

Education Select Inquiry into Key Stage 2 Tests 2016: Mathematics

This response relates solely to the mathematics tests at the end of Key Stage 2 in 2016, available at <https://www.gov.uk/government/publications/key-stage-2-tests-2016-mathematics-test-materials>. There are three papers: paper 1 tests arithmetic (40 marks), and papers 2 and 3 (35 marks each) are intended to test reasoning and problem solving. This response to the Education Select Committee questions draws on the experience of the 69 members of the joint Mathematical Association/Association of Teachers of Mathematics¹ Primary committee, which comprises primary teachers and others working in primary mathematics education. This group has also written to the Secretary of State for Education, the Rt Hon Justine Greening, about their concerns and the impact those are likely to have on primary teaching and learning of mathematics.

The Select Committee inquiry asks about:

- ***The purpose of primary assessment and how well the current system meets this:***

The primary purpose of statutory KS2 testing is² to ascertain what pupils have achieved in relation to the attainment targets outlined in the national curriculum (2014) in mathematics. The Bew report (2011)³ argues that one key use of this is for school-level accountability – but that must be fair and representative. This report suggests accountability is better achieved by greater use of teacher assessment as opposed to external assessment – and that, further, teacher assessment is better able to support formative use to promote optimum pupil learning over the transition from primary to secondary education.

- There is widespread concern among primary practitioners that the 2016 Year 6 (and Year 2) external tests were in many ways not fit for purpose and that, additionally, the Year 6 threshold for ‘expected progress’ was set at a level incommensurate with DfE claims that this would be at the equivalent of the previous level 4b. Details of such claims are given below.
- Other intended uses are stated to be (Bew, *ibid*) for school benchmarking and for parent information. Any of these uses is of course justified only if the assessments are both valid and reliable.

- ***The advantages and disadvantages of assessing pupils at primary school:***

As at all transition points, assessment for formative purposes at the end of primary school, as well as during it, is essential if secondary schools are to know what students entering Year 7 have met and what skills, knowledge and understanding they have evidenced. In this way secondary schools can more effectively both build on prior learning and enable young people to fill gaps in their learning. Teachers will understand that students in a new environment, surrounded by new people and expectations and maybe having made little use of ideas since they were assessed, might not show the same characteristics once in secondary schools – or might indeed exceed them: communication of formative assessment can give a starting-point but should be neither a ceiling nor an absolute floor.

- There have for some years been concerns by both primary and secondary teachers that a given ‘level’ achieved in Year 6 external assessments could be achieved in a variety of constructive or less constructive ways, resulting in more or less embedded or deep learning, and that therefore even breakdowns of pupil performance in Year 6 tests were of very limited use to secondary teachers – and of limited validity for purposes of school accountability (e.g. Ofsted 2012)⁴.
- Nevertheless, it is widely recognised that school accountability is desirable, and the Bew report suggest other ways in which this could be achieved, including by external sampling of performance.
- Validity of timed written tests is also an issue: Year 6 tests are intended to assess mastery of the 2014 National Curriculum Programme of Study in Mathematics for KS2⁵. Such validity will always be questioned when the curriculum requires the development of skills such as problem solving, reasoning, practical application and communication in a variety of formats, none of which is easy to assess under short written test conditions, especially with relatively young children.
- Further, the 2016 Year 6 cohort only experienced two years of a programme of study intended to last six years, so will inevitably have mastered different subsets of the curriculum at the point of assessment. Which parts have been mastered, and how deeply, comprise very important formative information for secondary teachers receiving students in September 2016, but hardly form the basis for robust summative assessment of school or attribution of a crude level to an individual.

¹ See www.m-a.org.uk; www.atm.org.uk

² Standards and Testing Agency (2014) *Key stage 2 mathematics test framework* available at <https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study>

³ <https://www.gov.uk/government/news/key-stage-2-testing-assessment-and-accountability-review-final-report-published>

⁴ Ofsted (2012). *Mathematics made to measure*. Her Majesty’s Stationery Office, London

⁵ Available at <https://www.gov.uk/government/publications/national-curriculum-in-england-mathematics-programmes-of-study>

- However, it is recognised that while there is concern at the level at which ‘expected’ thresholds have been set, particularly for this year group, there is some merit in a system which ‘flags up’ a *small* minority of students who are not yet in a position to access the secondary curriculum, and encourages teachers to look more closely at individual mathematics needs rather than labelling a young person with a level or artificial sub-level.
- **How the most recent reforms have affected teaching and learning logistics and delivery of the SATs:**
The 2014 curriculum has at times been challenged as being overly - or inappropriately - ambitious⁶. Nevertheless, outcomes from recent Ofsted inspections and the Summer 2016 tests suggest teachers and pupils have worked hard to try to adjust to the increased expectations, with about 70% of Year 6 achieving the ‘expected standard’. That still represents an increase in the number of young people beginning secondary school who are being widely labelled ‘failures’, and is a departure from the DfE commitment to comparability with a level 4b, since, in 2015, 77% achieved a level 4b or above (and 87% a level 4 or above). Such labelling can prove disheartening to both pupils and teachers. These early years of a more aspirational curriculum are of course unusually challenging for all parties.
 - Teachers widely report continued high pressure to ‘teach to the tests’ given these are high stakes tests for schools. That would be fine if the tests were valid, but see comments below.
 - Teachers also report being widely challenged to know how to interpret the new curriculum, with its renewed emphasis on problem solving and reasoning.
 - That, and increased expectations re content coverage and mastery, inevitably make considerable demands on teachers (and on pupils), and this has been exacerbated by pressure to adapt to a new curriculum in a short time period. Teachers are being extensively supported by NCETM (www.ncetm.org.uk) but it is early days to understand the nature of the long-term trajectory of teacher skills and confidence with the new curriculum. Teachers are largely confident about the physical delivery of the SATs and support from the sample materials and instructions provided.⁷
- **Training and support needed for teachers and senior leaders to design and implement effective assessment systems:**
Teachers report that the NCETM (www.ncetm.org.uk) is trying hard to develop common understandings of the new curricula, although these are not fully reflected in the 2016 tests. Such shared understandings are a prerequisite for the development of effective assessment systems, and would have been better supported by greater exemplification in the 2014 curriculum. For details see below. As suggested above, the range of the new curriculum is not validly assessed by timed written papers although these are suitable for some aspects of it.
 - Teachers report they are developing ways of effectively assessing most aspects of the new curriculum formatively, but that this will take time.
 - Some recent approaches to mastery (such as those in some schools in Singapore or Shanghai), supported by high quality CPD and high quality expert-developed textbooks, are claimed to be helpful, particularly where there is already a good depth and breadth of subject knowledge, but Ofsted (2011)⁸ suggests some primary schools are achieving this by other means and with other resources (and CPD).
 - Greater subject-specific expertise is also needed if schools are to deliver the aspirations described in Ofsted (2011) and support young people in achieving at a level commensurate with those in the highest-performing jurisdictions. Teachers report that where there are MaST (trained maths specialist) teachers in place, teachers have much more confidence and support to develop their practice. Unfortunately, funding for such development has been withdrawn.
- **Next steps following the most recent reforms to primary assessment:**
The Bew inquiry recommended a move to external sampling complemented by extensive teacher formative assessment. Additionally, such external tests need to be developed to ensure greater validity, as described below, as well as to find ways to assess aspects of the curriculum not well assessed by timed written papers. See also 6 below.

In addition to the answers to the specific SC questions above, the MA/ATM Primary committee say that

1. Is the format and content of these tests appropriate

- for learners in this age group?

- to assess curriculum coverage at the end of the primary cycle?

- The format is largely manageable for pupils at this age.
- The tests assessed the content of the curriculum, but with too little weight given to key processes of reasoning, problem solving and communication: see point 2 below.

⁶ For example, see <http://www.acme-uk.org/media/10145/collated.pdf>

⁷ Available at <https://www.gov.uk/government/publications/2016-key-stage-2-mathematics-sample-test-materials-mark-schemes-and-test-administration-instructions>

⁸ Ofsted (2011). *Good practice in primary mathematics: evidence from 20 successful schools*. HMSO, London.

- The Key Stage 2 curriculum itself is widely considered to be overly-demanding/inappropriate in some respects, as communicated during the National Curriculum consultation, and the mathematics education community widely considers more would be achieved with less content to cover by the end of Year 6, enabling the achievement of greater depth on which to build further mathematical understanding.

2. Are they good assessment tools?

-What do they measure vis a vis curriculum content and how do they measure that?.

-Do they measure the right things in the right ways?

- The intended curriculum content is adequately assessed by the 2016 mathematics tests.
- The 2016 tests are not consistent with the aims of the new curriculum: fluency and conceptual understanding, reasoning and problem solving.
- The exemplification of interim standards published in January 2016 was also inconsistent with the curriculum aims and contained many mathematical errors. The materials were not subject to the same rigorous quality assurance as National Curriculum Test items, which is unsatisfactory when teachers are insecure with a new curriculum. It is, further, surprising the same 'interim standards' recently re-issued (July 2016) without revision despite specific concerns having been raised with the DfE.
- In the 2016 tests, teachers consider over half of the written arithmetic test (40 marks) would in fact be better done mentally, but the curriculum no longer includes 'mental methods as a first resort'. This seems a retrograde step since mental maths builds up both conceptual understanding and fluency with number as a foundation for future mathematics learning, including algebra. Such an expectation should be restored as guidance to the national curriculum in mathematics, but meanwhile, more valid assessment of formal written methods would be achieved by using items for which those are the appropriate tools.
- Further, the use in the tests of a squared grid encourages children to set out a formal algorithm whether that is appropriate to the calculation or not. This does not encourage intelligent choice of application of the mathematics learnt.
- Fluency is far more than quick and accurate use of an algorithm, whether appropriate or not: it also requires highly-valued flexible and generalised use of procedures and their inverses. The arithmetic paper does not currently test these aspects at all, and much of the reasoning papers is simply procedural. If children are to be taught deep fluency, test papers must be seen to value it.
- Teachers estimate that only about 25% of the 'reasoning' papers (total 70 marks) comprise items that require mathematical thinking (i.e. problem solving and reasoning): the rest represent routine operations. The design of the tests needs to move to better reflect the aims of the curriculum, with an improved balance between 'fluency' in core skills (currently representing 75% of the 'reasoning' papers as well as all of the arithmetic paper), and problem solving/reasoning.
- Further, items on these latter papers often require a number of steps for solution, yet there are usually just two marks, meaning that a partially correct solution often receives no credit at all. The approach to marking these tests therefore needs re-thinking so that evidence of conceptual understanding is rewarded.
- Assessment approaches need to be developed so as to more validly assess problem solving and reasoning through a variety of approaches, including teacher assessment: at this age, these aspects of children's mathematical functioning, especially if embryonic, are not validly assessed by limited time written papers.
- Teachers suggest it may be worth considering awarding two outcome marks for mathematics – one for fluency and one for problem solving and reasoning.
- (The Key Stage 1 tests are even more inconsistent with the curriculum aims as they rely primarily on recall of facts and techniques with little problem solving or reasoning. They do, however, reflect the increased content in the Key Stage 1 curriculum that is in many ways age-appropriate.)

3. Are the tests robust in their design?

-Is the format well designed? Are questions well articulated and appropriately ramped through papers?

-Is there enough time for learners to show what they know?

-Are there enough questions to make valid judgements?

-Are mark schemes clear?

- Teachers report that these aspects are largely satisfactory: it is the inappropriateness of the question content and the mark schemes that concern them.

4. Where is the purpose of the tests articulated in policy documents? Do the tests fulfil the purposes for which they are set?

- Tests of such limited validity cannot robustly fulfil their intended purpose and uses as described in the references given.

5. Are there any unintended consequences of the tests that the SC need to be aware of and what is the evidence for those (e.g. any skewing of the mathematics curriculum)?

- The labelling of 30% of young people entering secondary school as 'not having met the expected standard' can have a very negative effect on both young people and their teachers, and is likely to lead to more 'teaching to the test' for any external assessment. It is therefore imperative that any external tests used have a greater validity than those used in 2016.

- If tests remain so skewed towards fluency at the expense of both problem solving and reasoning, then it is to be expected that teachers will skew their teaching to reflect that. The renewed emphasis on problem solving, reasoning and communication in the curriculum is in line with international best practice and supports the growth of good foundations for the enjoyment and mastery of mathematics at later stages. This mismatch of high stakes assessment with curriculum should therefore be addressed as a matter of urgency.

6. What specific actions could realistically be taken to improve tests for 2017?

- Questions intended to test formal written methods should require those only if they are the most appropriate method. A more valid test of conceptual mastery would be to ask pupils to choose formal or informal methods to answer a range of arithmetic questions.
- In the 'reasoning' papers, pupils will of course also use mathematical skills, knowledge and techniques – but these should be in genuinely problem solving situations or where reasoning is genuinely needed, and where it has to be communicated in an appropriate form.
- Careful consideration should be given to the balance of credit to be given to fluency and to other processes such as problem solving, reasoning and communication, so that tests better reflects the curriculum.
- The use of squared paper should be reviewed.
- In the medium-term, consideration should be given to the adoption of the Bew recommendations for teacher assessment complemented by sampling. This could allow more valid assessment of pupils' mathematical functioning, support the full enactment of the curriculum, and benefit the learning of young people as they move to secondary school, while also improving the validity of intended uses of the end of Key Stage tests. However, the effective development of such a system is predicated on continued efforts to further expand the mathematics subject-specific expertise of primary teachers.
- *As a complementary point it is widely felt that the proposed Year 7 'progress check' will distort learners' early secondary mathematics experience, as teachers will feel pressured, as Year 6 teachers do, to 'teach to the test'. Ofsted (2012)⁹ is clear that such teaching leads to superficial learning rather than building deep foundations that secure future progress. The Meeting of Mathematics Subject Associations (MMSA) has written to the Secretary of State for Education, and to the Minister for Schools, to draw attention to the concerns of the mathematics education community about the proposed Year 7 'progress check'. We are pleased to see (October 2016) that the DfE has now announced the use of such tests is to be optional and formative for internal use rather than compulsory, summative and public, at least for the next two years.*

⁹ Ofsted (2012). Mathematics Made to Measure. HMSO