Test papers up until 2016 were reported in levels which presented, according to the Standards and Testing Agency, the extent to which children understand the concepts and skills within a specific level. The expected standard for the key stage 2 test was a level 4; children achieving a level 5 were above the expected standard, whereas a level 3 meant that they were below the expected standard. Level 6 was tested in a separate paper (not included in our analysis).

The key stage 2 test was developed by the Standards and Testing Agency (STA; a government body), with the support of external contractors, according to a framework which described the nature of the test and how many marks are assigned to each level and each aspect of the national curriculum. The test was marked by expert markers external to the schools (although largely teachers), who were trained at national level. School-level data, reproduced in performance tables, were used to monitor the effectiveness of schools and to support school inspections. Schools had to meet 'floor standards' –including the requirement of at least 60% of their pupils performing at the 'national standard' level 4 or above- and would face intensive monitoring and consequences when performing below the floor standards. These floor standards have become increasingly challenging with a minimum percentage of 65 in 2014, and 85% in 2015.

KS2 Mathematics test 'new curriculum' (2016 onwards)

A new national curriculum was introduced in 2014 which was tested from 2016 onwards. The curriculum for mathematics aims to ensure that all pupils:

- become fluent in the fundamentals of mathematics, including through varied and frequent practice with increasingly complex problems over time, so that pupils develop conceptual understanding and the ability to recall and apply knowledge rapidly and accurately.
- reason mathematically by following a line of enquiry, conjecturing relationships and generalisations, and developing an argument, justification or proof using mathematical language

[1] The majority of this national curriculum was introduced in September 2014, with English and mathematics coming into force for all year groups from September 2016. Source: https://www.gov.uk/government/collections/national-curriculum

- can solve problems by applying their mathematics to a variety of routine and non-routine problems with increasing sophistication, including breaking down problems into a series of simpler steps and persevering in seeking solutions.

The 2014 national curriculum in mathematics was first tested in the key stage 2 Mathematics tests in 2016; the test included three test papers[2]: mathematics paper 1: arithmetic, mathematics paper 2: reasoning, mathematics paper 3: reasoning[3]. The 2016 and 2017 tests were reported in scaled scores instead of levels, and the level 6 paper was also abolished. A score of '100' represents the 'national standard'. Each pupil's raw test score is converted into a score on the scale, either at, above or below 100. The scale will have a lower end point somewhere below 100 and an upper end point above 100. A child who achieves the 'national standard' (a score of 100) will be judged to have demonstrated sufficient knowledge in the areas assessed by the tests[4]. Other changes in the Mathematics test are: children are no longer allowed to use a calculator when taking paper 2. Paper 3 used to be a 'mental maths' paper, but is replaced by an arithmetic test (now paper 1). The following table summarizes the new key stage 2 Mathematics test.

Table 1. Summary of key stage 2 Mathematics test (2016 onwards)

Component	Description	Number of papers	Number of marks	Timing of component
Paper 1	Arithmetic	1	30	30 minutes
Paper 2 and Paper 3	Mathematical fluency, solving problems and reasoning	2	80 overall 40 per paper	80 minutes 40 minutes per paper
	Total	**3**	**110**	**110 minutes**

Summary of changes[5]:

- From levels to scaled scores
- Removal of separate level 6 paper
- Paper 3 changed from mental mathematics to arithmetic
- Use of calculator no longer permitted
- New questions (in paper 1) which represent the new mathematics national curriculum, particularly to assess pupils' arithmetic ability in using formal methods to solve specific arithmetic questions, e.g. long multiplication and long division, and the introduction of operations involving fractions.

3. Methodology

Our research involved the following three steps:

Step 1: expert analysis of the key stage 2 Mathematics test

The first step entailed an expert analysis of the key stage 2 Mathematics test. A UK Mathematics test expert who previously was a principal manager at the QCA responsible for the development of the end of key

[2]http://www.satspapers.org/SATs%20papers/2016%20samples/2016%20Sample%20Maths/2016_Key_stage_2_Mathematics_test_framework.pdf

[3] https://www.gov.uk/guidance/2016-key-stage-2-assessment-and-reporting-arrangements-ara/section-2-key-changes

[4] STA (2017). Changes to Key Stage 2 SATs – 2016.

[5]https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/328886/2014_KS2_mathematics_sample_materials.pdf

stage mathematics tests (Jeff Goodwin) mapped each item in the 2010-2017 tests onto the 2016 key stage 2 mathematics test framework (2015). The map allowed us to analyse potential patterns in the test which may enable teachers to teach to the test.
The mathematics framework was developed by the Standards and Testing Agency (STA) and is based on the national curriculum programme of study (2014) for mathematics, introduced for teaching in schools from September 2014 and first assessed in May 2016.
The framework specifies the purpose, format, content and cognitive demand of the key stage 2 mathematics tests and was produced to aid the test development process. 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 national curriculum elements as specified in strands and sub-strands), the cognitive processes considered central to the mathematics tests (depth of understanding, computational complexity, spatial reasoning, data interpretation, response strategy), and types of item:

The strands that were tested between 2010 and 2017 are:
- N: Number and place value, approximation and estimation / rounding (6 sub-strands N1-N6)
- C: Calculation: Addition, subtraction, multiplication and division (9 sub-strands; C1-C9)
- F: Fractions, decimals and percentages (12 sub-strands, F1-F11)
- R: Ratio and proportion (4 sub-strands; R1-R4)
- A: Algebra (5 sub-strands; A1-A4)
- M: Measurement (9 sub-strands; M1-M9)
- G: Geometry – properties of shapes (5 sub-strands; G1-G5)
- P: Geometry – position and direction 3 sub-strands; P1-P3)
- S: Statistics (3 sub-strands; S1-S3)

The national curriculum (2014) aims of solving problems, fluency and reasoning are reflected within the cognitive domain and including them in the test framework ensures that the thinking skills and intellectual processes required for the key stage 2 mathematics test are made explicit. There are four cognitive domains, each of which is further detailed into four difficulty levels:

Depth of understanding: the demand associated with recalling facts and using procedures to solve problems.
Level 1: Recall of facts or application of procedures
Level 2: Use facts and procedures to solve simple problems
Level 3: Use facts and procedures to solve more complex problems
Level 4: Understand and use facts and procedures creatively to solve complex or unfamiliar problems

Computational complexity: the computational demand of problems
Level 1: No numeric steps
Level 2: One, or a small number of numeric steps
Level 3: A larger number of numeric steps. All steps are simple
Level 4: A larger number of numeric steps, at least one of which is more complex

Spatial reasoning and data interpretation: the demand associated with the representation of geometrical problems involving 2-dimensional and 3-dimensional shapes and position and movement. This strand is also used to assess the demand associated with interpreting data presented in tables, pictograms, charts and graphs. The levels reflect both spatial reasoning and data interpretation:
Level 1: No spatial reasoning required, no data interpretation required
Level 2: Manipulation of the geometric information is required, select and retrieve information
Level 3: Complex manipulation of the geometric information is required, select and interpret information
Level 4: Interpret, infer or generate new geometric information, Generate or infer new information from data

Response strategy: the demand associated with constructing a response to a question
Level 1: Select one or more responses or construct a simple response
Level 2: Construct a small set of responses
Level 3: Construct a straightforward explanation. Shows evidence of a method
Level 4: Construct a complex explanation

There were four different types of items on the 2010 to 2017 tests:

- multiple choice, where children are required to select their response from the options given
- matching, where children are expected to indicate which options match correctly
- true–false, where children are required to indicate whether each of a set of statements are true or false.
- constrained questions, where children are required to provide a single or best answer. These might involve giving the answer to a calculation, completing a chart or table, or drawing a shape. For questions worth more than one mark, partial credit is available.
- open (less constrained) questions, where children are required to communicate their approach to evaluating a statement or problem, or where there are more than one correct responses (as in items that have a level 3 and 4 response strategy).

Step 2: Teacher interviews

In the second step of our study we interviewed year 6 teachers about their understanding of the design of the KS2 mathematics test and their instructional strategies to respond to choices in test design. Teachers were interviewed between February and May 2015, prior to new curriculum tests, and again in June/July 2016 after the administration of the new test. The second round of interviews allowed us to understand how, and how quickly teachers respond to changes in tests.

We interviewed 30 Year 6 teachers from schools across Greater London and Kent, performing above and below the floor standards and with pupils from a range of backgrounds, including both advantaged and disadvantaged. The sample is not representative for the country as only 767 schools are below the primary school floor standard, representing 6% of state-funded mainstream schools[6]. Including a larger group of teachers from schools below the floor standard however allowed us to capture and understand the practices of teachers working in potentially higher stakes settings and to compare those to teachers in higher performing schools. Our aim is not to generalize findings but capture a broad array of practices and build theory about how teachers' analyse tests and come to understand patterns on the test. Table 2 summarizes the initial sample of teachers for the first round of interviews in 2015.

Table 2. Sample of teachers

Above floor targets					
Transcript No	FSM	Attainment %	Progress %	NOR	No of Y6 classes
3	1.9	100	100	216	1
5	5.4	96	100	242	1
25, 26	7.7	90	98	699	6
2	10.6	100	83	94	0.5
9	9.5	86	92	377	3
8	10.5	97	97	219	1
6	4.3	85	100	255	2
10	5.1	81	100	313	1.5

[6] source: https://www.gov.uk/government/statistics/national-curriculum-assessments-at-key-stage-2-2012-to-2013; http://www.education.gov.uk/schools/performance/2013/download_data.html

27, 20	16.1	85	100	459	2
17	33	65	87	234	1
16	33.2	83	90	210	1
28	6.1	82	84	643	3
29	10.5	85	92	243	1
30	59.6	70	91	343	1.5
Below floor targets					
12, 13	38.3	43	65	215	1
7	53.4	45	73	189	1
15	27.4	58	78	194	1
4	37.9	41	58	338	3
23, 22, 21	49.3	53	79	353	2
1	47.6	47	72	212	1
11	40.7	46	64	167	1
19, 24	44	51	80	286	1.5
14	34.2	45	60	274	1
18	33.5	29	64	279	1

Notes: FSM is percentage of children on free school meals; Attainment % is percentage of children achieving Level 4 or above in mathematics in 2013; progress % is percentage of children making expected progress in mathematics in 2013; NOR is number on roll (number of pupils in the school);

During the first round of interviews, teachers were first asked to describe clusters of similar items based on their characteristics, the difficulties their pupils face in answering (clusters of) items and how they would prepare pupils for them (e.g. the strategies they would use to prepare pupils to correctly answer items in each cluster, the specific time frame within which these strategies would be used, whether their strategy use might differ in the absence of testing and how school level arrangements, such as buy-in of external coaching services, affect their instruction). In a second part of the interview they were provided with copies of all the items from the 2014 test (paper A and B) and were asked to classify each item as appearing (almost) every year on the test (EY), being high weight in number of marks for this type of item (HW), or random (R). The interview guideline can be found in Appendix 1.

Two researchers coded all the interviews for clusters of similar test items, teachers' descriptions of student performance and their instructional strategies in relation to each cluster. The summary of clusters from the interviews was subsequently validated in a follow-up questionnaire to the same sample of teachers in which we asked them to what extent they recognize the clusters. 19 teachers completed the survey (8 from schools performing below the floor target, and 11 from schools above the floor target). Almost all of the clusters (apart from clusters of items 'measuring with a ruler', and 'calendar') are recognized by the majority of teachers in our follow-up questionnaire.

In 2016 (after the administration of the new key stage 2 mathematics test), the same teachers from the first year of interviews were approached. Unfortunately, in a third of the schools, the year 6 teacher was no longer teaching year 6 or no longer teaching in that school. We asked the new Year 6 teacher to participate instead. Appendix 2 summarizes our sample of teachers in the first and second round of interviews; each teacher has a unique code (T1, T2 etc).

In the second round of interviews we asked teachers specifically about:

- Changes in maths teaching in response to the new national curriculum;

- Changes to the key stage 2 tests: whether they can describe changes, how they learned about changes, how they have prepared children for the test, how changes have affected their teaching, and how children performed on the test;
- Changes in the clusters: we asked teachers to analyse the 2016 test and which items fit the 43 clusters identified in the first interview round.

Step 3: Review of effectiveness of teachers' instructional strategies
The final third step of our study included a review of the effectiveness of teachers' instructional strategies. A UK and US-based mathematics expert (Jeff Goodwin and Jon Star) reviewed the literature on teachers' instructional responses to word problems and context-free calculations (the dominant clusters of frequently occurring test items according to teachers in both round of interviews) and discussed the effectiveness of these responses as well as implications for test design.

4. Findings

This chapter first presents our expert analysis of the KS2 mathematics tests, administered between 2010 and 2017 and patterns in the test. We then present the interview results which first includes teachers' descriptions of clusters of recurring items and their instructional responses to these clusters, and second their description of general changes in the 2016 test. Chapter 5 reflects on the effectiveness of teachers' instructional strategies.

4.1 Expert analysis of 2010-2017 tests

Our analysis showed that the curriculum strand 'calculation' has the most items over the years with 'ratio', 'algebra' and 'position' the fewest. The number of items to test the various strands varied over the years, where the change in test seemed to have led to an increase in number of items to test 'frequency' and a reduction in 'statistics'.

Our analysis indicated no clear change in items for the various sub-strands. The sub-strands that had the most items before 2016 (more than 23 in total for the years 2010-2015) are:

- C4: Add / subtract to solve problems
- C8: Solve problems (commutative, associative, distributive and all four operations)
- M9: Solve problems (money, length, mass / weight, capacity / volume)
- G2: Describe properties and classify shapes
- S1: Interpret and represent data
- S2: Solve problems involving data.

The 2016 and 2017 tests saw relatively many items (more than 10 in total) to test C6 (multiply / divide mentally), C7 (multiply /divide using written method) and F4 (add / subtract fractions).

Across the years, there were also 7 out of the 56 sub-strands with no, or only one item.

Depth of understanding was mostly tested on level 2 and 3 between 2010 and 2015, while the 2016 and 2017 test included more items on level 1 and level 4. The increase in level 4 items can be explained by the discontinuation of the separate level 6 test and incorporating these more complex items in the regular test papers.

Most items on the 2010-2017 tests included one numeric steps where the 2016 and 2017 tests saw an increase in items with a larger number of simple numeric steps and a decrease in items with no numeric steps. Most items required no spatial reasoning across the years, nor any data interpretation, whereas most items required children to select or construct a response. The number of items requiring children to construct a set of responses seemed to have decreased over the years.

Test papers between 2010 and 2017 primarily included constrained questions. The change in test in 2016 didn't have any effect on the choice of question format, although two years of data for the changed test doesn't allow for any confirmative conclusions on a change in pattern.

Detailed tables of number of items per strand, sub-strand, cognitive domain and item format can be found in appendix 3.

4.2 Teachers' report of clusters of similar test items

We asked teachers in the first round of interviews to describe clusters of similar test items on the 2010-2015 KS2 mathematics test. In the second round of interviews we asked teachers which of these clusters were tested again in the 2016 test. We then presented teachers with the 2016 Mathematics test and asked them to allocate each item to one of the clusters. The findings here represent a small sample of teachers

from schools performing below and above the floor standard. The sample was chosen to present a wide-ranging set of understandings of the test and instructional strategies, but not for the potential to generalize findings to year 6 teachers in general.
When describing the test and recurring types of items, teachers talked about a set of skills or a cluster of items (e.g. in ‘data handling’, ‘understanding time’, or ‘shape’), while other descriptions would refer to specific types of items, such as when describing items that require children to read data from a table or a graph. These descriptions suggest that teachers have a different logic or understanding of patterns on the test compared to the framework used by test developers which samples sub-strands from strands, and strands from the mathematics domain.

Table 3 presents the 30 clusters of similar test items described by teachers and compares the number of teachers confirming the clusters for the 2010-2015 tests versus the 2016 test. The comparison suggests a decrease in 2-step money problems, context-free calculations, probability, explanation, matching fractions and decimals, data handling (line graphs, bar chart and pictogram) and transforming shapes, while more teachers recognized missing digits, negative numbers, percentage, Venn diagrams and ‘coins to make a total’ as typical items on the 2016 test.

However, the comparison of teachers’ reports of clusters of similar test items before and after scrutinizing the 2016 test indicates that our findings are greatly impacted by teachers’ recall of the content of the test and how each test paper is named. The extent to which teachers recognize the 27 clusters increases once papers are scrutinized, whereas including all the ‘context-free calculations’ in a separate paper, named ‘arithmetic’, seems to have teacher relabel these types of items. As teachers have only administered the new test once (whereas most teachers will have seen multiple versions of the old curriculum tests), this will likely have biased their perception of clusters of similar items on the new test. A detailed description of the 30 clusters can be found in appendix 4.

Table 3. Comparison of clusters of similar test items 2010-2015 versus 2016 test

Clusters described by teachers	**Total number of teachers mentioning/ describing the cluster in round 1 interviews**	**Total number of teachers confirming clusters for 2016 test before reviewing papers, round 2 interviews**	**Total number of teachers confirming clusters for 2016 test (based on analysis of individual items), , round 2 interviews**
General Word Problems	16	15	23
2-step money problem	23	12	25
Context-free Calculation (horizontal layout)	23	8	9
Multiplication Grid	1	2	0
Missing Symbol	1	5	0
Missing Digits (to complete calculation)	3	10	24
Number Cards	2	3	0
Negative Numbers	3	11	20
Function Machine	1	2	18
Missing Numbers (to complete number sentence)	8	7	25
Number Sequence	7	8	27
Algebra	9	11	28

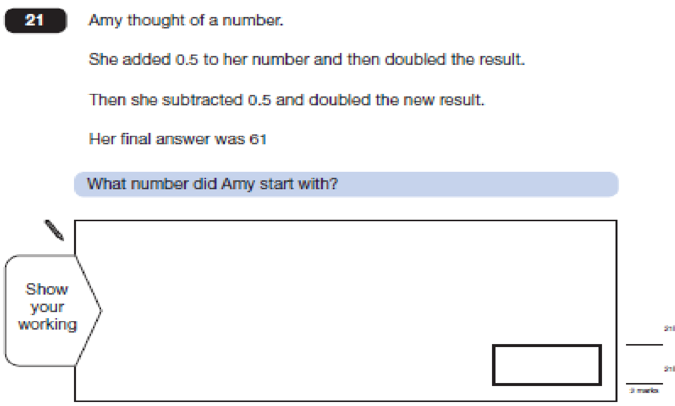
Ordering mixed fractions, decimals and/or percentages	6	2	0
Ordering Decimals (Different number of decimal places)	9	5	23
Ordering General (other than decimals/fractions; e.g. time)	3	5	26
Ratio and Proportion (often cooking)	13	9	4[7]
Percentage	4	13	24
Probability: Spinner, Marbles and Beads	12	0	0
Explanation (e.g. probability)	12	4	24
Matching fractions and decimals	12	4	0
Equivalent fractions (shaded diagram and numbers)	14	15	29
Matching shapes to names	5	0	0
Length Problem	3	3	14
Measuring with a Ruler	2	3	0
Reading a Scale	5	9	12
Conversion of Weights and Measures	10	4	20
Data handling: Line graph	23	0	0
Data handling: Bar chart	10	2	0
Data handling: Pie Chart	8	1	0
Data handling: pictogram	8	0	0
Carroll diagram	1	1	0
Venn Diagram	5	11	27
Reading a Table	4	8	25
Properties of Shapes (linked to matching shapes)	7	6	13
Angles & Protractor	15	13	28
Area and Perimeter	8	7	27
Coordinates (linked to line graphs and negative numbers)	8	7	25
Transforming shapes (rotation, symmetry and reflection, translate shape)	16	12	28
Time Problem and timetable	13	15	29
Calendar	1	0	0
Coins to make total	2	9	11

The three clusters described by most teachers in the first and second round of interviews (general word problems, 2-step money problems, context-free calculation) are outlined in more detail below. The description shows how teachers in our sample understood and analysed tests, and which features of the test were relevant to them. Further quotes of teachers describing the cluster are included in appendix 5.

[7] N.B. R&P identified from 'cooking' context

4.2.1 Word problems (general)

In the first round of interviews, half of the teachers in our sample talked about ‘word problems’ and described these as questions or calculations that are set in a context or as a problem. The picture below shows an example of a general word problem.



Teachers describing this type of item would emphasize the ‘wordiness’ of items which require children to read and decide on which calculation they need to do, instead of straightforwardly presenting the numbers and calculation. Two teachers in the first round of interviews for example say: ‘just reading all of that, you can just see the children going ‘aah, where do I start’? and ‘there aren’t very many (questions) in terms of straightforward calculations that the children are asked to solve, it usually is through some sort of problem solving’.

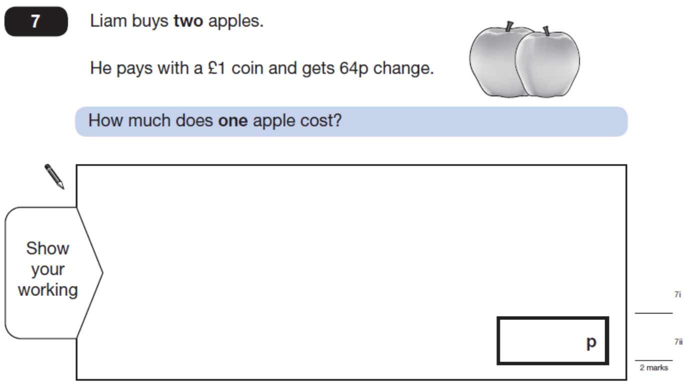
In the second round of interviews, most teachers confirmed that the new test continues to have word problems which fit this description, but they explained how some items that can be characterized as word problems have new features or are different in the following ways:

- Some of the word problems (e.g. B19) are wordier and have more information than previous tests; children need to unpick more information.
- Some of the word problems (e.g. B19) require children to do more steps, for which they also have to access multiple skills.
- Difficult word problems (with many steps or complex numbers) are now also placed at the start of the paper, where, in the past, they would come up at the end of paper.

4.2.2 Money Problem (often 2-step)

Most teachers in the first round of interviews talked about 2-step money problems as a type of item often included on the 2010-2015 test. In their explanation of the item they talked about problems requiring more than one calculation involving money (e.g. addition then subtraction), a problem in which children are going shopping, purchasing more than one item and requiring change. Items within this cluster would, according to some teachers, usually receive two marks, and include a 'show your working' box. Some teachers described how this cluster would ‘involve a picture’, introducing a problem around shopping and money where pupils have to do two different operations, involving addition, subtraction, multiplication or division to solve the problem. A teacher in the first round of interviews for example says ‘(…) multistep problem, strawberries cost seven pounds fifty, sugar costs seventy nine p per kilogram and ten glass jars cost six pounds ninety, so we are multiplying here, we are dividing, multisteps..’.

The following is an example of a 2-step money problem:

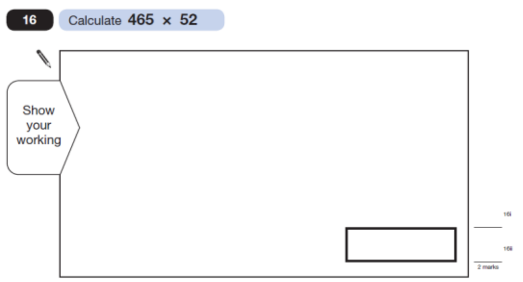


Some teachers described how the problem would generally be presented at the start of the paper when the operation is relatively simple (e.g. small numbers to add or subtract) or towards the end when larger numbers are introduced requiring multiplication and division. Items would follow the same format of presenting information with a picture, introducing the question and asking pupils to explain their answer in a box, where they would generally get 2 marks for the correct answer (1 mark for choosing the correct operation, and 1 mark for the correct answer). We considered this a sub-cluster of the cluster 'General Word Problems' as it is a specific type of calculation in a context.

In the second round of interviews, a third of the teachers also talked about multi-step problems, sometimes specifically referring to 'money problems' but often also to other types of contexts for calculations (e.g. measure, ratio). These teachers explained how the new test seemed to have more steps than before (e.g. 3 or 4 steps instead of 2), how these steps would be 'trickier' and require more 'reasoning' than before as children had to break down more information to understand which calculations were required, the language of the problem being more sophisticated, and the answer also needed more calculations, with some of the amounts in decimals (in item B19: strawberries, sugar and jars; mixing the weight with the cost of jars), involving fractions, crossing the 100s barrier, or requiring children to work backwords, and the question would be worth 3 marks instead of 2. Some teachers also said that some of the questions (e.g. B19) would have been a level 6 question and included in a separate paper in the previous test.

4.2.3 Context-free calculation (horizontal layout)

Most teachers described this cluster in the first round of interviews, referring to 'your bog standard addition and subtraction', 'straight sums', or 'written calculation with no context whatsoever'. Features they would talk about were: items requiring pupils to complete a context-free calculation, items set out in horizontal format, typically involving three-digit number(s), items including a 'show your working' box. Some teachers explain that a correct answer usually scores 2 points, except when the calculation involves addition or subtraction, which were more likely to be worth one mark only and have no designated working out space. The following picture is an example of a context-free calculation:



The new test had a separate arithmetic paper which only had context-free calculations and a small number of teachers explained, in the second round of interviews, how the new test would therefore include many more of these types of items. Even the old 'mental maths' paper would often have items with some contextual information according to these teachers, so the number of actual context-free calculations has risen substantially. One of the teachers explained how the 'grid background' in the space where children note down their answer is new, while another teacher explained how items are now more often in a vertical lay-out than horizontal. Two teachers said that context-free calculations are new types of items that did not appear on the previous test.

4.3 Teachers' instructional responses to clusters of similar test items 2010-2015

In our first round of interviews we asked teachers how they prepare children to do well on the clusters they described. Here we focus on their responses to the above three clusters: word problems, 2-step money problems and context-free calculation. Exemplary quotes are included in appendix 5.

4.3.1 Word problems (general)

The teachers who mentioned word problem as a frequently tested item on the 2010-2015 test described a number of strategies to improve pupils' performance on word problems. They would for example teach them to unpick the problem by underlining key words that would signal a specific operation, where they would talk to peers about a word problem to learn how to identify relevant information, or where they are asked to write word problems themselves. Key vocabulary would include words like 'shared' which would signal 'division', looking for key numbers, or units of measure they have to provide their answer in. One teacher for example explains: 'I've just given them a sheet of problems and said we are not even going to work it out, but just write with it whether it's a multiplication or a division, or it's an addition or a subtraction, so it's just kind of getting them to think of key words, like how many more than the number, or altogether'.

Some of these teachers also make a habit of introducing any kind of mathematical skill in a real-life context (sometimes as part of other subjects) to familiarize pupils with word problems and make sure they can access questions that are set in word problems. One teacher even explained how this would involve some practical work in a context that would show up in the word problem, such as having children cook something so they understand what is going on in the situations presented in word problems, and would know when their answer is potentially 'ridiculous'.

Another teacher also referred to the 'show your answer' box in word problems and how she taught pupils to make notes on any part of the test paper, but only include the correct answer and operation in the answer box. One teacher explained how she has discussed potential answers to word problems with her pupils, highlighting the fact that answers to word problems generally do not have a remainder: '
'The fact that it's a word problem often puts it in context, which means a remainder doesn't actually make sense, most of the time anyway, you don't have point two of a child I say to them..'.

A final strategy to allow pupils to answer word problems is reading out problems during test administration. One teacher explained how reading out word problems and emphasizing parts of the question helps pupils understand the item and helps them process what they need to do.

4.3.2 Money problem (often 2-step)

The first instructional strategy described by teachers in our sample who mention 2-step money problems involves pupils reading the item closely and underlining key words to understand the operations (in two steps) they have to do and which words in the problem are relevant for identifying the operation.

One teacher mentioned the ‘five finger rule’ of steps pupils need to go through when solving a two-step world problem. These steps were mentioned by other teachers as well and include: read the question, underline the key vocabulary, work out what operation it is, solve it, do the inverse.

A small number of teachers would also ask pupils to think about the plausibility of the answers they come up with, or ask pupils to draw out the situation and visualize it as a way to understand the problem and the operations involved, or they would set up games in which pupils have to use money to learn and understand the value of coins. One teacher for example would for example say to pupils: ‘If you had 10 pounds to start with is it realistic to say you’ve got 11 pound 23 now?’.

Some teachers would also ask pupils to convert pence into pounds, or pounds into pence to ensure they have the correct numbers when subtracting, adding, multiplying or dividing the numbers:
‘I always say to them look if it’s got pence there and pounds there you have a choice, you either everything becomes pence or everything becomes pounds, it’s making those decisions and choices.’

Two teachers also explained how they try to make pupils aware of cues in the question that would suggest they need to do two different operations; cues are the number of marking points and the inclusion of a box in which they have to show their work.

Some teachers also have pupils practise with the specific questions, often at the end of a week of teaching specific operations. These teachers would go onto Testbase to download example questions for pupils to practise with.

Some of the teachers have pupils write their own money word problems to familiarize them with these types of questions. Only one teacher however explained how she would use such a writing exercise to make pupils aware of the structure of the item and the operations that would be required to solve the problem. She explains:
‘So I think that’s what we have to do, break it down, underline it and say right, if I took this out could you still do the question? And then we’ll reverse it when they are more confident, and I’ll put up a kind of closed text and just say right, I’ll write this, and wherever I’ve put a blank you fill in a name here, you fill in an amount here, and another amount here, and they’ll sort of see oh they are the key bits, actually the fact that she buys coffee, that’s not relevant, so it’s kind of breaking it down, getting them to write it themselves.’

4.3.3 Context-free calculation

Teachers who described context-free calculations mentioned a number of methods they would teach pupils to calculate the correct answer on these types of items, such as chunking, standard algorithm, the grid method, the bus shelter method, the ladder method, a formal method of long division, the column method. Some of the teachers explained that they would particularly teach the grid method to lower ability pupils and how lower ability pupils would use a number line to add and subtract, while the higher ability pupils would be able to use the standard algorithm. A number of teachers explained how they decided to teach children to use the chunking method or the bus shelter method for division, depending on the size of the numbers involved. As lower ability pupils would often get the place value wrong, teachers would do a lot of work with place value grids, where pupils have to physically take the number and put it in the correct place value, teaching them to start from right to left. Some teachers also specifically talked about a sequence of skills they would teach, starting with the grid method and small round numbers, and then introducing large numbers and the column method, before moving on to calculations with decimals.

A number of teachers also described how they would teach pupils to write down how they calculated the answer as they are required to show their working on the test. These teachers would model the process of writing down the steps in multiplying, dividing, adding or subtracting numbers, and would ask pupils to use pencil and paper when solving these items. They would teach pupils to always write something down on the test paper as they may receive one mark for part of their work, even if they don't have the correct answer. Pupils need to learn how to show their work as the item is asking them to do this; they may miss one mark if they only do a mental calculation. One of the teachers also explained how she asks pupils to first estimate the correct answer (rounding up the numbers and doing a mental calculation) before working out the calculation on paper as that would allow them to check their answer. One teacher furthermore explained how she teaches pupils to count the zeros when they are multiplying two numbers with tens or hundreds, to make sure they have enough zeros in their answer.

Some teachers encouraged pupils to use any method they are comfortable with, although after the change in test they seemed to instruct pupils to use traditional methods as those would get them higher marks. Another strategy described by two teachers is to teach pupils to transfer a calculation from a horizontal format to a vertical one as that would be easier to work with. Pupils are taught how to correctly line up the numbers, purposefully asking them to transfer numbers with a different number of decimal places into a vertical format, as they would often get the place value wrong. One teacher mentioned how this is a specific skill pupils need to learn as they often feel that they can't change the outline of the numbers on the test and need to solve it horizontally:
'One thing we have started doing is you know when I was saying about writing the questions, addition question, they never write their addition questions as a column, they always write the numbers beside each other, and we have made a conscious effort to write, if we are ever doing anything on the board we write it out like that and then put it into a column, because a lot of the children will just see the two numbers next to each other and then will think that's how I have to solve it. They sometimes worry about changing things, because the test says this is how it's done, so we have made a concerted effort on that as well.'

Most of the teachers also frequently practise times tables with pupils, and go over number facts every week to make sure that pupils can recall these quickly when tested. One teacher stated that she spends at least 50% of her time teaching calculation as most of the items on the test would involve number calculation, while another teacher explained that in the Autumn term she would start each lesson with straightforward calculations 'to keep them really, really fresh on their toes'.

4.4 Teachers' responses to changes in the test

In the second round of interviews we asked teachers to describe the changes in the 2016 test to us. This section summarizes teachers' more general understanding of how the test changed in their views, and how this affected their teaching. Three changes were mentioned by most of the teachers: the removal of levels, the introduction of an arithmetic paper (and the removal of the mental maths paper), and the introduction of 'reasoning' papers.

4.4.1 Removal of levels

In the interviews in 2015, levels were one of the key topics teachers talked about when we asked them about notable features of the test that would inform their teaching. Most teachers explained how each of the two written Maths test papers would start with easy level 3 questions, have level 4 questions in the middle and finish with the difficult level 5 items at the end. This order of questions according to difficulty level would allow the lower attaining children to access the test, according to these teachers, and would build their confidence in answering the questions and their motivation to do well on the test. After the introduction of the new curriculum tests in 2016, performance of children is no longer reported in levels (where children are expected to reach at least level 4 and a separate level 6 test was administered to highest performing pupils), but instead of scaled scores.

The majority of teachers told us in the second round of interviews, how all the questions are now 'at level 5' and how some of their lower attaining children 'stared at them in horror' when opening their test booklet, asking them where the easy questions had gone. Some of these teachers described how the change in order of items according to difficulty had an impact on children's motivation and confidence in test taking.

The removal of levels also appears to have a profound impact on how some teachers come to understand and teach mathematics. Prior to the introduction of scaled scores, some teachers would talk about gradually building up the level of difficulty when teaching specific mathematical content areas, such as 'number sense and calculation', 'data handling' or 'shape and space'. Level 3, 4 and 5 test items on past key stage 2 test papers would help them understand the hierarchical nature of mathematics and how to introduce children to for example increasingly more difficult calculations (e.g. moving from one to multistep problems, or from adding and subtracting whole numbers to adding and subtracting decimals). Resources such as Test Base would allow them to access available questions according to content area and difficulty level and they could simply download relevant questions when teaching a specific skill. Now that the levels have been removed, some of the teachers explained that they just focus on getting all pupils to perform at level 5 in number and calculation as this is where most of the marks on the test are given, where some teachers say that they now hardly teach shape and space at all. These teachers also talked about moving towards a more 'mastery style' of teaching where they ensure that all pupils master the basics before they move on to teach more complex skills or other (more complex) content domains, such as algebra or geometry.

4.4.2 Introduction of an arithmetic paper

The mental mathematics paper, administered before 2016, involved a 10-minute assessment, administered by playing a CD, in which 11-year-old pupils were expected to carry out 20 calculations in their heads, given 5 seconds, 10 seconds or 15 seconds for each one, and asked to write down the answer to each question, without access to paper to make jottings for working out. Instead, in addition to two papers testing reasoning, children now have to sit an arithmetic paper with 36 questions, lasting 30 minutes. The arithmetic paper covers context-free calculations for all four operations, including the use of fractions, percentages and decimals. Squared paper is provided in the answer area, for children to show their working. 'Working out' is necessary in some cases, as the arithmetic questions now include much more complicated calculations which cannot be answered by mental calculation alone. In some questions, credit may be given for using a traditional algorithm, such as long multiplication.

The change in types of questions seems to have triggered a change in the planning of a teaching week for some teachers. A small number of teachers in the second round of interviews talked about how they have moved away from mental maths sessions, which they would do on a weekly basis where children would rehearse for the mental mathematics test using one of the many past papers. Instead, these teachers now seem to rehearse for the arithmetic paper each week, such as practising eight arithmetic questions every day, particularly between January and May.

Some teachers also talked about the need for children to work at speed to complete 36 questions in 30 minutes and the huge amount of practice required to achieve this, as some of the types of calculations require much more working out. Some teachers explained how they have taught traditional algorithms to children (as this is considered to be the preferred method that would get the highest marks on the test) where they might previously have allowed them to use a less formal, but slower, method with which they were more comfortable.

4.4.3 Testing 'reasoning' in paper 2 and 3

Paper 2 and 3 in the new 2016 test are titled 'reasoning', where before the first two test papers were a non-calculator paper 1, and a calculator paper 2. Teachers have various interpretations of the change where some talked about how pupils are now required to explain their answer and communicate their thought processes more effectively, whereas other teachers referred to an increased emphasis on problem solving where pupils have to learn to work more systematically (e.g. hypothesizing, looking at conjectures, trying to prove or disprove).

As test papers 2 and 3 (with problem-solving items) are titled 'reasoning', it seems that this will have an effect on the how problem solving is viewed in the future, where the second aim in the national curriculum on reasoning is considered a similar skill by some of the teachers in our study.

5. Reflections on the effectiveness of teachers' instructional strategies

The national curriculum as introduced in 2014 has three aims for key stage 1 and 2:

1. become fluent in the fundamentals of mathematics, including through varied and frequent practice with increasingly complex problems over time, so that pupils develop conceptual understanding and the ability to recall and apply knowledge rapidly and accurately.
2. reason mathematically by following a line of enquiry, conjecturing relationships and generalisations, and developing an argument, justification or proof using mathematical language
3. can solve problems by applying their mathematics to a variety of routine and nonroutine problems with increasing sophistication, including breaking down problems into a series of simpler steps and persevering in seeking solutions.

Word problems particularly aim to teach the second and third aim, while context-free calculations assess fluency in the fundamentals of mathematics. In this chapter we reflect on the effectiveness of teachers' instructional strategies to meet these aims, and whether the specific format and content of word problems and context-free calculations incites potentially effective strategies. As the literature on teaching primary mathematics is rather conflicted and our work did not include a systematic literature review, we can only offer a set of reflections for further discussion.

5.1 Fluency and context-free calculations

In our study, teachers described a range of strategies to teach pupils to calculate the correct answer on context-free items. Here we focus on two frequently mentioned methods: practicing number facts and times tables to improve fluency, and a standard algorithm (e.g. column method) to do calculations.

These methods are often referred to as 'traditional methods' to teach mathematics where pupils are taught to use one method to complete a calculation through direct instruction and rote learning, and where that method is taught in isolation rather than in a context. Practicing each method, the number facts and times tables, is expected to help pupils develop their fluency and support quick and accurate execution (Hewitt, 1996; Gersten, and Chard, 1999; Pellegrino and Goldman, 1987).

These views have been much contested in the mathematics education community since the early to mid-20th century. Advocates of the 'reform movement' argue that traditional methods of instruction over-emphasize memorization and repetition and fail to promote conceptual understanding of mathematics (Hiebert and Carpenter, 1992; National Council of Teachers of Mathematics, 1989, 2000). A discussion paper from the University of Chicago[8] on the changing place of algorithms in school mathematics for example argues that traditional methods fail with a large number of pupils who never become fully proficient in carrying out algorithms for the basic operations because of an overemphasis on procedural proficiency with insufficient attention to the conceptual basis for the procedures. This unbalanced approach produces, according to colleagues at the University of Chicago, pupils who are plagued by "bugs," such as always deducting the smaller digit from the larger in subtraction.

The Association of Teachers of Mathematics4 also argues that the focus on column methods of division and multiplication for multi-digit numbers skews the curriculum against enjoyment and understanding of mathematics and towards following procedures, and that column methods are unnecessary as a foundation for secondary mathematics.

[8] http://everydaymath.uchicago.edu/about/research-results/algorithms.pdf

As an alternative, reform mathematicians have advocated a reduced emphasis on instruction in one set of traditional algorithms in favor of student-generated approaches and alternative strategies. Arguably, a major thread of the curricular and instructional reforms in primary mathematics in the US and elsewhere over the past decades has been to provide pupils with a chance to contribute their own ideas and strategies for problem solving, rather than being directed by the teacher to learn/use/memorize one particular way. Teachers would introduce a problem (e.g. through schema-based instruction, see 5.2), and then pupils spend time discussing how they solved it. Pupils may use a variety of strategies for solving the problem – some idiosyncratic, some inefficient, but some interesting. The teacher then uses the range of strategies presented to engage pupils in rich mathematical conversations, discussing the value of strategies used by pupils. This 'thread' is usually carried by the term "student-generated" and there is a strong belief that the use of student-generated and non-traditional algorithms improves short-term and long-term mathematical understanding, improves pupils' attitudes and beliefs about mathematics, and allows teachers to build open and multi-cultural communities of mathematics learners (e.g., NCTM, 2000; Forman and Ansell, 2002; Philipp, 1996; Randolph and Sherman, 2001; Groth, 2007; Ellis and Yeh, 2008; Carroll, 1996). Interestingly, the word problems (discussed below) would offer teachers with a variety of representations to support pupils' understanding of the various algorithms and procedures.

5.2 Word problems to test problem solving

In our interviews, teachers talked about how they employed various strategies for helping pupils read words and understand the problem context, including reading the problem aloud, drawing pictures to represent the problem context, and/or providing pupils with practical work in the context to aid in developing familiarity (e.g. cooking). Once the pupil understands the problem context, he/she must be able to identify the most important mathematical words in the problem statement, look for (and underline) words that have mathematical meanings or connotations (e.g. divide, more), and/or identify words that appear to be irrelevant to the problem and eliminate them (e.g. striking through) from consideration.

When these two tasks are accomplished, the perception among teachers seemed to be that the hardest work of solving a word problem had been completed, given that understanding the words/context and identifying the most important mathematical words essentially allowed a word problem to be transformed to a more straightforward pure computation problem.

These tasks – reading and understanding the words in a problem – play central roles in commonly used heuristics for word problem solving. Heuristics are multi-step strategies that are presumed to help pupils solve word problems. General heuristic strategies such as drawing a picture, underlining key words, and re-reading the problem are derivative of Pólya's (1990/1945) four-step problem solving model (i.e., *understand the problem*, *devise a plan*, *carry out the plan*, and *look back and reflect*). The most commonly used heuristic among teachers in this study was referred to as RUCSAC, which stands for Read, Understand, Choose, Solve, Answer, and Check.

Teachers' reported reliance on problem solving heuristics is potentially problematic. These heuristics have come under scrutiny in mathematics education research for the failure to reliably lead to improvements in pupils' performance in solving word problems (Lesh & Zawojewski, 2007; Schoenfeld, 1992). Schoenfeld (1992) argued that these types of problem solving strategies are descriptive and do not provide the necessary detail for pupils who are not already familiar with the strategies to effectively use them.

This critique of problem solving heuristics is particularly relevant to what teachers report in the present analysis. Note that the strategies that teachers most frequently indicate using lie within the initial steps of the heuristics: *read* and *understand* in RUCSAC, and *understand the problem* in Polya's framework. Schoenfeld's (1992) concerns about heuristics are particularly relevant for these steps. Understanding the problem context via reading the problem's words may not provide sufficient detail about the mathematics of the problem to aid problem solving. Drawing a picture is only helpful when the picture conveys important mathematical information about what the problem says and asks for. Similarly, drawing a picture of pence and pound coins only helps when the depiction of (and quantities of) the coins relates to the mathematics of the problem. (And if the student understands enough about the mathematics of the problem to draw a helpful picture, then he/she may not need the picture to solve the problem.)

In sum, reliance on the types of problem solving heuristics that teachers report using here – ones that focus on reading and understanding the words and situational context of a problem – have been found in the research literature to be neither necessary nor sufficient to promote problem solving success.
Rather, the literature suggests that pupils' difficulties with word problems result from a failure to activate relevant schema during problem solving and a more appropriate response would be schema-based instruction (Fuchs, Fuchs, et al., 2004a; Fuchs, Fuchs, et al., 2004b; Jitendra & Hoff, 1996; Jitendra, Star, et al., 2009; Jitendra, Star, et al., 2011; Jitendra, Star, et al., 2013; Xin, 2008).

Schema are knowledge structures that help the learner organize and categorize problem types in order to determine the best strategies for solving the problem (Marshall, 1995; Sweller, Chandler, Tierney, & Cooper, 1990). The mathematical structure of a problem may not be fully apparent from understanding the problem context, nor may it be illuminated by identifying the mathematical vocabulary of the problem. Features of schema-based instruction include a focus on multiple solution strategies (e.g., discussing multiple strategies for solving the same problem), the use of think-alouds to help pupils monitor their learning and understanding to build metacognitive strategy knowledge, and the use of schematic diagrams. With respect to diagrams, note that schematic diagrams are not merely pictures of the problem context, but rather they show relationships between elements of the problem's mathematical structure (Hegarty & Kozhevnikov, 1999).

A few teachers reported instructional strategies that begin to approximate schema-based instruction. In particular, in describing how she helps pupils with money problems, one teacher notes the following:
And with money problems, difference problems, we'll start putting them into multistep, we'll work from one step to multistep and we'll look for ways in which we can make it really complicated, or we'll ask them to make up their own types of questions, or we'll ask them to use things like the inverse to work a question, and answer back to the question and that sort of thing, and again making them familiar with the types of questions that have come up in the past.

By indicating her interest in making pupils "familiar with the types of questions that have come up in the past," it appears that this teacher's typology of past problems is not organized around context (e.g., shopping problems, money problems) or by generic strategy, as was more typical of other teachers who were interviewed (e.g., "*there is always a shopping type question*"; T29: "*it's the type of question where they can use a related fact*"). Rather, this teacher seems to be aiming to help pupils organize and remember past problems based on their underlying mathematical structure of the problems, which is consistent with schema-based instruction.

Similarly, a second teacher notes:
Shopping and money; ones about shopping, it always follows the same structure of where you've got two structures, and it either will be so and so has bought these things, how much change will they get from ten pounds, or along other lines; if this much has cost this much, that's the same structure. there's two structures that I can think of anyway, yeah, either the money ones always follow the same pattern.

This teacher is the only one who uses the word "structure" to describe a feature of word problems. She describes a structure common to many shopping problems (a person buys some things, and the task is to determine her change from ten pounds), presumably working with pupils on seeing the mathematics behind this structure (e.g., subtracting the total cost of purchases from ten pounds).

6. Conclusion and discussion

External examinations have grown as instruments of control over educational systems in the past decade. Many countries and states nowadays use student assessment results for school and teacher accountability and to monitor and improve the performance of individual pupils, schools and education systems (Eurydice, 2015). The aim of such assessments is to provide teachers with incentives to focus on central standards and to encourage them to adopt more effective practices (Stecher, 2002).

Previous studies have indicated many problematic consequences of high stakes testing. Studies for example show how high stakes tests predictably emphasize some national curriculum standards while consistently excluding others, enabling teachers to teach to the test and inflate test scores. These studies suggest that teachers are familiar with patterns in the test and adapt their teaching in similar ways. As of yet there is little empirical evidence to support such a claim. This report presented the findings of a small explorative study (expert analysis of the key stage 2 mathematics tests 2010-2017, teacher interviews, review of teachers' instructional strategies) of the following research questions:

1. To which features of key stage 2 maths tests are teachers responsive?
2. What instructional strategies do teachers use in response to key stage 2 maths tests?
3. How beneficial or harmful are these strategies in promoting student learning?

6.1 To which features of key stage 2 mathematics tests are teachers responsive?

Our findings suggest that teachers have a different understanding of patterns in the test, compared to test developers. They describe clusters of frequently tested items that incorporate a mixture of strands, sub-strands, representations of content and item formats instead of referring to the hierarchical testing framework which samples items within sub strands and strands. There are three clusters which are described by most teachers in both round of interviews: general word problems, 2-step money problems and context-free calculation. The description of the clusters suggests that teachers' mental model of relevant patterns in the test seems to follow a much more idiosyncratic horizontal structure where they would sometimes describe tested content areas, but also focus on types of items that frequently occur on the tests.

6.2 Teachers' instructional strategies in response to key stage 2 mathematics tests

We asked teachers how they prepare their children for the three clusters of recurring test items (word problems with a subset of 2-step money problems and context-free calculations).

Word problems

Our findings indicate that teachers believe that success on word problems and 2-step money problems comes about when 1) pupils are able to read the words in the problem in order to understand the problem context, and once the student understands the problem context, 2) he/she must be able to identify the most important words in the problem statement, particularly the important *mathematical* words.

Teachers indicated that they employed various strategies for helping pupils read words and understand the problem context, including reading the problem aloud, drawing pictures to represent the problem context, and/or providing pupils with practical work in the context to aid in developing familiarity (e.g., cooking). They supported pupils in identifying important mathematical words by looking for (and underlining) words that have mathematical meanings or connotations (e.g., divide, more), and/or by identifying words that appear to be irrelevant to the problem and eliminating them from consideration.

When these two tasks are accomplished, the perception among teachers was that the hardest work of solving a word problem had been completed, given that understanding the words/context and identifying the most important mathematical words essentially allowed a word problem to be transformed to a more straightforward pure computation problem.

Context-free calculations
Teachers mention a number of methods they would teach pupils to calculate the correct answer on context-free items, such as chunking, standard algorithm, the grid method, the bus shelter method, the ladder method, a formal method of long division, the column method. The use of strategies varies according to ability level of children; e.g. using the grid method for multiplication and number line to add and subtract with lower ability pupils, while the higher ability pupils would be able to use the standard algorithm. Strategies also vary according to the size of the numbers involved, or whether the calculation involves round numbers or decimals.
The 'write down your answer' box motivates teachers to instruct children to write down how they calculated the answer, where teachers sometimes model the process. Pupils are also instructed on strategies to check their answer, how to transfer a calculation from a horizontal format to an (easier) vertical one, how to correctly line up numbers with different decimal places into a vertical format, or how to count the zeros when they are multiplying two numbers with tens or hundreds (to make sure they have enough zeros in their answer). Teachers also frequently practise times tables with pupils, and go over number facts every week to make sure that pupils can recall these quickly when tested. After the introduction of the new test in 2016, teachers spent even more time on improving pupils' fluency in mathematical operations, where traditional formal methods are now prioritized for more complex calculations.

6.3 Effectiveness of teachers' instructional strategies

In our review of teachers' instructional strategies we discussed the focus of some teachers on problem-solving heuristics to support pupils' word-problem solving, as well as the emphasis of some teachers on traditional methods to teach context-free calculation. We discussed how problem-solving heuristics (focusing on reading and understanding the words and situational context of a problem) have been found in the research literature to be neither necessary nor sufficient to promote problem solving success. In our review we explained how schema-based instruction is found to be more effective. Such instruction would focus on the strategy to solve problems, offering pupils the opportunity to look for related problems. This would also allow for problems unfamiliar to the children to be considered and strategies established which could be related to problems in the tests.

Traditional methods to teach (context-free) calculation (e.g. using the column method for division) is much contested in the mathematics education literature where advocates of the reform movement argue that this method fails to enhance children's calculation skills when there is an overemphasis on procedural proficiency with insufficient attention to the conceptual basis for the procedure. Our review indicated how other studies suggest that a large number of pupils find these procedures difficult and would, as a result, make the same mistakes on these items, such as always deducting the smaller digit from the larger in subtraction. As an alternative, many in mathematics education have advocated a reduced emphasis on instruction in traditional algorithms in favour of student-generated approaches and alternative strategies. Such approaches would allow pupils to develop their own strategies to solve problems, interrogating the appropriateness and effectiveness of these strategies in classroom discussions.

6.4 Teachers responses to changes in the test

Following the introduction of the new national curriculum in …, the structure and content of the key stage 2 mathematics test was also revised in 2016. Teachers particularly talked about the removal of levels where the order of items on the test according to difficulty is now random, and where there are changes in the types of papers pupils have to sit: no mental mathematics paper, but instead an arithmetic paper with context-free calculations where formal methods are believed to get higher marks and the introduction of two 'reasoning' papers with word problems. These changes appear to affect both pupils' test motivation, as well as affect teachers' understanding of mathematics and how they teach specific skills. Pupils, particularly those with lower ability in mathematics are used to start with easy questions and would now be discouraged and anxious of failing when encountering difficult items at the start of the test. The effect is however likely going to fade when pupils become familiar with the new tests.

The removal of levels and the random order of items according to difficulty level also seems to have an effect on how teachers understand the hierarchical nature of mathematics and how they introduce pupils to increasingly difficult content (e.g. moving from one to multistep problems, or from adding and subtracting whole numbers to adding and subtracting decimals). After the introduction of the new test, teachers focus on getting pupils to perform at, what would previously be, level 5 questions, particularly in the content domain of 'number' and 'calculations'. They explain that pupils get the most marks in this area and that all pupils now need to reach this benchmark. Given the limited instructional time to prepare pupils for the test, they seem to focus on calculations to ensure pupils pass the test, where some teachers now seem to ignore other, less frequently tested, content domains such as shape, algebra or geometry, or the most difficult types of (previously level 6) skills.

6.5 Implications for policy and practice

Our findings suggest that tests and changes in tests have a profound effect on how teachers understand mathematics and teach mathematics where both the structure of the paper, as well as the format of specific questions informs the instructional strategies they use. Available external resources, such as Test Base, support their interpretation and use of tests and have an important role in how teachers prepare their pupils for the test. Our results provide suggestive evidence that tests have direct effects on how resources are allocated within schools, though we cannot explicitly measure the instructional processes that produced these results. This also means that policy decisions about curriculum standards, how these are tested and how cut scores for proficiency are determined potentially affect school-level decisions about resource allocation, and which content domains are emphasized or de-emphasized. While our results may represent the short-term effects of changes in the test, we expect these consequences to increase over time as teachers become more familiar with the new test.

This small scale explorative study also suggests that the government in England has been successful in achieving its stated aim "to ensure that all pupils are taught efficient calculation methods - rather than spending too much time on confusing, time-consuming methods like chunking and gridding [...] and KS2 tests will be designed to reward pupils whose working shows they have used the efficient methods" (Truss, 2013) despite literature which support the alternative methods (see for example Anghileri, 2000). It is clear from our interviews that, in response to the change from the Mental Mathematics paper to the Arithmetic Paper, teachers have changed their practice considerably – particularly evident in how teachers told us that they plan their week. These changes, with an emphasis on formal written methods, have led to a reduced focus on teaching the 'mental strategies' advocated under the National Numeracy Strategy (DfEE, 1999) and which support a conceptual understanding of Number and Calculation.

There is some evidence of an unbalanced approach according to our research, seemingly a direct result of the changes in the tests. The changes to the test have caused a shift in emphasis from conceptual knowledge towards procedural knowledge in some classrooms, with the potential for instrumental rather than relational understanding. Our research suggests that teachers have not only reduced the emphasis on 'chunking' and 'gridding' but have increased the proportion of time spent on practicing efficient methods to the exclusion of a range of mental strategies that support understanding of number. This is concerning for, as Anghileri notes, "all pupils are expected ultimately to use efficient written methods for calculating but the only way such methods can be meaningful is if they are developed progressively to support and extend mental strategies." (Anghileri, 2000, p.108)

6.6 Implications for further research

Our findings of teachers' description of clusters of similar test items suggest that current methods used to study score inflation do not capture 'teaching to the test' very well. These methods, which are primarily used in the U.S. to understand potential adverse effects of high stakes testing, predominantly compare performance of pupils on frequently tested standards or on high stakes tests to their performance on non-tested standards or on similar low stakes tests. Such studies have for example included an analysis of test-item level data over a number of years to see if pupils perform better on predictably sampled content (see Jennings et al, 2014; Klein et al, 2000; Koretz et al, 2002). Although this methodology provides valuable insights into potential score inflation, our findings suggest that the pattern of pupils' performance across items and content areas does not automatically point to practices of teaching to the test. The methodology assumes that teachers analyse and understand the hierarchical sampling pattern of the test used by the testing agency, while our study indicates that teachers do not interpret the test from such a vertical view of standards down to individual items.

We therefore suggest that the research methodology to study teaching to the test needs to follow a different approach. Instead of looking at patterns in performance of pupils in frequently tested strands and sub-strands, we need to capture the clusters of items and content areas that teachers would describe and analyse their reallocation of (curriculum/instructional) time and resources to these content areas and clusters of test items to understand if and how they are 'teaching to the test'. Such an understanding will help us design tests and support teachers' use of these tests to improve their instruction of mathematics.

7. References

- Advisory Committee for Mathematics Education. (2011a). *Mathematical Needs: The Mathematical Needs of Learners*, London: ACME, The Royal Society
- Advisory Committee on Mathematics Education (2016) Problem solving in mathematics: realising the vision through better assessment. Royal Society
- Alexander, R. (2004) Still no pedagogy? Principle, pragmatism and compliance in primary education. *Cambridge Journal of Education,* 34, 7-33.
- Anghileri, J.: 2000, *Teaching Number Sense*. London: Continuum.
- Anghileri, J. (1989) An investigation of young children's understanding of multiplication. *Educational Studies in Mathematics,* 20, 367-385.
- Anghileri, J. (2004) Disciplined calculators or flexible problem solvers. *Proc. 28th Conf. of the Int. Group for the Psychology of Mathematics Education.*
- Anghileri, J., Beishuizen, M. & Van Putten, K. (2002) From informal strategies to structured procedures: mind the gap! *Educational Studies in Mathematics,* 49, 149-170.
- Anthony, G. and Walshaw, M. (2007). *Effective pedagogy in mathematics/Pàngarau: Best evidence synthesis iteration*, Wellington: Ministry of Education.
- Askew, M. (2008) Mathematical discipline knowledge requirements for prospective primary teachers, and the structure and teaching approaches of programs designed to develop that knowledge. *Knowledge and beliefs in mathematics teaching and teaching development,* 1, 13-35.
- Askew, M. (2012) *Transforming primary mathematics,* Abingdon, Routledge.
- Askew, M., Brown, M., Rhodes, V., Johnson, D., and William, D. (1997) *Effective teachers of numeracy*, London: Kings College.
- Ball, D. L. (2000) Bridging practices intertwining content and pedagogy in teaching and learning to teach. *Journal of teacher education,* 51, 241-247.
- Ball, D. L., Hill, H. C. & Bass, H. (2005) Knowing mathematics for teaching: Who knows mathematics well enough to teach third grade, and how can we decide?
- Barmby, P., Bolden, D., Raine, S. & Thompson, L. (2012) Assessing young children's understanding of multiplication. *36th Conference of the International Group for the Psychology of Mathematics Education.*
- Blinko J. (2007). Mathematics in context. Article in Association of Teachers of Mathematics Journal MT188.
- Boaler J. (1994). When do girls prefer football to fashion? An analysis of female underachievement in relation to 'realistic' mathematics contexts. British Educational Research Journal 20 (5) 551-564.
- Boaler, J. (2008). 'Promoting 'relational equity' and high mathematics achievement through an innovative mixed-ability approach'. *British Educational Research Journal*, 34 (2), 167-194.
- Brown, M. (2002) Researching Primary Numeracy. *Proceedings of the Annual Meeting of the International Group for the Psychology of Mathematics Education.* Norwich.
- Brown, M., Askew, M., Millett, A. & Rhodes, V. (2003) The key role of educational research in the development and evaluation of the national numeracy strategy. *British Educational Research Journal,* 29, 655-667.
- Carroll, W. M. (1996). Mental Computation of Pupils in a Reform-Based Mathematics Curriculum. School Science and Mathematics, 96(6), 305-311.
- Clausen-May T. (2001). An approach to test development. NFER
- Clifton, J.and Cook, W. (2012*). A Long Division: Closing the Attainment Gap in England's Secondary Schools,* London: Institute for Public Policy Research.
- Cockcroft, W.H. (1982). *Mathematics counts: Report of the Committee of Inquiry into the Teaching of Mathematics in Schools under the Chairmanship of Dr WH Cockcroft*, London: Her Majesty's Stationery Office.
- Coles, J. (2012). Putting learning first: from tactics to strategy. Presented at Cambridge Assessment seminar: Learning comes first: Shifting the focus from examining to the curriculum. One Birdcage

Walk, Westminster, London. 31 January 2012.
http://www.cambridgeassessment.org.uk/ca/digitalAssets/199663_Jon_Coles.pdf
- Collins, S., Reiss, M. and Stobart, G. (2010). 'What happens when high-stakes testing stops?: teachers' perceptions of the impact of compulsory national testing in science of 11-year-olds in England and its abolition in Wales'. *Assessment in Education: Principles, Policy and Practice*, 17 (3), 273-286.
- De Corte, E., Greer, B. & Verschaffel, L. (1996) Mathematics teaching and learning. In Berliner, D. & Calfee, R. (Eds.) *Handbook of educational psychology.* New York, Simon & Schuster Macmillan.
- Department for Education (2013) Mathematics programmes of study: key stages 1 and 2.
- Department for Business, Innovation and Skills. (2010). *Science and Mathematics Secondary Education for the 21st Century Report of the Science and Learning Expert Group,* London: DBIS.. IPSOS Mori Social Research Institute. (2010). *Becoming a Teacher: The experiences of STEM teacher,* London: The Wellcome Trust.
- Department for Education and Employment. (1999). *Framework for teaching mathematics from reception to year 6*. London: Department for Education and Employment.
- Department of Education Northern Ireland. (2011a). *Count, Read, Succeed: A strategy to improve outcomes in Literacy and Numeracy,* Belfast: DENI.
- Department of Education Northern Ireland and Department for Employment and Learning Northern Ireland. (2009) *Report of the STEM Review,* Bangor and Belfast: DENI and DELNI.
- Denvir, B. & Brown, M. (1986a) Understanding of number concepts in low attaining 7–9 year olds: part I. Development of descriptive framework and diagnostic instrument. *Educational Studies in Mathematics,* 17, 15-36.
- Denvir, B. & Brown, M. (1986b) Understanding of number concepts in low attaining 7–9 year olds: Part II. The teaching studies. *Educational Studies in Mathematics,* 17, 143-164.
- Dickson, L., Gibson, O. & Brown, M. (1984) *Children learning mathematics: a teacher's guide to recent research,* London, Cassell.
- Education Endowment Foundation. (2017). Improving mathematics in key stages two and three. www.educationendowmentfoundation.org.uk.
- Education and Training Inspectorate.(2010). *Issues from Inspection Transition in Mathematics: Primary to Post-Primary,* Belfast: ETI.
- Ellis, M., & Yeh, C. (2008). Creative arithmetic: Exploring alternative methods. Teaching Children Mathematics, 14, (6), 367-368.
- Evans, A., Hawksley, F., Holland, M.R. and Caillou, I. (2008).'Improving subject knowledge and subject pedagogic knowledge in employment based secondary initial teacher training in England'. *Annual Conference of the Association of Teacher Education in Europe*, Brussels, 13-27 August, 2008.
- Forman, E., & Ansell, E. (2002). The multiple voices of a mathematics classroom community. In Learning Discourse (pp. 115-142). Springer, Dordrecht.
- Fuchs, L. S., Fuchs, D., Finelli, R., Courey, S. J., & Hamlett, C. L. (2004). Expanding schema-based transfer instruction to help third graders solve real-life mathematical problems. *American Educational Research Journal*, *41*(2), 419-445.
- Fuchs, L. S., Fuchs, D., Prentice, K., Hamlett, C. L., Finelli, R., & Courey, S. J. (2004). Enhancing mathematical problem solving among third-grade pupils with schema-based instruction. *Journal of Educational Psychology*, *96*(4), 635.
- Golding, J. (2012). '"You weren't expected to be creative": policy-practice tensions in GCSE mathematics'. *Informal Proceedings of the British Society for Learning Mathematics,* 32 (1) Available at www.bsrlm.org.uk
- Goulding, M., Rowland, T. and Barber, P. (2002). 'Does it matter? Primary teacher trainees' subject knowledge of mathematics'. *British Education Research Journal*, 28 (5), 689-704.
- Gooding S. (2009). Children's difficulties with mathematical word problems. In Joubert M. (Ed) Proceedings of the British Society for Research into Learning Mathematics 29(3).

- Gray, E. M. & Tall, D. O. (1994) Duality, ambiguity, and flexibility: A "proceptual" view of simple arithmetic. *Journal for Research in Mathematics Education,* 25, 116-140.
- Greer, B. (1992) Multiplication and division as models of situations. IN GROUWS, D. A. (Ed.) *Handbook of Research on Mathematics Teaching and Learning.* New York, Macmillan.
- Groth, R. (2007). Understanding teachers resistance to the curricular inclusion of alternative algorithms. Mathematics Education Research Journal, 19, (1), 3-28.
- Hamilton, L.S. and Koretz, D.M. (2002). Tests and their use in test-based accountability systems. In Hamilton, L.S., Stecher, B.M., Klein, S.P. (Eds.). *Making sense of Test-based Accountability in Education*. Santa Monica: Rand cooperation. http://www.rand.org/pubs/monograph_reports/MR1554/
- Harris, J. (2012). *Rational Numbers: Investigating compulsion for mathematics study to 18*. London: The Pearson Think Tank.
- Hegarty, M., & Kozhevnikov, M. (1999). Types of visual-spatial representations and
- mathematical problem solving. *Journal of Educational Psychology, 91*, 684-689
- Her Majesty's Inspectorate for Education and Training in Wales.(2010). *Improving numeracy in key stage 2 and key stage 3,* Cardiff: Estyn.
- Her Majesty's Inspectorate of Education. (2010a). *Learning Together: Mathematics*, Livingston: HMIE.
- Hiebert, J., & Carpenter, T. (1992). Learning and teaching with understanding. In D. Grouws (Ed.), Handbook of research on mathematics teaching and learning (pp. 65-97). New York: Simon & Schuster Macmillan.
- Hill, H. C., Blunk, M. L., Charalambous, C. Y., Lewis, J. M., Phelps, G. C., Sleep, L. & Ball, D. L. (2008) Mathematical knowledge for teaching and the mathematical quality of instruction: An exploratory study. *Cognition and Instruction,* 26, 430-511.
- Hill, H. C., Ball, D. L. & Schilling, S. G. (2008) Unpacking pedagogical content knowledge: Conceptualizing and measuring teachers' topic-specific knowledge of pupils. *Journal for research in mathematics education,* 372-400.
- Hill, H. C., Rowan, B. & Ball, D. L. (2005) Effects of teachers' mathematical knowledge for teaching on student achievement. *American educational research journal,* 42, 371-406.
- Hill, H. C., Schilling, S. G. & Ball, D. L. (2004) Developing measures of teachers' mathematics knowledge for teaching. *The Elementary School Journal,* 105, 11-30.
- Holcombe, R.W., Jennings, J.L. and Koretz, D. (2012). The Roots of Score Inflation: An Examination of Opportunities in Two State's Tests. To appear in G. Sunderman (Ed.), *Charting Reform, Achieving Equity in a Diverse Nation*. Greenwich, CT: Information Age Publishing.
- Holcombe, R.W. (2011). *Can the 10th Grade Mathematics MCAS Test be Gamed? Opportunities for Score Inflation.* Unpublished manuscript. Harvard University, Cambridge, Massachusetts.
- Jennings, J. L., & Bearak, J. M. (2014). "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educational Researcher*, *43*(8), 381-389.
- Jitendra, A. K., & Hoff, K. (1996). The effects of schema-based instruction on the mathematical word-problem-solving performance of pupils with learning disabilities. *Journal of Learning Disabilities*, *29*(4), 422-431.
- Jitendra, A. K., Star, J. R., Starosta, K., Leh, J. M., Sood, S., Caskie, G., ... & Mack, T. R. (2009). Improving seventh grade pupils' learning of ratio and proportion: The role of schema-based instruction. *Contemporary Educational Psychology*, *34*(3), 250-264.
- Jitendra, A. K., Star, J. R., Rodriguez, M., Lindell, M., & Someki, F. (2011). Improving pupils' proportional thinking using schema-based instruction. *Learning and Instruction*, *21*(6), 731-745.
- Jitendra, A. K., Star, J. R., Dupuis, D. N., & Rodriguez, M. C. (2013). Effectiveness of schema-based instruction for improving seventh-grade pupils' proportional reasoning: A randomized experiment. *Journal of Research on Educational Effectiveness*, *6*(2), 114-136.

- Lesh, R. & Zawojewski, J. (2007) Problem solving and modeling. In Lester, F. K. (Ed.) *Second handbook of research on mathematics teaching and learning.* Charlotte, NC, Information Age Publishing.
- Lester F., Garofalo J. and Kroll D. (1989). The role of metacognition in mathematical problem solving in Schoenfeld 1992.
- Mansell, W., and M. James. "Assessment Reform Group (2009) Assessment in schools. Fit for
- purpose? A Commentary by the Teaching and Learning Research Programme. London, Economic and Social Research Council." *Teaching and Learning Research Programme*.
- Marshall, S. P. (1995). Schemas in problem solving. New York: Cambridge University Press. Sweller, J., Chandler, P., Tierney, P., Cooper, M. (1990). Cognitive load as a factor in the structuring of technical material. Journal *of Experimental Psychology: General, 119*, 176- 192.
- McClure L. (2013) Problem solving and the new curriculum. Nrich.maths.org.
- National Audit Office. (2008). *Department for Children, Schools and Families - Mathematics*
- *performance in primary schools: Getting the best results*, London: HMSO.
- National Centre for Excellence in the Teaching of Mathematics. (2010). *Developing mathematics in Primary Schools: Headteachers talk about creating Excellence in mathematics,* NCETM. Available via www.ncetm.org.uk accessed 22 June 2012.
- National Council of Teachers (2006) Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for coherence. National.
- National Research Council. (2011*). Successful K-12 STEM Education: Identifying Effective Approaches in Science, Technology, Engineering and Mathematics. Committee on Highly Successful Science Programs for K-12 Science Education. Board on Science Education and Board of Testing Assessment, Division of Behavioural and Social Sciences and Education,* Washington DC: The National Academies Press.
- National Council of Teachers of Mathematics. (1989). Curriculum and evaluation standards for school mathematics. Reston, VA: Author.
- National Council of Teachers of Mathematics. (2000). Principles and standards for school mathematics. Reston, VA: Author.
- Nickson M. (1998). What is the difference between a pizza and a relay race? The role of assessing KS2 Mathematics. British Journal of Curriculum and Assessment 7 (3) 19-23.
- Norris, E. (2012). *Solving the maths problem: international perspectives on mathematics education*, London: RSA.
- Noyes, A., Drake, P., Wake, G. and Murphy, R. (2011). *Evaluating Mathematics Pathways, Final Report*, *RR 143*. London: DfE.
- Nunes, T. & Bryant, P. (1996) *Children doing mathematics*, Wiley-Blackwell.
- Ocean, J. (2009). *The "Top Ten": ten (now eight) important factors for coaches of mathematics in elementary and high schools*, Paper presented at the British Educational Research Association Annual Conference, University of Manchester, 2-5 September 2009. Available at http://www.leeds.ac.uk/educol/documents/187705.doc
- Office for Standards in Education, Children's Services and Skills. (2012) Mathematics: Made to Measure.
- Office for Standards in Education, Children's Services and Skills. (2010a). *Finnish pupils' success in mathematics*. London: Ofsted.
- Office for Standards in Education, Children's Services and Skills. (2011d). *Good practice in Primary mathematics: evidence from 20 successful schools*, London: Ofsted.
- Office for Standards in Education, Children's Services and Skills. (2012b). *Good practice resource – Engaging able mathematics pupils: King Edward VI Camp Hill School for Boys,* Manchester: Ofsted.
- Ontario Ministry of Education. (n.d.). *Research Synopsis Connecting Practice and Research in Mathematics.* Accessed via www.edu.gov.on.ca/eng/research/projects 27 November 2012.
- Perry, C. (2012). *Science, Technology, Engineering and Mathematics (STEM). Research and Information Service Briefing Paper,* Belfast: Northern Ireland Assembly.

- Phelps, G. (2009) Just knowing how to read isn't enough! Assessing knowledge for teaching reading. *Educational Assessment, Evaluation and Accountability*, 21, 137-154.
- Philipp, R. A. (1996). Multicultural mathematics and alternative algorithms. Teaching Children Mathematics, 3(3), 128-134.
- PÓLYA, G. (1990) *How to Solve It,* London:, Penguin. (Originally published in 1945).
- Pollitt A. and Ahmed A. (2000). Comprehensive failures in educational assessment. Paper presented at the European Conference on Educational Research 2000. http://www.cambridgeassessment.org.uk/Images/109671-comprehension-failures-in-educational-assessment.pdf
- Randolph, T. D., & Sherman, H. J. (2001). Alternative algorithms: Increasing options, reducing errors. Teaching Children Mathematics, 7, (8), 480-484.
- Royal Society. (2007). *The UK's science and mathematics teaching workforce. A State of the Nation report,* London: Royal Society.
- Royal Society. (2010). *Science and mathematics education 5-14, A State of the Nation report*. London: Royal Society.
- Schoenfeld, A. H. (1992) Learning to think mathematically: Problem solving, metacognition, and sense making in mathematics. IN GROUWS, D. (Ed.) *Handbook of research on mathematics teaching and learning.* New York, McMillan.
- Silver, E. A., Mesa, V.M., Morris, K.A., Star, J. R. and Benken, B.M. (2009*).* 'Teaching Mathematics for Understanding: the analysis of lessons submitted by Teachers Seeking NBPTS Certification'. *American Research Journal*, 46 (2), 501- 531.
- Siraj-Blatchford, I., Shepherd, D-L., Melhuish, E., Taggart, B., Sammons, P., Sylva, K. (2011). *Effective Primary Pedagogical Strategies in English and Mathematics in Key Stage 2: A study of Year 5 classroom practice drawn from the EPPSE 3-16 longitudinal study*, London: DfE.
- SKEMP, R. R. (1987) *The psychology of learning mathematics,* Hillsdale, N.J. ; Hove, Lawrence Erlbaum.
- Standards and Testing Agency (2015). Key stage 2 mathematics test framework. www.gov.uk/sta.
- Stanic G. and Kilpatrick J (1988) Historical perspectives on problem solving in the mathematics curriculum. In Charles R. and Silver E. (Eds) The teaching and assessment of mathematical problem solving. Reston VA, National Council of Teachers of Mathematics.
- Stigler J.W. and Hiebert J. The teaching gap. Free Press, Simon and Schuster
- Stecher, B. M. & Mitchell, K. J. (1995) Portfolio-Driven Reform: Vermont Teachers' Understanding of Mathematical Problem Solving and Related Changes in Classroom Practice.
- Swain, J., and Swan, M. (2007). *Thinking Through Mathematics research report*. http://dera.ioe.ac.uk/22296/1/doc_3631.pdf
- Turner, J. C., Warzon, K. B. and Christenson, (2011). 'Motivating Mathematics Learning: Changes in Teachers' Practices and Beliefs during a nine month collaboration', *American Education Research Journal* 48: 718-762.
- Truss, E. (2013). *North of England Education Conference*. London: DfE.
- Verschaffel L, Greer B, and De Corte E (2000). Making sense of word problems. Swets and Zeitlinger.
- Wilburne, J. M., Marinak, B. A., & Strickland, M. J. (2011). Addressing Cultural Bias--Reflect and Discuss. *Mathematics Teaching in the Middle School*, *16*(8), 460-465.
- Wiliam, D. and Bartholomew, H. (2004). 'It's not which school you're in but which set you're in that matters: the influence of ability grouping practices on student progress in mathematics'. *British Education Research Journal,* 30(2), 279-293.
- Wiliam, D. and Thoreson, O. (2011*). Excellence in Mathematics, Report from the Mathematics Excellence Group,* Edinburgh: The Scottish Government.
- Williams, P. (2008) Independent Review of Mathematics Teaching in Early Years Settings and Primary Schools. DCSF.

- Woodward, J., Beckmann, S., Driscoll, M., Franke, M., Herzig, P., Jitendra, A., Koedinger, K. R. & Ogbuehi, P. (2012) Improving Mathematical Problem Solving in Grades 4 through 8. IES Practice Guide. NCEE 2012-4055. *What Works Clearinghouse*.
- Xin, Y. P. (2008). The effect of schema-based instruction in solving mathematics word problems: An emphasis on prealgebraic conceptualization of multiplicative relations. *Journal for Research in Mathematics Education*, 526-551.

Appendix 1. Interview guidelines

Interview guideline first interview (March/April 2015)
The nature, prevalence and effectiveness of key stage 2 test preparation strategies in mathematics

Interview Schedule

Name of Teacher:

Name of School:

Date:

Thank you for taking part in our study – we are really grateful to you for giving up your time to talk to us.

I wonder if you could start by giving us some idea of your background in teaching and your experience. Can you tell me about that?

Prompt for
Time spent teaching different year groups? Time spent in Year 6?
Any mathematics specialism: at A-level? in degree? in ITE? Later CPD?

I would first, very briefly, like to get a sense of how you plan your teaching in mathematics focusing on how you plan your daily/weekly lessons and how you plan the order of topics.

Prompt for resourcing
Is there a particular curriculum framework that you follow?
What is that? Who chose it?
Relationship to Old NC or New NC?
Are there particular textbooks and other instruction materials that you use?
Who chose these? Can you use others if you prefer?

Prompt for patterns of work
How does the teaching of the mathematics topics vary across the year? Is there much variation between terms?
How do you make decisions about how to distribute the curriculum across the year?
If you feel you want to vary that – say you feel the children need to spend more time on a topic than the scheme allows, what do you do?
How do you go about making your weekly/daily plans?
Would it be possible to see examples of your long/medium/short term planning?

Prompt for class and individual levels of planning
Do you plan specifically for different groups and/or individuals?
What groups do you think about when you plan?
How often do you undertake this level of planning?
How do you target pupils for extra support in mathematics?
Which pupils are targeted?
How do you choose instructional strategies or decide on the specific choice of learning settings in the classroom for these groups of pupils

How do you decide which pupils receive small group instruction, pull out/push in programs, individualized instruction etc.?
Is this within the mathematics lesson/classroom, or outside?

Prompt for target setting
Do you use target setting?
How do you set targets? Are these individual or group targets?
Who decides what kinds of targets to use?
How are targets informed?
How do you monitor progress?

Prompt for ongoing asssessment
How do you track performance of pupils?
Do you use formative assessment? Can you give me an example?
Do you use short end of topic type tests?
Do you link these assessments to NC levels in any way? How?

I'd like to spend some time with you thinking about the way in which you teach pupils multiplication and division. Can you talk to me about that?

Prompt for multiplication
How do you teach multiplication in Year 6?
Are extended written methods used?
Are children taught to use the grid method? If so, when? Or Why not?
Do children use the standard algorithm for long multiplication? If so, when? Or Why not?
Is there a transition period? How do you manage the transition between the two? Has this changed in the last two years? How?
What do you know about how marks are awarded for working out in multiplication?
What has been the effect of recent changes in awarding marks for working?

Prompts for division
How do you teach division in Year 6?
Are extended written methods used?
Are children taught chunking? If so, when? Or Why not?
Do children use the standard algorithm for long division? If so, when? Or Why not?
Is there a transition period? How do you manage the transition between the two? Has this changed in the last two years? How?
What do you know about how marks are awarded for working out in division?
What has been the effect of recent changes in awarding marks for working?

Has the key stage 2 test had any impact on how you teach multiplication and division?

I'd like to spend some time with you thinking about the KS2 math tests themselves. How would you describe the content and the style of questions that are tested in the key stage 2 math tests?

Prompt for test content
How would you describe the three test papers, in terms of content and structure?
Are there strands/sub-strands that are tested each year, or topics that are never tested?
How have you learned about the content of the current mathematics tests?

Have these topics changed over the last years; e.g. in how frequently they are tested, how they are tested (types of items) or where they show up on the test?
If not mentioned, how about the following changes:
Removal of use of calculator in Paper B?
Requirement for children to use standard algorithm to get marks for working?
Reintroduction of Level 6 paper?

Prompt for similarity of items, and how frequently or consistently these show up on the test
Are there types of items that always show up on the test?
If yes, how would you describe these items?
For each of the types of items: do they show up on the test every year, and/or would there be one or more of those items every year?
Have these items changed over the last years; e.g. in when and how frequent they are tested or where the items show up on the test?

Sorting exercise
[If this is the first year of teaching in Y6, but they have previously taught Y2-Y5 and used KS1 SATs or Optional NC Tests, draw on that experience.]

Note: hand out appendix 1 at the start of the sorting exercise

Please look at the following test papers (2014 L3-5 Paper 1 and Paper 2) and give an example of the types of similar items you just described. Please look through each test paper and mark the items:
EY: this item shows up on the test (almost) every year / high recurrence over years
HW: there are many of these items on the test every year (high weight in points)
R: items that don’t stand out/are not visually similar and seem random

Are there any other items that you recognize/that you have seen on key stage 2 mathematics tests across other years that you didn’t mention before? (Please also mark those with EY, HW or R).

Please state the question number and describe what you see for each item.

NOTE DOWN CLUSTERS:

ASK FOR EACH CLUSTER

a. Prompt for clusters of visually similar items
How are these items similar?
Are there meaningful clusters of items on this test/previous tests?
What would they look like? What do these clusters mean to you?

How have you learned about these clusters?

b. Prompts for pupil performance on visually similar items
Which pupils do well with this cluster of items?
Which pupils do not do so well with this cluster of items?
How do you factor that into you teaching?

c. Prompts for aligning teaching to clusters of visually similar items
Does your knowledge about these visually similar cluster of items affect your teaching in any way? If so, how? If not, why not?
Are these items significant in relation to how you teach and prepare children for the tests? If so, how? (grouping of pupils, formative assessments, scheduling of content, specific instructional strategies)
Are the visually similar items easier for pupils to answer compared to the random ones? Why? Why not?

Prompts for other relevant patterns
Are there other relevant patterns on the test; e.g. in ordering of items within each paper? Or in what is tested in either paper A or B?
Does this knowledge affect your teaching? If yes, how? If not, why not?

Describe for me the more general ways in which you prepare children for the KS2 mathematics tests. Please note that we are asking questions specifically about the mathematics test and not about the KS2 test in general.

Prompts for preparation and planning
Are there particular books you use? Or do you use past papers?
If you use past papers, how do you use them?
Do you analyse them or break them up at all? How?
Does this analysis feed back into your other mathematics teaching at all?
How does planning for the KS2 math test preparation fit with your other planning activities?

Prompt for patterns of work
When do you start KS2 test preparation?
Does the way you teach and engage with children change at this time?
In what ways? What do you think the children notice?

Prompt for ways of working
Is the way you plan these lessons the same or different to the way you plan at other times?
How do you choose how to approach particular topics?
Do you target particular pupils or groups of pupils for extra teaching? How do you decide which pupils to target? What extra teaching do you typically provide?
If a child is struggling to learn something, how do you work with them? Is there support available for you in thinking about specific difficulties you may encounter?

Prompt for test data use
How are student outcomes/ achievement results of the key stage 2 test analyzed: collectively or individually?
What do you/does the school look for?
How has this informed your teaching?
Does your school explore patterns of performance on the mathematics tests? And feed back to you?
Does your school examine how individual children performed on individual questions?

Does your school modify teaching and test preparation strategies in response to how previous classes performed?

Prompt for parental support
Do parents in this school take any measures to prepare their children for the key stage 2 test in mathematics? What are those? How do they affect your teaching?

The tests are changing for 2016. Have you heard anything about this? What have you heard?

Prompts
Potential changes in the content? format? levels being tested? L6 paper? performance descriptors?
How have you learned about the content of the current and future test?
Has this knowledge started to affect your planning and delivery of your teaching? If yes, how and why? If not, why not?

Prompts for use if they mention the 2016 Mathematics Test Framework? (if not – do not pursue)
How would you describe the different elements of the framework (cognitive domains etc.)?
Are they helpful to you in planning and delivering your teaching?
If yes, how and why? If not, why not?
Has this knowledge affected your planning and delivery of your teaching?
If yes, how and why? If not, why not?

Thank you for taking part in our study – we are really grateful to you for giving up your time to talk to us. Is there anything else that you would like to tell me that we have not covered in the interview?

Enquire about colleagues/friends (esp. for < floor targets)

[ENDS]

Hand out sorting exercise

Please look at the following test papers (2014 L3-5 Paper 1 and Paper 2) and give an example of the types of similar items you just described. Please look through each test paper and mark the items:

EY: this item shows up on the test (almost) every year / high recurrence over years
HW: there are many of these items on the test every year (high weight in points)
R: items that don't stand out/are not visually similar and seem random

Are there any other items that you recognize/that you have seen on key stage 2 mathematics tests across other years that you didn't mention before? (Please also mark those with EY, HW or R).

Please state the question number and describe what you see for each item.

Interview guideline second interview (June 2016)
Name of Teacher:

Name of School:

Date:

Thank you for taking part in our study – we are really grateful to you for giving up your time to talk to us.

Part A – new participants only

1. I wonder if you could start by giving us some idea of your background in teaching and your experience. Can you tell me about that? [10 minutes]

Prompt for
Time spent teaching different year groups? Time spent in Year 6?
Any maths specialism: at A-level? in degree? in ITE? Later CPD?

Part B – all participants

2. How has your teaching of mathematics changed this year, since the new National Curriculum came into force for Year 6 in September? [10 minutes]

General Prompts
Could you say a bit more about that?
Can you explain that in a bit more detail?

Prompt for mastery
What do you understand by the term 'mastery'?
Has the concept of mastery influenced the way you teach this year?

Prompt for resources
Is there a particular curriculum framework that you follow?
Are there particular textbooks?
Any other instruction materials that you use?

Prompt for patterns of work
How has the teaching of the maths topics varied across the year?
Was there much variation between terms?
How do you go about making your weekly/daily plans?

Prompt for class and individual levels of planning
Has the way in which you (or the school) plan specifically for different groups and/or individuals changed?
How has the way in which you target pupils for extra support in mathematics changed?

Prompt for ongoing assessment
How have you tracked the performance of pupils this year?

Can you tell me anything about the new Teacher Assessment Framework for KS2 mathematics and how this has affected the way you teach this year?

Prompt for target setting
Have there been any changes to the way in which you (or the school) use target setting? If yes, how and why?

3. I'd like to spend some time with you thinking specifically about multiplication. First I'd like to do a little exercise to get you thinking about multiplication. Can you sort the following cards into groups and explain how you have sorted them?

Can you sort them in a different way? And tell me how you have done it this time? Repeat as time allows. [10 minutes]

(The following scenarios to be presented as a set of images. Notes in brackets indicate differences - for researcher reference only. Researcher to photograph results of sorting activity)

Several pairs of boots (meta-entity)
The fingers on several hands (fixed ratio)
Placing five paintbrushes in each of several pots (allocation - grouping)
Dealing cards to people in turn (allocation - sharing)
A shop label showing the cost per kilo (fixed rate)
An array with m rows and n columns (array)
A beaker containing three times as much liquid as another beaker (comparison)
The Grid Method for multiplication
The traditional algorithm

4. Here are some word problems which involve multiplication. How would you represent these problems to help pupils to solve them? What method would you expect children to use to find the answer? [10 minutes]

(The following problems to be presented as a set of cards, with a space to show working - set out in the same format as a SATs question. Up to six problems to be explored, as time permits.)

If 3.5kg of sugar is required for 5 kg of strawberries, how much sugar is required for 8kg of strawberries?
What is the maximum number of possible outfits from 3 pairs of trousers and 4 blouses?
If elastic can be stretched by 2.5 times, how far can you stretch a 3.5m length?
If three children each have 2.4l of orange juice, how much altogether?
Calculate the area of a rectangle 3.3m x 4.2m.
If 3/5 pupils pass an exam, and if there are 80 pupils, how many pass?

5. I'd like to spend some time with you thinking about the way in which you teach pupils multiplication. Can you talk to me about how you have taught multiplication in Year 6 over the last year? [10 minutes]

Prompts for methods
Are extended written methods used?
Are children taught to use the grid method? If so, when? Or Why not?
Do children use the standard algorithm for long multiplication? If so, when? Or Why not?
Is there a transition period? How do you manage the transition between the two? Has this changed in the last two years? How?

What visual images do you use in your teaching?
What concrete materials do you use in your teaching?
Are text books used?

Prompts for influences
What are the school's expectations about how you should teach multiplication?
Is there a Calculation Policy? How helpful do you find this?
What do you think the National Curriculum say about how to teach multiplication?
Is there Local Authority/Academy Trust guidance?
What about the influence of test books?
What training have you received?
ITT
INSET/Courses
Policy Review
Planning meetings
Peer observation
Team teaching
What do you remember about being taught multiplication when you were at school?

Prompts for recent changes?
Have the changes to the NC had any impact on how you teach multiplication?
Have the changes to the key stage 2 SATs had any impact on how you teach multiplication?

6. I'd like you to talk now about the changes to the KS2 tests [10 minutes].

Can you describe changes you have seen to the recent key stage 2 test?

How and when did you learn about these changes?

Have they affected your teaching in any way? If so, how?

Which materials/strategies are you using to teach pupils in relation to these changes?

How do you expect pupils to perform on these changed aspects of the test? Can you explain why you expect pupils to be struggling or performing well? Do you think performance of pupils will improve over the next years (as pupils and teachers become more familiar with the new curriculum and test)? If so why?

7. I'd like to talk now about changes in relation to 43 commonly recurring types of SATs question which teachers identified during the interviews we conducted last year. We have called these commonly recurring types of SATs question 'clusters'. [10 minutes]

Present list of clusters

Did the recent test paper have questions in the 43 clusters?
[ ] no [ ] a few [ ] many

Which of these 43 clusters (and the new ones) featured highly in your teaching this year?

Do your pupils have particular difficulty answering items in any of the 43 clusters? How do you address this?

Did the test include new types of questions? How would you describe these?

8. I'd like us to take look at this year's KS2 Maths test.

Present copies of 2016 SATs Papers

Can you look at the items on the recent test paper and tell us in which of the 43 cluster(s) they fit (more than one option allowed)? [10 minutes]

9. Thank you for taking part in our study – we are really grateful to you for giving up your time to talk to us. Is there anything else that you would like to tell me that we have not covered in the interview? [10 minutes]

Appendix 2. Sample of teachers round 1 and 2

Postcode	**FSM**	**Progress: % making expected progress in maths in 2013**	**No of Y6 classes**	**Y6 teacher round 1**	**Y6 teacher round 2**
Above floor target					
1	1.9	100	1	T1	T1
2	5.4	100	1	T2	T31
3	7.7	98	6	T3, T4	T32, T4
4	10.6	83	0.5	T5	T5, T33
5	9.5	92	3	T6	T34
6	10.5	97	1	T7	T7
7	4.3	100	2	T8	T8
8	5.1	100	1.5	T9	T35
9	16.1	100	2	T10, T11	T36, T11
10	33	87	1	T12	T37
11	33.2	90	1	T13	T13
12	6.1	84	3	T14	T38
13	10.5	92	1	T15	T15
14	59.6	91	1.5	T16	T39
Below floor target					
1	38.3	65	1	T17, T30	T17
2	53.4	73	1	T18	T18
3	27.4	78	1	T19	T19
4	37.9	58	3	T20	T20, T40
5	49.3	79	2	T21, T22, T23	T21, T22, T23
6	47.6	72	1	T24	T24
7	40.7	64	1	T25	T25
8	44	80	1.5	T26, T27	T41, T27
9	34.2	60	1	T28	T28
10	33.5	64	1	T29	N/A

Appendix 3. Expert analysis of key stage 2 mathematics test 2010-2017

Total number of items per strand on 2010-2017 KS2 Mathematics test

Strand	**2010 total number of items**	**2011 total number of items**	**2012 total number of items**	**2013 total number of items**	**2014 total number of items**	**2015 total number of items**	**2016 total number of items**	**2017 total number of items**
N=Number	6	5	3	8	2	6	10	7
C=Calculation	25	19	20	22	28	27	30	28
F=Fractions	7	9	9	9	12	10	20	22
R=Ratio	1	3	2	1	3	3	4	7
A=Algebra	1	4	3	2	2	0	4	4
M=Measurement	20	17	21	19	13	14	8	8
G=Geometry	9	11	14	5	7	10	6	5
P=Position	3	2	0	2	3	2	2	2
S=Statistics	11	15	11	11	12	13	4	4
U=Unclassified	1	1	1	3	1	1	0	0

<u>Number of items for each sub-strand (old curriculum tests 2010-2015)</u>

	2010			**2011**			**2012**			**2013**			**2014**			**2015**		
Sub-strand	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**
N1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
N2	0	0	0	0	0	2	1	0	0	1	0	1	0	0	1	0	0	1
N3	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	1
N4	1	1	1	0	1	0	1	0	0	1	1	0	0	0	0	1	0	0
N5	2	1	0	0	1	0	0	0	1	0	2	1	0	0	0	1	1	0
N6	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
C1	0	0	3	0	0	2	0	0	3	2	0	1	0	0	3	1	0	2
C2	1	0	0	0	0	0	2	0	0	0	0	0	0	0	0	1	0	0
C3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
C4	2	1	1	2	1	0	1	2	0	0	3	0	2	9	0	6	1	2
C5	0	1	1	0	0	0	0	0	1	0	1	0	0	0	2	0	1	0
C6	0	0	4	0	0	3	0	0	3	2	0	4	1	0	2	0	0	3
C7	1	1	1	1	0	1	2	0	1	1	1	2	1	0	0	0	1	0
C8	2	3	1	3	4	1	0	2	2	1	3	1	4	4	0	3	6	0
C9	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1
F1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0	2	0
F2	0	0	1	0	0	0	0	0	0	1	0	0	2	0	1	0	0	0
F3	1	1	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
F4	1	0	0	0	0	0	0	0	1	0	0	0	2	0	0	0	0	0
F5	0	1	0	0	0	1	0	0	1	0	0	2	0	0	1	0	0	2
F6	0	0	1	0	0	0	0	0	0	2	0	0	0	1	0	0	0	0
F7	0	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1
F8	0	0	0	0	0	0	0	2	0	0	1	0	1	1	0	0	0	0
F9	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0
F10	0	0	0	3	1	2	1	0	0	0	2	1	0	0	1	0	2	2

	2010			2011			2012			2013			2014			2015		
Sub-strand	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**
F11	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
F12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
R1	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	1	0	0
R2	1	0	0	0	1	1	0	0	1	1	0	0	1	1	0	1	0	0
R3	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
R4	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0
A1	0	0	0	0	0	0	2	0	0	0	0	0	1	0	0	0	0	0
A2	0	0	0	0	0	1	0	0	0	0	0	1	0	0	1	0	0	0
A3	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
A4	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
A5	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	0
M1	1	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0	0
M2	0	0	0	0	0	0	0	1	0	0	2	0	0	2	0	1	0	0
M3	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
M4	1	3	1	2	1	0	1	2	1	0	2	1	2	0	1	2	0	0
M5	0	0	0	0	1	2	0	1	1	1	1	0	0	0	1	0	0	3
M6	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
M7	1	0	0	2	0	0	0	1	0	1	1	0	0	0	0	0	1	0
M8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
M9	6	4	2	3	4	1	5	5	2	3	3	2	3	1	3	1	5	1
G1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
G2	3	2	1	2	3	2	4	5	0	2	0	0	1	2	1	4	0	1
G3	0	0	0	3	1	0	2	1	0	1	0	0	0	1	0	1	2	0
G4	1	1	1	0	1	0	0	1	1	0	1	1	1	1	0	1	1	0
G5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

	2010			2011			2012			2013			2014			2015		
Sub-strand	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**	**L3-L5 Paper 1**	**L3-L5 Paper 2**	**L3-L5 Paper 3**
P2	1	1	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0
P3	0	1	0	2	0	0	0	0	0	1	0	0	0	2	0	0	1	0
S1	3	1	0	1	2	0	3	3	0	4	2	0	0	6	0	2	3	0
S2	2	5	0	6	6	0	2	3	0	1	4	0	5	1	0	4	4	0
S3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Unclassified																	1	0

Number of items for each sub-strand (new curriculum tests: 2016-2017)

	2016			2017		
Substrand	**Paper 1**	**Paper 2**	**Paper 3**	**Paper 1**	**Paper 2**	**Paper 3**
N1	0	0	0	0	0	0
N2	1	5	0	2	1	0
N3	0	0	0	0	0	0
N4	0	0	1	0	1	0
N5	0	0	2	0	0	2
N6	0	0	1	0	0	1
C1	2	0	1	1	0	0
C2	4	1	0	3	0	0
C3	0	0	0	0	0	0
C4	0	0	0	1	2	0
C5	0	2	0	0	0	2
C6	4	0	0	3	2	1
C7	9	0	1	7	0	0
C8	0	2	2	0	1	4
C9	2	0	0	1	0	0
F1	0	0	0	0	0	0
F2	0	3	0	0	1	0
F3	0	0	0	0	0	0
F4	4	0	0	6	1	2
F5	2	0	0	4	0	0
F6	0	0	0	0	0	0
F7	0	0	0	0	0	0
F8	4	0	1	3	1	0
F9	2	0	0	2	0	0
F10	0	1	3	0	0	0
F11	0	0	0	0	1	1
F12	0	0	0	0	0	0
R1	0	1	0	0	0	1
R2	2	0	0	3	1	0
R3	0	0	0	0	1	0
R4	0	0	1	0	1	0
A1	0	0	0	0	0	1
A2	0	2	2	0	2	0
A3	0	0	0	0	0	0
A4	0	0	0	0	0	1
A5	0	0	0	0	0	0
M1	0	0	0	0	0	0
M2	0	0	0	0	0	0
M3	0	0	0	0	1	0
M4	0	0	1	0	1	0

	2016			2017		
Substrand	**Paper 1**	**Paper 2**	**Paper 3**	**Paper 1**	**Paper 2**	**Paper 3**
M5	0	1	2	0	0	1
M6	0	0	0	0	0	0
M7	0	0	1	0	0	1
M8	0	0	1	0	0	1
M9	0	2	0	0	2	1
G1	0	0	0	0	0	0
G2	0	1	1	0	1	3
G3	0	0	0	0	0	0
G4	0	2	2	0	1	0
G5	0	0	0	0	0	1
P1	0	0	0	0	0	0
P2	0	1	1	0	0	1
P3	0	0	0	0	1	0
S1	0	2	2	0	0	2
S2	0	0	0	0	2	0
S3	0	0	0	0	0	0

Total number of items per cognitive domain 2010-2017 KS2 Mathematics test

Cognitive domain	**2010 total number of**	**2011 total number of**	**2012 total number of**	**2013 total number of**	**2014 total number of**	**2015 total number of**	**2016 total number of**	**2017 total number of**
Depth of Understanding								
1=Recall of Facts or Procedures	9	4	12	8	9	10	37	39
2=Use to Solve Simple Problems	38	45	40	42	39	47	28	22
3=Use to Solve Complex Problems	32	33	28	28	31	25	14	16
4=Understand and Use Creatively	5	4	4	3	4	3	9	10
Computational Complexity								
1=No Numeric Steps	18	11	20	9	13	9	6	6
2=One Numeric Step (or small number)	47	50	49	50	49	50	40	45
3=Larger Number of Simple Numeric Steps	16	23	14	19	19	23	26	31
4=Larger Number of More Complex Steps	3	2	1	3	2	3	16	5
Spatial Reasoning								
1=No Spatial Reasoning	70	70	66	43	70	73	78	77
2=Manipulation Required	8	10	12	23	9	5	5	8
3=Complex Manipulation	6	6	6	12	4	7	4	0
4=Interpret, Infer or Generate	0	0	0	3	0	0	1	2
Data Interpretation								
1=No Data Interpretation	66	64	72	73	65	65	82	74
2=Select and Retrieve	14	13	9	6	9	17	6	9
3=Select and Interpret	2	8	3	2	7	3	0	3

4=Generate or Infer	2	1	0	0	1	0	0	1
Response Strategy								
1=Select or Construct Response	43	41	36	52	26	66	69	71
2=Construct Set of Responses	30	35	37	20	45	18	18	14
3=Construct Straightforward Explanation	11	8	9	7	11	1	0	2
4=Construct Complex Explanation	0	2	2	2	1	0	1	0

<u>Number of items according to format, 2010-2017 KS2 Mathematics test</u>

Format of Question & Response	**2010: total number of items**	**2011: total number of items**	**2012: total number of items**	**2013: total number of items**	**2014: total number of items**	**2015: total number of items**	**2016: total number of items**	**2017: total number of items**
MC=Multiple Choice	1	1	2	4	4	1	4	5
M=Matching	0	0	1	0	3	1	0	1
TF=True/False	3	2	3	4	1	1	0	1
C=Constrained	75	81	75	70	75	79	84	80
O=Open	5	2	3	3	0	4	0	0

Appendix 4. Clusters of similar test items

3. Word Problems (general): Word problems require children to read a problem and decide which operation(s) to carry out.
4. 2-step money problem: Word problems involving money, requiring children to do more than one calculation (e.g. addition then subtraction). This type of item would often concern children going shopping, purchasing more than one item and requiring change. Items within this cluster would usually receive two marks, and include a 'show your working' box.
5. Context-free Calculation (horizontal layout): Items requiring pupils to complete a context-free calculation, set out in horizontal format, typically involving three-digit number(s). The calculation may involve multiplication or division which would get pupils two marks for a correct answer, while answers need to be explained in a 'Show your working' box. Calculations may also involve addition or subtraction, which are more likely to be worth one mark only and have no designated working out space.
6. Multiplication Grid: Items requiring children to use their knowledge of times tables to complete a multiplication grid.
7. Missing Symbol: Items requiring children to add missing symbols to complete number sentences (usually empty circles rather than boxes).
8. Missing Digits (to complete calculation): Items displayed as a calculation or number sentence with one or more empty boxes for one or more digits, or perhaps to complete column subtraction.
9. Number Cards: Items presented using number cards, usually single digit numbers. Children are required to use some or all of the number cards to make a three digit number or add to make a given total. Can involve properties of numbers or a calculation.
10. Negative Numbers: Items requiring children to manipulate positive and negative numbers, often to find the difference between a positive and a negative number.
11. Function Machine: Items requiring children to carry out more than one operation on a number, perhaps requiring children to work backwards, possibly presented in the form of a diagram.
12. Missing Numbers (to complete number sentence): Items displayed as a number sentence with one or more empty boxes for one or more numbers. This type of item: 1. May require inverse operation. 2. May require two numbers to balance equation. 3. May require trial and error to find up to three numbers.
13. Number Sequence: Items requiring children to work out the rule for a number sequence (e.g. add 10) and then apply the rule to continue the sequence, perhaps working backwards or finding the Nth term. Could include decimals or fractions on a number line.
14. Algebra: Items including a letter representing a number, such as k, often bold and italic. Always towards the very end of the SATs Paper.
15. Ordering mixed fractions, decimals and/or percentages: This type of item includes a collection of four fractions and decimals (e.g. two fractions, one decimal and one percentage) which the children are required to order smallest to largest. The item is often placed at the end of the paper.
16. Ordering Decimals (Different number of decimal places): This type of item asks children to order numbers with different numbers of decimal places, typically a collection of four or five numbers with different numbers of decimal places which have to be ordered from smallest to largest.
17. Ordering General (other than decimals/fractions; e.g. time): Items asking children to order items (e.g. length of time), usually from smallest to largest
18. Ratio and Proportion (often cooking): Items requiring children to scale up or scale down, e.g. cooking for 12 people using a recipe for 4 people. The item sometimes requires two steps, such as when to divide to scale down and then multiply to scale up. Children have to show their working in a box and would receive two marks in total. The item often appears at the end of the test paper.
19. Percentage: Items requiring children to find one or more percentages of an amount, such as 10% and 25% of 200, asking children to compare different percentage values, or presenting percentages in a word problem.
20. Probability: Spinner, Marbles and Beads: Items about probability often depicting a spinner, where children have to consider the likelihood of the spinner landing in a particular way, Or marbles, beads, discs, or other objects, often placed inside a bag, and requiring children to consider the likelihood of a particular object being selected from the bag.

21. Explanation: Items including a statement from one of the three characters in the SATs paper, in inverted commas, which is true or false, and asks children (in a clouded shaped space) to explain why the statement is correct or incorrect.
22. Matching fractions and decimals: Items typically including four or five pairs of fractions and equivalent decimals arranged in two columns, requiring children to draw lines across the page to match the correct pairs. Usually 'The first one is done for you'. This type of item will generally appear at the beginning of the test paper.
23. Matching equivalent fractions (shaded diagram and numbers): Items involving equivalent fractions, e.g. 3/4 = 6/8. These fractions are often presented with a range of different shapes with different and/or equivalent fractions shaded, or where children have to understand how to calculate the common denominator to compare and find the equivalent fraction.
24. Matching shapes to names: This item typically includes four or five pairs of shapes and names of shapes arranged in two columns, requiring children to draw lines across the page to match the correct pairs. Usually 'The first one is done for you'. The item would appear nearer the beginning of the paper.
25. Length Problem: Items requiring children to solve problems involving length. Sometimes includes picture 'Not actual size'. Calculation needed sometimes involves more than one step and has a 'Show your working' box.
26. Measuring with a Ruler: Items requiring children to use a ruler to measure given lines or draw lines of given lengths. Lengths expressed in centimetres.
27. Reading a Scale: Items requiring children to read numbers from a scale, such as weight in kilograms or capacity in litres. Often requiring children to calculate intervals between smaller graduations that have not been labelled on the scale.
28. Conversion of Weights and Measures: Items where children have to convert units of measure (weight, length, capacity, and/or time), e.g. from millilitres to litres. Items can be expressed as word problems, or requiring comparisons of different units of measure.
29. Data handling: Line graph: Items requiring children to interpret data from a line graph where the horizontal axis usually refers to time, and the vertical axis is usually labelled in multiples. The item normally includes a series of two or three question, requiring pupils to work out smaller intervals on the axis to answer at least one question. Items about graphs tend to show up towards the end of the test paper.
30. Data handling: Bar chart: Items requiring children to interpret data from a bar chart. The bar chart can have a horizontal or vertical layout and is normally followed by a series of questions, often near the beginning of the paper. Questions often require children to find the difference between two bars to find 'how many more than'.
31. Data handling: Pie Chart: Items requiring children to interpret data from a pie chart divided into three or four sections. The interpretation requires them to calculate percentages and sometimes explain their answer. The item usually shows up towards the very end of the test.
32. Data handling: pictogram: Item including pictograms with complete and fractional images representing whole numbers, followed by a series of questions. Children are often required to find the difference between two values to find 'how many more than' or to add values for 'altogether'. This item is usually seen early in the test.
33. Carroll diagram: Items requiring children to complete a Carroll diagram.
34. Venn Diagram: Items requiring children to place numbers or other things in the correct positions in a Venn diagram. Shape/orientation of Venn diagrams varies from year to year. Often involves properties of numbers.
35. Reading a Table: Items requiring children to interpret information presented in a table. Often presented with a table at the top of the page with questions below. Often seen early in the paper.
36. Properties of Shapes: Items requiring children to know names of shapes and/or understand about properties of shapes, such as number of equal length sides or number of lines of symmetry.
37. Angles & Protractor: Items requiring children to: 1. identify types of angles (often nearer the beginning of the paper) 2. use angle rules such as sum of angles in a triangle, or 3. use a protractor (often towards the end of the paper).

38. Area and Perimeter: Items about area requiring children to find the area of a given shape, which is usually an irregular or composite shape. Children can find the answer by counting whole or part squares.
39. Coordinates: Items presenting children with a picture of a shape or a line in the first quadrant, where some points are labelled, but the axes are often ungraduated (not marked with divisions or units of measurement). Children are required to identify coordinates of a given point.
40. Rotation: Items requiring children to rotate a shape or pattern, often a pattern of circles or triangles on a square. Children need visualisation skills and/or tracing paper to rotate the pattern.
41. Symmetry and reflection: 1. Items requiring children to complete a shape reflecting in a mirror line, and 2. Items requiring children to identify lines of symmetry.
42. Translate shape: Items requiring children to move a shape and draw it in its new position, but without rotating or reflecting it.
43. Time Problem and timetable: Word problems about time, requiring children to find the difference between two times crossing the hour boundary, or items requiring children to interpret information presented in a time table.
44. Calendar: Items requiring children to use/understand a calendar.
45. Coins to make total: Items requiring children to use their knowledge of coins to make a given total.

Appendix 5. Teachers quotes to describe (instructional responses to) clusters of similar test items

Description of clusters

Quotes to describe the cluster ‘word problems’:

- ‘You get a lot of questions around cakes and food and decisions to make, there aren’t very many in terms of straightforward calculations that the children are asked to solve, it usually is through some sort of problem solving.’ (first round of interviews)
- ‘There’s usually some kind of word problem that involves multiplication and division and possible multiplication and addition, that kind of thing.’ (first round of interviews)
- “Just reading all of that, you can just see the children going ‘aah, where do I start?’ I think that’s a really tough question, I really do. Yeah, we are into level six definitely now with that. Just keeping track of all of that information, there are so many numbers there, and it’s money obviously coming in again, weights and measures, scaling, multiplication, division, it’s all in there, there’s a lot, it’s a laden question, it really is, so difficult.” (second round of interviews)

Quotes to describe the cluster ‘2-step money problem’

- ‘You get a lot of questions around cakes and food and decisions to make, there aren’t very many in terms of straightforward calculations that the children are asked to solve, it usually is through some sort of problem solving.’ (first round of interviews)
- ‘There’s usually some kind of word problem that involves multiplication and division and possible multiplication and addition, that kind of thing.’ (first round of interviews)
- ‘Number nineteen, again I think this, I think the problems in them this year have got more steps than before, I think last year we probably would have had a two step, whereas this is actually a fairly tricky three step question to do here, so I think that will be, so they have had problems like that before, general word problems probably would come under, but I think there’s an extra step in it these days.’ (first round of interviews)
- ‘It was the jars though that I think threw a lot of the children, the fact that they weren’t just using the strawberries and the sugar, they had to remember the jars, so it did become three steps.’ (first round of interviews)
- ‘Nineteen, so this is like a multistep word problem, like you’ve got your two step money problems there, thought this was hard because it’s a lot of writing, no pictures with it, and you also had to convert between, you had like a mixed unit, so you had your pounds, and then you had pence as well, so was quite a lot to do in this, so it was, there were lots of steps, and also the fact that it’s worth three marks.’ (first round of interviews)
- ‘Number nineteen, Miss Mills is making jam to sell at the school fair, strawberries, again this is a tough one, again a multistep problem, strawberries cost seven pounds fifty, sugar costs seventy nine p per kilogram and ten glass jars cost six pounds ninety, so we are multiplying here, we are dividing, multisteps, twelve kilograms of strawberries and ten kilograms of sugar to make twenty jars full of jam. Total cost to make twenty jars of jam. Just reading all of that, you can just see the children going aah, where do I start? I think that’s a really tough question, I really do. Yeah, we are into level six definitely now with that. Just keeping track of all of that information, there are so many numbers there, and it’s money obviously coming in again, weights and measures, scaling, multiplication, division, it’s all in there, there’s a lot, it’s a laden question, it really is, so difficult.’ (second round of interviews)
- ‘Number eleven, toyshop orders eleven boxes of marbles, each box contains six bags of marbles, each bag contains forty five marbles, how many..? OK, familiar type question, but it’s more complex in some way and I can’t think how. I think there’s an extra step to it. I think I might only expect to see two steps in that and there’s another one.’ (second round of interviews)
- ‘Number nine is almost a pictogram type question, number nine’s kind of a horrible…number nine is a question which is designed to confuse, I think, the children, it’s one of those nice ones which is doing a two step money problem, so it is in the clusters, but the way it’s been organised is perhaps

more complicated than it would have been in the past, that might have been instead of six pencils cost one sixty eight, three pencils and a rubber cost this, it might have been six pencils cost this, how much would four cost? Not how much would three? So having to halve it and then work out a difference, so actually it's mixing together, it's a two step money problem but it's adding more depth to it.' (second round of interviews)

Quotes to describe the cluster 'context-free calculation'

- 'There's always your bog standard addition and subtraction, just straight sums that always come up'. (first round of interviews)
- 'There's always a written calculation with no sort of context whatsoever, but I'll say that I've always found there to be a strong bias towards multiplication, for all we sort of teach all these division strategies I can't think too many times where we've just gotten a straight old sort of divide this by this. If they do that it's usually that style, and that's a fairly standard objective isn't it, for key stage 2 and 3, by a 2 digit number'. (first round of interviews)
- 'At least one addition, like just calculation, this number add this number, or this number multiply by this number; calculations they are always written horizontally.' (first round of interviews)
- 'I think about multiplication and division, I think most years there's always a standard long division, long multiplication question, always…. You know I think, you know, in most papers, or most years, there's a standard long multiplication and standard long division in there to do at some point, and it's normally those ones where you have a box to show your working or show your method, so it's normally the two mark questions there.' (first round of interviews)
- 'Um…I don't think so, no, it's hard, because you didn't have a paper like this before, because it was all kind of in context, even the mental maths pretty much was most of them were in a context before, so that's hard really, to have anything, you know. In the old papers, I seem to remember you got very, very few of these actual context free ones, just the odd one or two, but not many of them. So that's actually very, quite different.' (second round of interviews)
- 'Obviously we didn't have an arithmetic paper before, but you would always have some sort of calculation papers, so question one just a straightforward addition calculation, so that I would say is similar to what we would normally have, the only difference being obviously they wouldn't have had this grid background, so that's applicable for all of these questions, but that is typical and we've seen that before.' (second round of interviews)

Description of instructional strategies

Quotes for instructional strategies to teach word problems

- 'I have done some where I've just given them a sheet of problems and said we are not even going to work it out, but just write with it whether it's a multiplication or a division, or it's an addition or a subtraction, so it's just kind of getting them to think of key words, like how many more than the number, or altogether. So it's identifying the key words, but it is hard to take out, I always say take out the irrelevant things like the names, does it matter whether it's Susan, James, or Ishmar, or whoever it is, so take out what's irrelevant, look at the figures, and look at these key words, think about what it is. I've also done it with children drawing, so they can actually draw a little picture, even if it's only little fiddly men, you know, there are so many in the class who like ice cream, so many like this, how many altogether kind of thing. So I've done that, with the drawing, as well, especially with division questions and things on multiplication questions, how many of so much, how many cars do we need, well put four in each car and see what we've got left. So I have done that with them. Yeah, but it is difficult, and as I say that's one of the problems with teaching, unless you are dedicated enough to sit down and write all the word problems, which I used to be years ago, yeah resources, they already know what they are doing. And that's the problem I think with mathematics, the children can't compartmentalise things, and they don't necessarily transfer from one thing to another.'
- 'We do actually download an image of a Sats question which we might start the lesson off. We might start it with nothing at all, we might cover up the question, we might cover up some of the numbers, and get the children to actually unpick the question. What do you think this is going to be

about? You know, Tom had thirty six pounds, that's all you can see, OK, talk to somebody, what's that going to be about? And then they'll come up with oh somebody had more, or somebody had less, or so and so, and how much had they had altogether? So we get a lot of talking around the questions without the children actually having to do the mathematics and really thinking about what they think it's going to be.'
- 'Yes, that was something just one of the boys spotted, he said is there ever a remainder? And I said not that I know of. I could happen, but if you...usually it's a chunking calculation and it's usually within a word problem. The fact that it's a word problem often puts it in context, which means a remainder doesn't actually make sense, most of the time anyway, you don't have point two of a child I say to them, and they sort of think about it like that. Often they do come out with a remainder and I say think about it logically, oh no that doesn't make sense, go back and check it, and actually most of the time there isn't one. So they know to double check if they end up with one because it's probably wrong.'

Quotes for instructional strategies to teach 2-step money problems
- 'I get kids to read the question properly'. And 'we go through word problem strategies, read the problem twice, underline the key things, is there anything that is not important, things that you don't really need to worry about, well I suppose the fact that they are apples doesn't really matter in terms of your calculations, so we've got a sort of a protocol of how we work through a word problem. I would encourage the children to show all their calculations on this sort of question'.
- 'well there's one thumb and four fingers, but we talk to them about read the question, underline the key vocabulary, work out what operation it is, solve it, do the inverse'.
- 'If you had 10 pounds to start with is it realistic to say you've got 11 pound 23 now? So actually thinking about real-life situations and whether your answer is going to fit.'
- 'So I think that's what we have to do, break it down, underline it and say right, if I took this out could you still do the question? And then we'll reverse it when they are more confident, and I'll put up a kind of closed text and just say right, I'll write this, and wherever I've put a blank you fill in a name here, you fill in an amount here, and another amount here, and they'll sort of see oh they are the key bits, actually the fact that she buys coffee, that's not relevant, so it's kind of breaking it down, getting them to write it themselves.'
- 'so giving them exposure to using money and getting the coins out and manipulating the coins and actually doing the transactions with their partner, so they get a real sense of oh that's what it means, and they soon tell you if they've been short-changed and they get that concept pretty quickly'.
- I always say to them look if it's got pence there and pounds there you have a choice, you either everything becomes pence or everything becomes pounds, it's making those decisions and choices.
- This one I would say line, number line, the jumps, seeing it, especially those less able, seeing where do you get from two twenty five to five pounds, counting up. Not subtraction, even though they are finding the difference, I would not say do a compact method of subtraction, no, because they are likely to get the wrong answer because it's 2 point 00 and 0 point 05. So you are dealing with numbers that are likely to cause you issues, and you are likely to get the wrong answer.
- 'I always say there's not a straight division one and it is the hidden one, it is division and recognising it and knowing the clues. I always say to them the one, the one is the sharing, you are picking out one, that is your division, so just spotting those clues within the question...'
- 'But I have sort of said to them, look at the mark points as well, make sure you do your workings out because it's one point for that and one point for the answer. If you do your workings out you do get a point. So I try and tell them, especially when it's got a question where you have to explain what you've done.'
- 'And with money problems, difference problems, we'll start putting them into multistep, we'll work from one step to multistep and we'll look for ways in which we can make it really complicated, or we'll ask them to make up their own types of questions, or we'll ask them to use things like the inverse to work a question, and answer back to the question and that sort of thing, and again making them familiar with the types of questions that have come up in the past.'.

Quotes for instructional strategies to teach context-free calculation

- ‘One thing we have started doing is you know when I was saying about writing the questions, addition question, they never write their addition questions as a column, they always write the numbers beside each other, and we have made a conscious effort to write, if we are ever doing anything on the board we write it out like that and then put it into a column, because a lot of the children will just see the two numbers next to each other and then will think that’s how I have to solve it. They sometimes worry about changing things, because the test says this is how it’s done, so we have made a concerted effort on that as well.’
- ‘Fifty percent of my time should be spent doing number calculation; you kind of end up going along with that, so you in those tests at the end of the year I’m going to get the majority on calculation and shape, so if my children are good at place value and they are good at calculation then they can potentially get half the marks and then that’s why these boosters that I’m doing now.’

Quotes for general changes in the test (introduction of reasoning papers)

- “For me it’s that kind of, that explanation as to how they’ve got to an answer, what their thought process of that is, and being able to communicate that effectively.”
- “Yes, it has, and I think it’s changed quite drastically, with a much larger approach on problem solving …”
- “We have tried to put in more puzzles, more problems”
- “Reasoning is something, actually we put a lot of importance on reasoning because we were under the impression that these children needed to be able to explain word for word exactly why they’d chosen a method, how they’d used that method, and why they came to the answer that they had reached.”
- “So they are able to explain their answers fully, explain how they have chosen
- so teaching the curriculum and then allowing lots of opportunities for problem solving and open-ended tasks and trying to incorporate multistep tasks.”
- “… that children need to be able to apply what they can do in different contexts, so we try to expose them to situations where they need to work out one, two, even three step problems,”
- “We did a lot around reasoning and developing those skills so working systematically, hypothesising, looking at conjectures, whether trying to prove or disprove, and also making up our own as well …”

Appendix 6. Characteristics of questions used to assess problem solving (ACME)

ACME sets out a list of desirable characteristics of questions used to assess problem solving (ACME, June 2016[9]). The report recognises that not all characteristics will be found in a written test but it is worth discussing which of these are to be found in Key Stage 2 tests and which are not:

Varying the presentation
ACME recommends varying the presentation of questions by including:
- Questions set within unfamiliar contexts or formats
- Questions based on authentic scenarios
- Questions which are open-ended
- Questions requiring the translation of text into mathematical forms

An analysis of the KS2 mathematics test indicates that there are questions set within unfamiliar contexts or formats, with some of these being based on authentic scenarios. As pupils will have had different life experiences, what may be authentic for one pupil, will not be for another.
The KS2 test is time dependent, which means that open-ended questions are limited to those that have a restricted number of responses. As all the questions are written, they all require translating into mathematical forms.

Making choices
The second criteria is ‘making choices’; ACME recommends including:
- Questions in which a method of approach is not immediately obvious
- Questions which can be solved by a range of methods
- Questions requiring the selection of relevant information
- Questions with multiple steps but little or no scaffolding

Our analysis of the KS2 test indicates that there are questions where a method of approach is not immediately obvious, although this will depend on the experiences of the pupil. A range of methods could be used. Most word problems will require the selection of relevant information, but it is usual for all the information to be needed. Whilst some question need only one step, there are some that need multiple steps; a maximum of two-steps is usual.

Thinking mathematically
According to ACME, tests need to assess whether pupils can think mathematically, including:
- Questions requiring abstract thinking
- Questions requiring assumptions to be made
- Questions requiring movement between mathematical representations; e.g. numerical, graphical, diagrammatic, algebraic
- Questions requiring the synthesis of mathematical ideas or approaches

The KS2 test has items which require abstract thinking, but it will depend on pupils’ experience if they solve the question through abstract thinking or a more scripted approach. There are no examples where assumptions need to be made, but this does not mean that some pupils will not make assumptions. Some questions do require movement between mathematical representations, and this would also lead to a synthesis of approaches.

Obtaining results
ACME recommends:
- Questions leading to a range of different possible solutions
- Questions involving the interpretation of solutions
- Questions requiring the communication of solutions

[9] http://www.acme-uk.org/media/35168/acme%20assessment%20of%20problem%20solving%20report%20-%20june%202016%20-%20final.pdf

As the Key Stage tests are marked externally, the process does not lend itself to a range of different solutions. There are some questions where given solutions need interpreting, but these are limited to items that ask for an explanation to be given. There is no expectation that solutions will be communicated.

Making modifications
Last, ACME recommends:
- Questions enabling the critical analysis of solutions
- Questions requiring the evaluation of solutions
- Questions in which information can be revised
- Questions in which approaches can be refined

These questions cannot be found in the KS2 test as making modifications to test questions does not lend itself easily to a test situation.

Whilst there are limitations to the assessment of problem solving within a Key Stage 2 mathematics test, there are aspects which are addressed, particularly in the types of items which are referred to by teachers as ‘word problems’ or ‘2-step money problems’. Below we’ll discuss the extent to which these types of short word questions, which are set in a context, are suitable to test problem-solving.