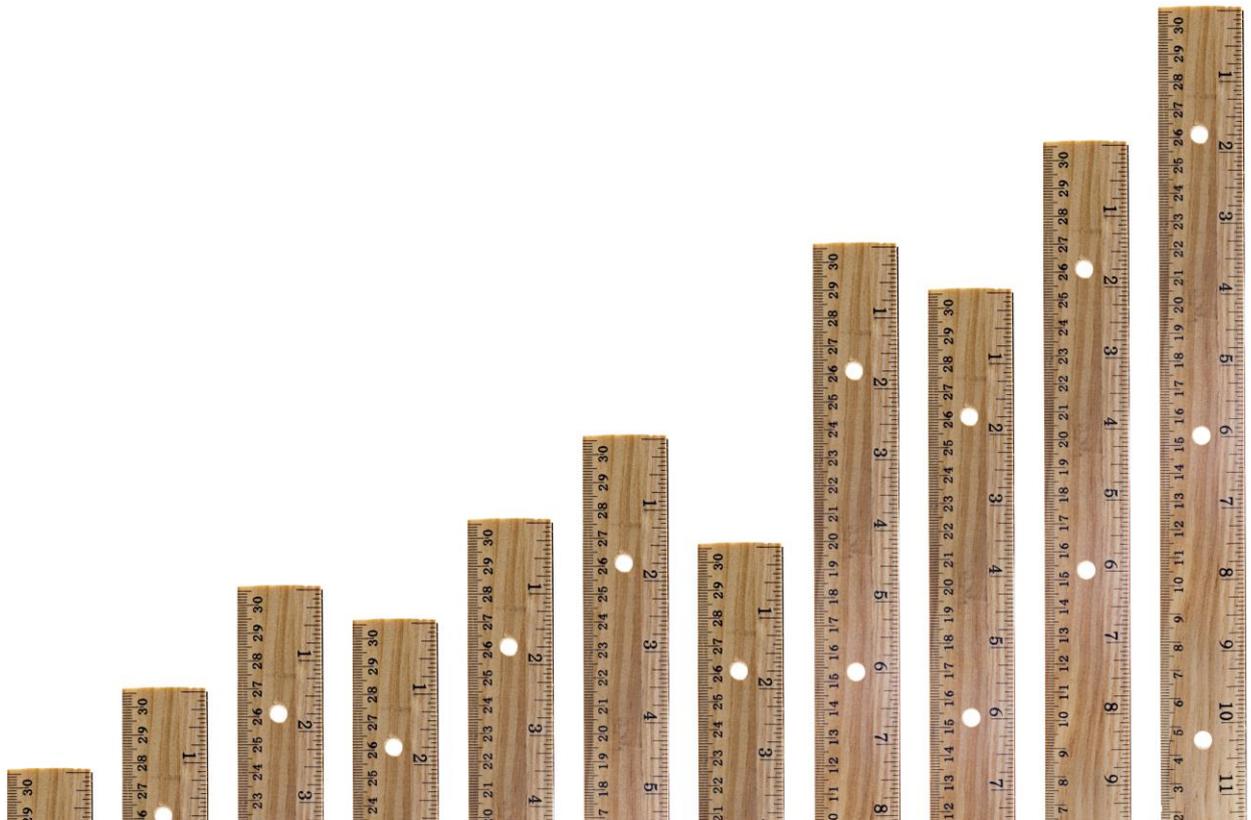


# Making progress

The future of assessment and accountability in primary schools

Tom Richmond and Eleanor Regan



## About the author

**Tom Richmond** is the director of the EDSK think tank.

Tom has spent almost 20 years in the world of education. He began his career teaching A-level Psychology at one of the country's leading state schools, having gained a BSc in Psychology from the University of Birmingham and an MSc in Child Development from the Institute of Education in London.

After three years in teaching, he moved into politics to work on policy development and research across the education, skills and welfare sector. This included roles at think tanks such as Policy Exchange and the Social Market Foundation, Pearson, G4S and working for an MP.

He subsequently spent two years as an advisor to ministers at the Department for Education, first under Michael Gove and then Nicky Morgan, where he helped to design and deliver new policies as well as improve existing ones. After leaving the Department for Education, he spent two years teaching at a Sixth Form College before moving back into education policy and research, first at the Reform think tank and then at Policy Exchange before deciding to launch EDSK.

He has also written extensively for publications such as the TES and Schools Week and has appeared on numerous media outlets, including the BBC News Channel, Sky News, BBC Radio 4, BBC Radio 5 Live, LBC and TalkRADIO.

**Eleanor Regan** is a researcher at the EDSK think tank.

She completed a BA in Geography at the University of Southampton, where she developed a strong interest in issues of inequality, particularly surrounding social class.

In her spare time, she is an admin volunteer for a social enterprise that focuses on improving children's attainment in mathematics.

## Acknowledgements

EDSK would like to express our sincere thanks to the National Association of Headteachers, Pearson and GL Assessment who kindly sponsored this research project. It should not be inferred from their sponsorship that these organisations necessarily agree with the report's conclusions and recommendations as these are independently determined by EDSK.

Thanks also go to those who commented on specific sections of the report or provided expert input for one or more of the recommendations.

The views expressed in the report are attributable to the authors and do not necessarily reflect those of EDSK's advisory board or EDSK's supporters. Any errors remain the responsibility of the authors.

## About EDSK

EDSK is an independent, impartial think tank.

Our mission is to design new and better ways of helping learners of all ages succeed, particularly those from disadvantaged backgrounds.

We aim to achieve this by conducting research on topics across the education and skills landscape and then producing evidence-based recommendations for how the system needs to change, both now and in the future.

Our research covers primary schools, secondary schools, colleges, apprenticeships and universities. Whichever topic or policy area we are investigating, our priority is always to produce better outcomes for learners while also ensuring that the greatest support is given to the most disadvantaged members of society.

We work closely with a range of individuals and organisations to ensure that our research reflects a wide range of evidence on any given issue. EDSK retains copyright and full editorial control over our research, irrespective of how it is funded.

EDSK is a not-for-profit think tank registered in England and Wales.  
(Company Registration No. 11795717)

Find out more at [www.edsk.org](http://www.edsk.org)

# Contents

	Executive Summary	1
1	Introduction	10
2	Current primary school assessments	14
3	The new ‘reception baseline assessment’	18
4	The impact of SATs on teachers and schools	24
5	The impact of SATs on pupils	30
6	The phonics check and multiplication check	36
7	Alternative forms of assessment	43
8	Recommendations	50
9	Areas for further consideration	63
	Conclusion	66
	References	68

## Executive Summary

Having been introduced 30 years ago, SATs ('standardised assessment tasks') are one of the most prominent features of primary education in England, yet they remain as controversial now as they were back in the early 1990's. To their supporters, SATs are an invaluable tool for measuring the attainment and progress of both pupils and schools in an objective and consistent manner across England's 16,800 state-funded primary schools. To their critics, SATs are a stressful, unnecessary and burdensome way to monitor pupils and schools and should therefore be scrapped. In recent years, SATs for 7-year-olds and 11-year-olds have been joined by other formal nationwide assessments. A 'phonics check' was created for Year 1 pupils in 2012 to test their basic reading skills, and the current academic year will see two new assessments rolled out: a 'reception baseline assessment' to test the cognitive skills of reception pupils and an online 'multiplication check' to assess Year 4 pupils on their multiplication tables up to  $12 \times 12$ .

As evidenced by the last General Election campaign, debates over the importance and usefulness of primary school assessments (and the associated accountability system) are still as strongly contested as ever. The collapse of all national assessments in 2020 and 2021 due to the tragic outbreak of COVID-19 has revived these disagreements once again, with both supporters and critics of SATs claiming that the events of the past 18 months have demonstrated why they were right all along. In truth, England's recent performances in international comparative tests appear to suggest that some reforms to primary assessments over the past decade are indeed bearing fruit, but those same tests show that England remains well short of being a world-class education system. Consequently, after analysing the available research evidence on the full range of primary school assessments being used in the 2021/22 academic year, this report set out to design a new assessment and accountability system for primary education that would deliver the following improvements:

- Promoting high standards for pupils in all year groups in terms of their progress and attainment
- Reflecting the contribution that schools make to their pupils' education in a fairer and more proportionate manner
- Ensuring that the assessment system supports high-quality teaching and learning rather than encouraging excessive test preparation
- Providing more accurate information on the performance of pupils and schools as well as performance at a national level
- Reducing the assessment burden on teachers and schools

## The reception baseline assessment (RBA)

At present, the progress of primary school pupils is measured by the improvement shown between their Key Stage 1 (KS1) SATs results in Year 2 (age 6-7) up to their Key Stage 2 (KS2) SATs results in Year 6 (age 10-11). As far back as 2013, the Department for Education (DfE) had questioned this arrangement, pointing out that “a baseline check early in reception would allow the crucial progress made in reception, year 1 and year 2 to be reflected in the accountability system and would reinforce the importance of early intervention”. Many stakeholders agreed that pupil progress should be measured from the earliest possible point, leading the DfE to announce that they would replace the current KS1 starting point with a ‘reception baseline’ to measure pupil progress from age 4 up to KS2 SATs.

Although the DfE’s first attempt to develop a new ‘reception baseline’ test was aborted after a pilot study in 2016, the National Foundation for Educational Research (NFER) was eventually contracted in 2018 to create the new RBA. The development of this new assessment involved over 7,000 primary schools, and the results of the subsequent pilot study were encouraging. The RBA is an activity-based assessment of language, communication and literacy (e.g. early vocabulary and comprehension) as well as mathematics (e.g. early understanding of numbers and patterns). Of the practitioners (mostly reception teachers) who tested the RBA, 82 per cent rated the mathematics tasks as ‘satisfactory’ or above and 65 per cent of practitioners trialling the literacy, communication and language tasks considered them to be at least ‘satisfactory’. What’s more, 84 per cent of practitioners rated the children’s interest and enjoyment of the tasks as at least ‘satisfactory’, and 89 per cent said that the children’s understanding of the tasks was ‘satisfactory’ or better.

Not all observers were convinced, though. A recent report from the British Educational Research Association criticised the Government’s plans to use the RBA as the starting point of a ‘progress measure’ for the whole of primary education. For example, the time lag between the RBA at age 4 and KS2 SATs at age 11 raises the question of how useful the results of the new 4-to-11 progress measure will be, seeing as many aspects of a primary school could change dramatically over six or seven years (e.g. a new head teacher and / or school leadership team). Presenting this new progress measure to parents as an indication of the level of progress that their child is likely to make at a school *now* (as opposed to several years previously) will be a questionable proposition.

Numerous other factors could distort the new progress measure. 20 per cent of primary pupils in England move school at ‘non-standard’ times of year, meaning that these pupils will end up being educated in a different school from the one in which they completed the RBA. It remains unclear how the DfE’s new progress measure will accommodate this degree of pupil mobility. In addition, any school with more autumn-born pupils could inadvertently be made

to look worse in terms of ‘pupil progress’ because summer-born children will on average have lower levels of attainment and thus potentially be able to demonstrate more ‘progress’ over the next seven years. Furthermore, in the absence of KS1 SATs, many primary schools will not deliver any other formal assessments after the RBA before their pupils move onto another school (e.g. switching from an infant school to a junior school at age seven). As a result, the introduction of the RBA means there will no longer be any measure of pupil attainment or progress for thousands of primary schools. An Education Select Committee investigation in 2017 also heard that the existing progress measures from KS1 to KS2 had been designed in such a way that they encouraged “deflating [KS1] results to demonstrate greater progress at the end of KS2”, yet the Government has not outlined how it will prevent this happening with the RBA in future. Such concerns do not undermine the validity of the RBA as an assessment of reception pupils, but they illustrate the pressure that will be placed on the RBA as it is delivered in primary schools from this year.

## The impact of SATs on teachers and schools

At present, KS2 SATs in Year 6 include three tests (reading; grammar, punctuation and spelling; mathematics) that are marked externally. Alongside these tests, teachers assess their pupils’ writing skills. KS1 SATs only cover mathematics and reading (with an optional grammar, punctuation and spelling test) and are marked by their teachers, along with teacher judgements of pupils’ level in science, writing and speaking and listening.

There have been many investigations of testing and assessments by parliamentary committees spanning multiple governments, yet they invariably conclude that national testing can play an important role in providing objective and consistent data about the standard of education across different schools. However, problems quickly arise when the data from any single test is used for too many different purposes, with one such Committee being told in 2008 that Key Stage tests such as SATs were being used for 14 separate purposes at the time. The same Committee also heard that a poor set of SATs results may result in “being perceived as a “failing school”, interventions by Ofsted and even closure in extreme cases” – illustrating how the high-stakes accountability system for primary schools is closely linked to SATs results.

One of the most worrying effects of this high-stakes accountability system is ‘teaching to the test’, whereby teachers ‘drill’ their pupils in a subject on which they will later be tested and devote a high proportion of time to test preparation, exam techniques and even question-spotting. The aforementioned Committee in 2008 “received substantial evidence that teaching to the test ...is widespread” and that “test results are pursued at the expense of a rounded education”. One survey found that for four months of the final year at primary school, teachers were spending nearly half their teaching time preparing pupils for KS2 tests.

In 2010, the Coalition Government expressed their concern that “especially in year six, there is excessive test preparation – with some children practising test questions for many weeks in advance of the tests.” Similarly, the Education Select Committee in 2017 heard evidence that ‘teaching to the test’ could mean KS2 SATs results were “severely inflated in being far larger than true gains in students’ learning”. The Committee recognised “the pressure that schools are under to achieve results at Key Stage 2” and that “many teachers reported ‘teaching to the test’ ...as a result of statutory assessment and accountability” – a point made by teaching unions almost a decade earlier when they told the 2008 Committee it was “hardly surprising that the focus is on ensuring that students produce the best results.”

A narrowing of the curriculum is another side-effect of the current assessment regime in primary schools, as non-tested subjects such as sport, art and music risk being neglected and having their curriculum time reduced to make more room for preparing for the next test. The Education Select Committee in 2017 cited evidence from Ofsted that most primary schools were spending four hours or more a week teaching English and maths, yet around two thirds of schools spent only one to two hours per week teaching science, and around a fifth spent less than one hour. In 2018, Ofsted reported that they had seen “curriculum narrowing, especially in upper Key Stage 2, with lessons disproportionately focused on English and mathematics” and “sometimes, this manifested as intensive, even obsessive, test preparation for Key Stage 2 SATs that in some cases started at Christmas in Year 6.”

Numerous committees and reviews in recent years have also expressed misgivings about whether performance tables (‘league tables’) give a fair reflection of what a primary school has contributed, particularly when schools operate in different contexts and sometimes with very different pupils. At the same time, the current assessment and accountability system has repeatedly been found to distract schools from improving teaching and learning as well as encouraging a ‘risk-averse culture’. Finding more accurate and reliable ways to measure pupil progress in a fair and proportionate manner would therefore represent a major step forward.

## **The impact of SATs on pupils**

In 2007, the independent Cambridge Primary Review (CPR) found that “for pupils in years 2 and 6 the notion of SATs looms large in pupils’ minds” and “some pupils feel that their learning is almost entirely focused on achieving good grades in SATs”. The following year, a parliamentary committee commented that “whilst some children undoubtedly find tests interesting, challenging and even enjoyable, others do not do their best under test conditions and become very distressed.” When the 2017 Education Select Committee inquiry spoke directly to pupils, they reported that “in general, the pupils were positive about taking SATs [as] they felt that SATs were a good opportunity to demonstrate what you knew”, although “some pupils told us they could get nervous or anxious about taking the tests”.

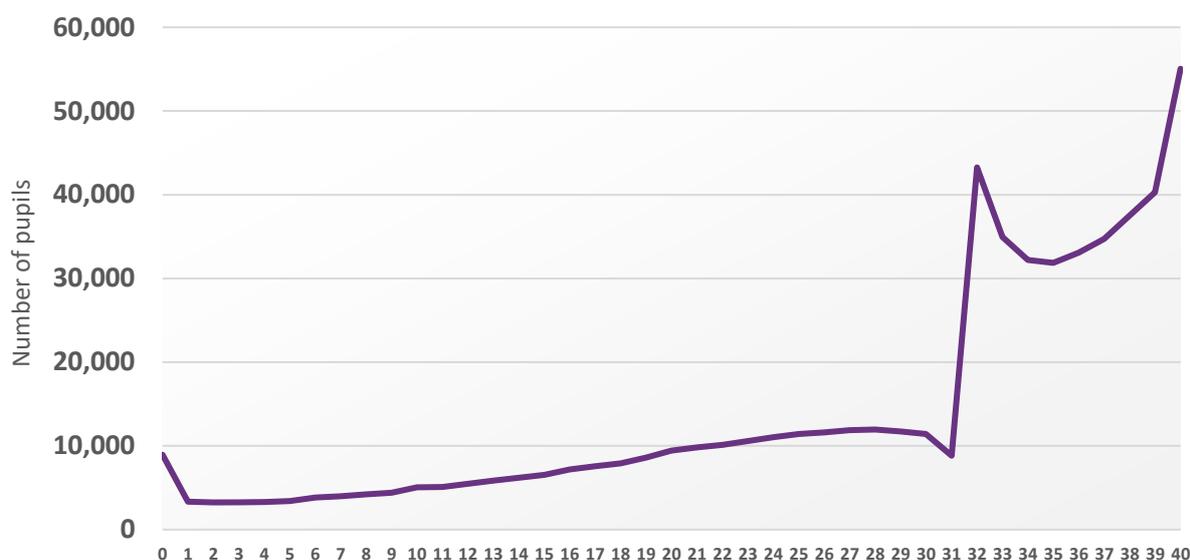
A survey by the polling firm ComRes in 2016 of 750 10 and 11-year-olds has provided one of the most comprehensive datasets on what pupils really think about testing in primary schools. When ComRes asked pupils how they felt about school tests, almost 60 per cent said they feel 'some pressure' to do well, with 28 per cent saying they felt 'a lot' of pressure and 11 per cent not feeling any pressure at all. When pupils were asked how they felt when taking tests at school and were given a list of words to choose from, the most common choices were 'nervous' (59 per cent), 'worried' (39 per cent) and 'stressed' (27 per cent). Even so, a notable (albeit smaller) proportion of pupils reported positive emotions such as being 'confident', 'excited' and 'happy'. Perhaps surprisingly, after some pupils acknowledged that tests at school could make them feel nervous or worried, 62 per cent said they either 'enjoy' tests or don't mind taking them. This indicates that primary pupils, like many adults, can understand the value and importance of tests even if they do not always relish them.

The question of what causes any stress or anxiety among pupils has also been investigated. When ComRes asked pupils who they would be most worried about knowing they did badly in a test, 'my parents' was named by 41 per cent – a bigger proportion than 'my friends' and 'my class teacher' combined. On a similar note, the CPR had earlier found that the effect of SATs on pupils varied considerably and depended, at least in part, on the actions of teachers and head teachers. For instance, previous research showed that "where schools have created a secure, non-threatening environment, high attainers begin to feel more confident and even exhilarated during the test period. However, under pressure, other pupils become demotivated and dysfunctional". In 2019 Amanda Spielman, the Chief Inspector at Ofsted, stated that "good primary schools manage to run key stage tests often with children not even knowing that they're being tested", and testing "only becomes a big deal for young children if people make it so for them."

## The phonics check and multiplication check

Since 2012, the 'phonics check' has tested whether Year 1 pupils can understand letters and sounds to an appropriate level, which would allow them to read many short words. The check lasts for about 5-10 minutes, during which time pupils must read a list of 40 words to their teacher. Pupils are scored against a national standard ('pass threshold'), which has been set at 32 marks ever since the test was introduced. The pass mark was originally "communicated to schools in advance of the screening check being administered so that schools could immediately put in place extra support for pupils who had not met the required standard." The national test scores in 2012 revealed – in the words of the DfE – "a spike in the distribution at a score of 32". The same pattern of results – with a steep rise at the exact pass mark – has been visible in every subsequent year of the phonics check.

## The number of Year 1 pupils (aged 6) in England scoring each possible mark on the first phonics check in 2012



Even though school-level results on the phonics check are not published, teachers and school leaders know that how well pupils are taught to read (including “how well the school is teaching phonics”) is classified as a ‘main inspection activity’ when Ofsted visit infant, junior and primary schools. In addition, for a school to be awarded a ‘Good’ or ‘Outstanding’ grade by Ofsted, inspectors must agree that, among other things, “staff are experts in teaching systematic, synthetic phonics”. The Ofsted inspection handbook refers to phonics 21 times in total. These demands offer a compelling explanation for the atypical results shown in the figure above. Such results also lend further weight to the argument that it is not possible for the phonics check to concurrently serve all the purposes for which it is being used – which includes monitoring individual pupils’ progress, school-level accountability, Ofsted inspections, national results and local authority standards.

The burden on schools has not been helped by the DfE interchangeably using the phrases ‘expected standard’ and ‘required standard’ to describe the pass mark. This lack of clarity emphasises how, despite there being no performance tables for schools’ phonics check results, the accountability system is almost intentionally placing considerable pressure on schools to ensure their pupils reach a certain level. When looking at the national proportion of pupils reaching 32 marks, there has been an improvement from 58 per cent in 2012 to 82 per cent in 2019. It is uncertain how much of this increase since 2012 is the result of genuine improvements in pupils’ ability to read as opposed to other factors such as greater familiarity with the test amongst teachers and pupils (a common occurrence in any new assessment). To illustrate the point, the proportion of Year 1 pupils passing the phonics check has increased by just a single percentage point since 2016.

From next summer, the new multiplication check (MTC) will require Year 4 pupils to complete an on-screen test of their recall of multiplication tables up to  $12 \times 12$ , with just six seconds to answer each of the 25 questions. Like the phonics check, a school's results on the MTC will not be published but will be made available to Ofsted while the national results will be reported separately. Inevitably, the fact that Ofsted will see each school's results draws this new test into the high-stakes accountability system, which is likely to reduce the accuracy of the MTC and thus compromise the DfE's attempts to track national standards. That said, the decision not to introduce a 'pass mark' (unlike the phonics check) means there is less likely to be the same clustering of results around a specific score. The DfE claim that the MTC is necessary because "knowledge and recall of multiplication tables is essential for the study of mathematics and for everyday life." That said, it is striking how the new MTC ignores addition, subtraction and division even though the National Curriculum says pupils should practice division and other related skills alongside multiplication. Perhaps unsurprisingly, the reception from stakeholders to the MTC has not been overly positive, with unions variously describing the new test as "unnecessary" and of "no benefit to pupils or teachers".

## Conclusion

Although existing primary school assessments appear to have made at least some contribution to improving standards, there are several weaknesses in the current arrangements that cannot be ignored. This report shows that primary schools are finding it increasingly difficult to focus on improving teaching and learning in the face of so many large and onerous assessments. Moreover, these assessments are often being used for too many purposes, which risks generating poor quality information for parents and policymakers. To overcome these issues, there is an urgent need to ensure that national assessments promote high standards but without distracting teachers and leaders from their core task of educating pupils.

To this end, this report puts forward a package of reforms (to be implemented by 2026/27) that aims to free up time for teaching and learning, reduce staff workload, track the progress of pupils and schools in a fair and proportionate way and monitor national standards over time. Three major shifts are required to achieve these goals over the coming years. First, there must be a concerted move away from the distorting and damaging effects of overbearing one-off tests such as SATs and instead use more frequent but shorter low-impact assessments. Second, pen-and-paper tests should be replaced by online assessments – as used in other countries – to make the tests more reliable as well as less burdensome for schools. Third, the one-size-fits-all standardised nature of SATs should be jettisoned in favour of 'adaptive' tests that provide a more personalised assessment. These changes would collectively represent a genuine step forward in terms of how, why and when we assess what primary school pupils know and understand. The report's recommendations therefore offer a positive and ambitious vision for the future of primary assessment and accountability in England.

## Recommendations

### A new approach to assessment

- **RECOMMENDATION 1:** The current assessment system – including the full range of Key Stage 1 and Key Stage 2 SATs as well as the ‘reception baseline assessment’ and multiplication tables check – should be scrapped by 2026.
- **RECOMMENDATION 2:** From 2026, assessments for reading, numeracy and spelling, punctuation and grammar (SPG) should be delivered through online adaptive tests. These tests automatically adjust the difficulty of the questions to match a pupil’s performance, with a correct answer leading to a harder subsequent question and an incorrect answer leading to an easier question.
- **RECOMMENDATION 3:** Pupils will take the new adaptive tests approximately once every two years. As a result, the tests will provide regular updates on how pupils are performing throughout primary education. Each test will last for around 30 minutes. School leaders will be able to choose how to schedule these tests during the academic year.
- **RECOMMENDATION 4:** To encourage pupils to develop their creative writing skills throughout primary school, pupils’ writing will be assessed in Year 2 (age 6-7) and Year 6 (age 10-11) through ‘comparative judgement’ exercises across all primary schools.
- **RECOMMENDATION 5:** To support parents’ understanding of how their child is progressing through primary education, they should be provided with a report at the end of Years 2, 4 and 6 that shows their child’s most recent results on the new adaptive tests and writing assessments.

### A new approach to school improvement and accountability

- **RECOMMENDATION 6:** To help benchmark their performance and identify areas for improvement, primary schools should be provided with an annual ‘profile’ of results for each year group to show how they are performing on the adaptive tests and writing assessments. The profile will also include national averages and local authority averages as well as the results achieved at ‘similar schools’ around the country.

- **RECOMMENDATION 7:** There will be eight headline measures of accountability for primary schools:
  - Pupils' average **attainment** and **progress** in reading
  - Pupils' average **attainment** and **progress** in mathematics
  - Pupils' average **attainment** and **progress** in spelling, punctuation and grammar
  - Pupils' average **attainment** and **progress** in writing

All eight measures will be reported annually for every primary school using descriptors from 'well above average' to 'well below average'.

- **RECOMMENDATION 8:** The phonics check should continue in Year 1 (age 5-6) but the concept of a 'pass threshold' should be removed. In addition, Ofsted should no longer be provided with each school's scores on the phonics check. These changes will reduce the likelihood of the results being distorted by having this assessment form part of the high-stakes accountability system.
- **RECOMMENDATION 9:** To ensure that national standards are measured separately from the performance of pupils and schools, the national standards for reading, numeracy and spelling, punctuation and grammar should be judged in future using 'sample testing'. This will involve placing a small number of identical (or very similar questions) into the adaptive tests every year so that standards can be monitored over time.

## 1. Introduction

*“Promoting children’s learning is a principal aim of schools. Assessment lies at the heart of this process. It can provide a framework in which educational objectives may be set, and pupils’ progress charted and expressed. ...The assessment process itself should not determine what is to be taught and learned. It should be the servant, not the master, of the curriculum. Yet it should not simply be a bolt-on addition at the end. Rather, it should be an integral part of the educational process.”<sup>1</sup>*

In July 1987, then Education Secretary Ken Baker appointed the Task Group on Assessment and Testing (TGAT) to advise the government on how the new ‘National Curriculum’ could be assessed. Although the National Curriculum would set out ‘attainment targets’ describing “the knowledge, skills, understanding and aptitudes which pupils of different abilities and maturity should be expected to have acquired at or near certain ages”,<sup>2</sup> the plans for how pupils would be assessed were constructed separately by TGAT. As shown in the above quote, TGAT were convinced of the value of using appropriate assessments to support both pupils and teachers.

When their final report was sent to the Department of Education and Science in December 1987, the TGAT was in favour of something “radically new”, in their own words.<sup>3</sup> Their main proposal was for national assessments to be introduced at the ages of 7, 11, 14 and 16, with these ages signifying the end of each ‘key stage’. In addition, these assessments would report on pupil performance using a system of ten ‘levels’ across hundreds of attainment targets in all ten National Curriculum subjects.<sup>4</sup> In the end, this enormous battery of tests went beyond the appetite of ministers. Nevertheless, the concept of national assessments at ages 7, 11 and 14 was born, accompanied by the phrase ‘standardised assessment tasks’ – more commonly known as ‘SATs’.

From 1991 to 1995, SATs were introduced at the end of Key Stage 1 (KS1) for 7-year-olds, the end of Key Stage 2 (KS2) for 11-year-olds and the end of Key Stage 3 (KS3) for 14-year-olds. Even before this rollout was complete, these new assessments began to cause controversy and there was a widespread boycott of SATs in 1993 in response to perceived concerns among teachers about the increased workload generated by the time-consuming assessments alongside the view that teacher assessments would be just as good.<sup>5</sup> The boycott forced the new Education Secretary John Patten to ask Sir Ron Dearing to review the National Curriculum and its accompanying tests. The outcome of this review was a slimmer and less prescriptive National Curriculum along with the call for government to “simplify the national tests in the core subjects as far as possible without sacrificing validity and reliability [and] in

particular, the time which is needed to administer the tests must continue to be reduced.”<sup>6</sup> SATs had thus survived their first protest, but the boycott was an ominous sign given how recently the assessments had been introduced.

The next major turning point came in 2003 when, faced with yet more boycotts, Labour’s Education Secretary Charles Clarke announced that SATs for 7-year-olds would be replaced by ‘teacher assessment’ to reduce the burdens on pupils and their teachers.<sup>7</sup> Another blow to the national testing regime came in the summer of 2008 when a marking fiasco led to ETS – an American non-profit testing company who were running the SATs marking process for the first time – being stripped of their contract.<sup>8</sup> Just a few months later, then Education Secretary Ed Balls decided to scrap KS3 SATs for 14-year-olds. More change came the following year, as the science tests taken by 11-year-olds every year were dropped as well.<sup>9</sup> To compound the recent troubles, another SATs boycott organised by the National Association of Head Teachers (NAHT) and the National Union of Teachers (NUT) took place in the summer of 2010 in response to the damage that the unions claimed these assessments were doing to the education system.<sup>10</sup>

Of all the SATs created in the early 1990’s, only English and maths tests for 11-year-olds remained in place as formal external assessments when the 2010 General Election came into view. However, rather than continuing down the same path of fewer and less frequent SATs, the Conservative Party manifesto advocated the opposite. A new “simple reading test” was to be introduced at the age of six because “every child who is capable of reading should be doing so after two years in primary school.”<sup>11</sup> The manifesto also declared that “we will ensure that our exam system is measured against the most rigorous systems in the world”, which translated into their decision to “keep Key Stage 2 tests and league tables [and] we will reform them to make them more rigorous.”<sup>12</sup> The newly-appointed Schools Minister Nick Gibb also defended SATs following the boycott in the summer of 2010, stating that “externally-validated tests give parents and professionals valuable information to gauge the standards of our primary schools and their pupils and play a vital role in accountability”. That said, he recognised that “we know the tests can be better, and we will be discussing with all parties how to improve the effectiveness, accuracy and rigour of the tests.”<sup>13</sup>

Under the Conservative-led Coalition Government, the pace of change rapidly increased. Following the delivery of the new ‘phonics check’ for reading in 2012, it was announced in 2014 that a new assessment at the start of reception (‘baseline assessment’) would be created. The intention was that this baseline assessment would allow the Department for Education (DfE) to measure the progress that pupils make from the start to the end of primary school when they sit their KS2 SATs.<sup>14</sup> In 2015, a new ‘multiplication tables check’ was outlined by the Conservative Government following their election manifesto pledge that “all children would be required to learn their multiplication tables, by heart, up to the 12 times table”.<sup>15</sup>

2015 also saw the demise of the concept of ‘levels’ within the National Curriculum, and there was another overhaul of primary testing in reading, writing and maths in 2016 to reflect the new National Curriculum implemented in 2014.

The evidence from international tests suggests that at least some of these reforms have begun to bear fruit. The latest Programme for International Student Assessment (PISA) in 2018, which assesses the abilities of 15-year-olds across 80 countries in reading, mathematics and science, found that the mean scores in England were significantly above average in all three subjects. The mean scores in reading and science in England had not changed significantly compared to previous PISA cycles, but in mathematics England’s overall mean score showed a statistically significant increase since 2015.<sup>16</sup> What’s more, the most recent Progress in International Reading Literacy Study (PIRLS) in 2016, which assesses the reading performance of pupils in their fourth year of formal schooling (Year 5 in England) across 50 countries, showed that England was significantly above the international median score. In addition, England’s score in 2016 was the highest average performance it had achieved since PIRLS began in 2001 and was also significantly higher than the majority of other countries. The improvements in England’s average score were largely attributable to increases in the average performance of boys and lower-achieving pupils.<sup>17</sup>

Although these international comparisons paint a broadly positive picture, it would be wrong to assume that the debate over primary school assessment is over. 30 years since their introduction, some teaching unions continue to insist that SATs have a detrimental impact on pupils, teachers, headteachers and even parents. Moreover, the Labour Party asserted in its 2019 election manifesto that “schools are being subjected to intensified testing, inspection, league tables and competition [that] aren’t improving pupil achievement or narrowing the attainment gap, but are contributing to a growing teacher recruitment and retention crisis.” Their solution was to “end the ‘high stakes’ testing culture of schools by scrapping Key Stage 1 and 2 SATs and baseline assessments, and refocussing assessment on supporting pupil progress.”<sup>18</sup> Similarly, the Liberal Democrat manifesto outlined their plan to “reduce unnecessary stress on pupils and teachers and end ‘teaching to the test’, by scrapping existing mandatory SATs and replacing them with a formal, moderated teacher assessment at the end of each phase and some lighter-touch testing.”<sup>19</sup>

Although the result of the 2019 General Election dampened such hopes for the foreseeable future, doubts over the suitability and effectiveness of the testing regime in primary schools unexpectedly resurfaced due to the tragic outbreak of COVID-19 last year. In effect, the assessment system for both primary and secondary schools collapsed in 2020 and 2021 following the Government’s decision to cancel all examinations and assessments scheduled for the summer months and to abandon league tables (‘performance tables’). While this reaction from ministers was entirely understandable given the circumstances, it emphasises

the inherent risks in relying on one-off and typically 'high stakes' tests to measure the attainment and progress of pupils as well as the performance of primary schools.

Rather than revisiting the debates of the past, this report will instead look to the future. To begin with, the report explores the qualitative and quantitative evidence related to the assessments that are set to be used in the 2021/22 academic year. When determining the effectiveness (or otherwise) of the current assessment arrangements, this report recognises that the accountability system in England is 'high stakes' i.e. there are potential consequences for a school if they perform poorly, and this is not likely to change in the foreseeable future. That said, there is still plenty of room for improvement in terms of how, when and why pupils are assessed during their journey through primary education. As a result, this report aims to reform the existing arrangements by designing a new assessment regime that delivers the following improvements:

- Promoting high standards for pupils in all year groups in terms of their progress and attainment
- Reflecting the contribution that schools make to their pupils' education in a fairer and more proportionate manner
- Ensuring that the assessment system supports high-quality teaching and learning rather than encouraging excessive test preparation
- Providing more accurate information on the performance of pupils and schools as well as performance at a national level
- Reducing the assessment burden on teachers and schools

By seeking to deliver these improvements, it is hoped that this report will be a valuable contribution to the debate over the future of assessment and accountability in primary schools across England.

## 2. Current primary school assessments

This chapter describes the statutory (required) assessments that pupils in state-funded primary schools in England must complete from the start of Reception up to Year 6. The Standards and Testing Agency (STA), part of the DfE, is responsible for developing and delivering assessments for pupils in this phase of their education. Two of the primary assessments described in this chapter – the multiplication tables check (MTC) in Year 4 and the ‘reception baseline assessment’ (RBA) – are new assessments that are set to be introduced in the 2021/22 academic year.

### ‘Baseline assessment’ – Reception (age 4-5)

Within the first six weeks of starting primary school, pupils will be required to take the new RBA. It is an activity-based assessment of pupils’ starting point in language, communication and literacy (e.g. early vocabulary and comprehension) as well as mathematics (e.g. early understanding of numbers and patterns). The assessment is presented in a task format, with several activities per task, that require pupils to respond in different ways such as oral responses, pointing and ordering / moving objects.

The RBA is a one-to-one assessment and lasts approximately 20 minutes. Teachers record the performance of each pupil as the assessment is taking place, but there is no ‘pass mark’. Although the total number of marks for the RBA is 39, the relationship between each activity and the number of marks available varies depending on the activity. Schools are not given the raw scores for each pupil, but teachers will receive a series of short narrative statements that tell them how their pupils performed in the assessment at that time. That said, the results of the RBA will be used by the DfE to establish a new starting point to measure the progress that pupils make by the time they complete their KS2 SATs in Year 6.

### Phonics check – Year 1 (age 5-6)

The ‘phonics check’ is a short assessment that tests whether a pupil has learnt to decode (i.e. understand) letters and sounds to an appropriate level, which will in turn allow them to read many one-syllable and two-syllable words. The goal of the phonics check is to ensure that pupils have the skills necessary to read fluently and to identify any pupils who may require additional support with their reading. It lasts for about 5-10 minutes, during which time pupils must read a list of 40 words to their teacher – half of which are real (e.g. cat, sum) and half are non-words (e.g. vap, osk). The reason that non-words are included is that pupils of

this age should have the skills to decode and read almost any unfamiliar word purely based on the rules for combining letters and sounds, even if they have never seen the word before.

Pupils are scored against a national standard ('pass threshold'), meaning they will be either below, at, or above this standard. This threshold has been set at 32, meaning that pupils had to read at least 32 words out of 40 correctly. Pupils who do not meet this threshold must take the check again at the end of Year 2. The results for individual schools are not published, but parents are informed about their child's result on the phonics check.

### Key Stage 1 SATs – Year 2 (age 6-7)

There are two compulsory SATs in Year 2:

- **Mathematics** (Paper 1: arithmetic test – approximately 20 minutes – worth 25 marks; Paper 2: reasoning test – approximately 35 minutes – worth 35 marks)
- **Reading** (Paper 1: reading texts of around 200-300 words with questions throughout – approximately 30 minutes – worth 20 marks; Paper 2: reading different texts of around 500-600 words and answering questions – approximately 40 minutes – worth 20 marks)

In addition, there is a grammar, punctuation and spelling test available to schools on an optional basis (it was compulsory until 2017). This 25-minute test is made up of two papers and is worth a total of 40 marks. Pupils are also assessed by their teacher on science, writing, and speaking and listening. However, these 'teacher assessments' are not included within a school's official performance data. Parents are not automatically sent their child's KS1 SATs results by schools, but they are available should they request them.

KS1 SATs are marked by teachers within each school. The grading system takes a pupil's raw score on any of the three tests (reading; mathematics; grammar, punctuation and spelling) and translates this into a 'scaled score' between 85 and 115. A score of 100 denotes the 'expected standard'. A score between 101 and 115 shows that the pupil has exceeded the expected standard, whereas a score between 85 and 99 shows they have not met the expected standard. Meanwhile, a separate grading system known as P-Scales is used to describe the level of performance for pupils with Special Educational Needs (SEN).

Currently, a key measure of how well a primary school is serving its pupils is the progress they make between their KS1 SATs (Year 2) and KS2 SATs (Year 6). As noted earlier, the new RBA will change this because it creates a new starting point to measure the progress that pupils make during their time at primary school. As a result, the DfE announced in 2017 that KS1 SATs will be non-statutory from 2023 following the introduction of the RBA. This means that schools will be able to choose whether or not to administer KS1 SATs from 2023 onwards.

## Multiplication Tables Check – Year 4 (age 8-9)

As part of the new MTC, all pupils must complete an online on-screen test to assess whether they can fluently recall their multiplication tables up to  $12 \times 12$ . Pupils only have six seconds to answer each of the 25 questions on the basis that the MTC is designed to test their ability to *recall* the answer, not work out the answer as they go along. The MTC typically takes less than five minutes to complete and it is automatically scored.

There will be no expected standard or ‘pass threshold’ for the MTC. The results of the MTC are calculated automatically and will be made available to schools once all their pupils have completed it (scheduled for June of each academic year) and schools can share pupils’ scores with parents as well. The DfE will publish data at an aggregate level to monitor national standards over time.

## Key Stage 2 SATs – Year 6 (age 10-11)

There are three compulsory SATs in Year 6:

- **Reading** (reading three texts of around 600-800 words and answering questions on them – 60 minutes – worth 50 marks)
- **Grammar, punctuation and spelling** (Paper 1: a range of questions covering grammar and punctuation – 45 minutes – worth 50 marks; Paper 2: an oral spelling test with 20 words read out by the test administrator – approximately 15 minutes – 20 marks)
- **Mathematics** (Paper 1: arithmetic test – 30 minutes – worth 40 marks; Papers 2 and 3: reasoning and problem-solving tests – 40 minutes each – worth 35 marks each)

Once the SATs are complete, they are sent away for external marking. Pupils receive their results towards the end of the summer term in Year 6. Parents are also notified of the results in terms of the same raw scores and ‘scaled scores’ described earlier for KS1 SATs, along with a comment on whether their child is working above or below the expected standard.

Alongside the three SATs, teacher assessment is used for judging pupils’ writing and science at the end of KS2. The assessments made by teachers are supposed to be based on a broad range of evidence regarding how a pupil has performed in a variety of contexts. The teachers’ judgements are made using a set of ‘frameworks’ that describe the standards a pupil must be assessed against. For English, pupils are determined to be either ‘working towards the expected standard’, ‘working at the expected standard’ or ‘working at greater depth’, whereas for science pupils can only be judged as ‘working at the expected standard’. In addition, ‘pre-key stage standards’ are used for pupils who are working below the overall standard of National Curriculum assessments and P-scales are again used for pupils with SEN.

Although the teacher assessments are not sent for external marking, they are required to undergo 'external moderation' to ensure teachers' judgements are accurate and consistent with national standards. To this end, 25 per cent of schools are subject to moderation by local authorities every year on a sample of their judgements on pupils' writing.

## Which results are published for each school?

As of December 2019, the main measures for primary schools that appear in performance tables included a mixture of 'progress' and 'attainment' measures.

### Progress

- Pupils' average progress from KS1 to KS2 in **reading**
- Pupils' average progress from KS1 to KS2 in **writing**
- Pupils' average progress from KS1 to KS2 in **mathematics**

These progress measures, introduced in 2016, are a type of 'value added' measure, which means that pupils' results are compared to the achievements of other pupils nationally with similar prior attainment (i.e. similar SATs results) at KS1. A progress score of 0 means pupils in a given school do, on average, about as well at KS2 as those pupils with similar prior attainment at KS1 in other schools in England. A positive progress score (e.g. +0.3) means pupils in a school on average do better at KS2 than pupils with similar prior attainment at KS1 nationally, whereas a negative score (e.g. -0.4) means pupils in a school on average do worse at KS2 than those with similar prior attainment.

### Attainment

- The percentage of pupils achieving the 'expected standard' in **reading, writing and mathematics** at the end of KS2
- The percentage of pupils who achieve at a 'higher standard' in **reading, writing and mathematics** at the end of KS2
- Pupils' average scaled score in **reading** at the end of KS2
- Pupils' average scaled score in **mathematics** at the end of KS2

Alongside an individual school's performance on each of these attainment measures, the DfE also publishes the 'local authority average' and the 'England average' for each measure to make it easier to identify a school's relative position.

### 3. The new ‘reception baseline assessment’

In 2013, the DfE launched a consultation on reforming primary school accountability, which stated their ambition that “all pupils, excepting some of those with particular learning needs, should be secondary ready at age 11”.<sup>20</sup> Even so, they recognised that “some schools have particularly demanding intakes and, in the short term, will find it challenging to reach the ambitious thresholds that we will set”.<sup>21</sup> To reflect this, they expressed their desire to “continue to look at the progress that pupils make, which will acknowledge the performance of schools whose pupils achieve well despite a low starting-point.”<sup>22</sup>

After noting that the current ‘baseline’ for measuring pupil progress were SATs results at KS1, the consultation suggested that “a baseline check early in reception would allow the crucial progress made in reception, year 1 and year 2 to be reflected in the accountability system and would reinforce the importance of early intervention.”<sup>23</sup> In their response to the consultation in 2014, the DfE reported that “many respondents ... supported the principle of schools being accountable for the progress of their pupils – and that progress should be measured from the earliest possible point in school.”<sup>24</sup> They went on to announce that, in future, they would use a ‘reception baseline’ as the starting point from which to measure the progress of pupils.

The DfE’s plan was to build on the assessments already available for children of this age and work with experts to create criteria for the baselines, with the aim of having the new reception baseline assessment (RBA) in place for September 2016. However, a comparability study of three different RBAs used by schools brought the DfE’s plans to a grinding halt in April 2016. It was discovered that the three different baseline assessments were “not sufficiently comparable to create a fair starting point from which to measure pupils’ progress”, meaning that “the results cannot be used as the baseline for progress measures, as it would be inappropriate and unfair to schools.”<sup>25</sup>

The next DfE consultation on primary assessment in 2017 reiterated that they felt “the case for a baseline in reception is strong”,<sup>26</sup> leading them to revive the idea of the RBA. In their response to the consultation, they cited the fact that “a majority of responses demonstrated clear support for moving the baseline assessment to reception”. Respondents such as the NAHT and the Association of School and College Leaders (ASCL) “agreed with the principle that a progress measure should cover as much as possible of the time that a pupil is in a school, including the important years from reception to year 2”.<sup>27</sup> In light of these responses, the DfE set out their intention to develop a new statutory RBA to measure pupil progress along with plans to test and evaluate the RBA before its introduction in the autumn of 2020. Crucially, the DfE chose to work with a single partner to design and deliver the new assessment rather than having competing assessments.

The National Foundation for Educational Research (NFER) was subsequently contracted by the STA in early 2018 to develop and deliver the RBA. The development process included various trials and pilots throughout 2018 and 2019 as well as consultations with early years practitioners and assessment experts along with feedback questionnaires being sent to schools involved in the trials. The pilot study that tested the RBA involved over 7,000 primary schools across England volunteering to take part. This extensive testing led to some changes during the development phase. For example, the original plans for the RBA included a test of how well children can persist with a task (known as ‘self-regulation’) but this was dropped in February 2019 as the initial trials showed the test was taking too long to administer.<sup>28</sup>

On the same day that the DfE confirmed the RBA was going ahead in 2020, they published a report that provided the evidence gathered throughout the development of the RBA. Of the practitioners (mostly reception teachers) who tested the RBA, over 80 per cent rated the suitability of the practical resources as ‘satisfactory’ or above. In addition, 82 per cent rated the mathematics tasks as ‘satisfactory’ or above and 65 per cent of practitioners trialling the literacy, communication and language tasks considered them to be at least ‘satisfactory’.<sup>29</sup> Practitioners were also asked about the children’s engagement with the tasks. 84 per cent rated the children’s interest and enjoyment of the tasks as at least ‘satisfactory’, and 89 per cent said that the children’s understanding of the tasks was ‘satisfactory’ or better, with the practitioners making comments such as:

- *“The children understood most activities. They enjoyed playing the games.”*
- *“The practical resources were effective in supporting the questions.”*
- *“Tasks were generally pitched appropriately and gave useful insights.”<sup>30</sup>*

Alongside these findings, Schools Minister Nick Gibb was keen to emphasise the value of the RBA, saying it was “hugely important that we understand how much progress primary schools help their pupils make” and the RBA would provide a “better understanding of a child’s starting point when they arrive at school”.<sup>31</sup>

## Questions over the value of the baseline assessment

The DfE stated in their 2017 consultation response that “the prime focus of the new assessment will be on skills which can be reliably assessed and which correlate with attainment in English and mathematics at the end of key stage 2, most notably early literacy and numeracy.”<sup>32</sup> This was later supported by the NFER’s report on the validity of the RBA, which cited numerous research studies that support the assessment of language, communication, literacy and mathematics in reception-aged pupils.<sup>33</sup>

Just a few months after the NFER had been chosen by the STA to design and deliver the RBA, a report from a panel assembled by the British Educational Research Association claimed that the proposed RBA would be flawed as a measure of pupil progress. For example, the time lag between the RBA at age 4 and KS2 tests at age 11 raises the question of how useful the results of the new progress measure will be for parents, schools and the DfE when they arrive. By definition, the RBA will lag behind KS2 SATs results by around six to seven years. As the report from the British Educational Research Association recognised, “any data used to inform a parent’s choice of school must be extrapolated from the results of a cohort of students who entered the schools six or so years prior to the current year of entry.”<sup>34</sup> The panel added that they “know of no other assessment internationally... in which the data lie dormant for such a substantial period.”<sup>35</sup> It is therefore unclear how informative the new RBA-to-KS2 progress measure will be, seeing as numerous aspects of a primary school could change dramatically in between pupils starting and finishing primary school (e.g. a new head teacher and / or school leadership team). As a result, presenting this new progress measure to parents as an indication of the level of progress that their child is likely to make at a school *now* (as opposed to several years previously) is a questionable proposition.

Another issue that could affect the RBA is pupil mobility between schools. The RBA is designed to capture pupil progress over the course of seven years, but 20 per cent of primary pupils in England move school at ‘non-standard’ times of year.<sup>36</sup> The British Educational Research Association panel also noted that mobility rates vary by region and by demographic characteristics such as socioeconomic background and ethnicity.<sup>37</sup> The panel commented that “either mobile pupils will have to be taken out of the progress measure in all schools, resulting in varying numbers of children being ‘missing’ from the accountability measure; or baseline assessment results will ‘follow’ pupils between schools, resulting in schools being held accountable for pupils’ progress despite being unaware of their starting points, and having been responsible for only part of pupils’ school lives.”<sup>38</sup> In fairness, pupil mobility already affects the existing accountability regime due to the gap between KS1 and KS2 statutory assessments, but the proposed seven-year gap between the RBA and KS2 SATs will almost certainly exacerbate the issue.

Moreover, the panel raised several concerns about the accuracy and fairness of the RBA. For example, “autumn-born children have demonstrated a strong advantage in attainment over their younger, summer-born peers in assessments similar to the one proposed”, meaning that “schools serving more children who are young for their year of entry may appear to have less favourable effects on children’s later attainment.”<sup>39</sup> This variable could be particularly influential in schools that have a small cohort of pupils in each year, where individual pupils may have a disproportionately large impact on a school’s overall progress score. As it stands, the RBA does not take pupils’ precise age into account, which could distort the validity of the results in terms of measuring progress.

The different structures of primary schools pose a further challenge to the value of the RBA as part of a progress measure. The blend of infant, first, junior, middle and all-through primary schools demonstrates how problematic a single measure of pupil progress from reception to Year 6 could become. Even if all schools with a reception year deliver the RBA as planned, many of them will not deliver any other formal assessments before pupils move to the next phase of their primary education. In fairness, this is not a new problem. The DfE's consultation response in 2017 had pointed out that "the current accountability system does not work as well as it could for some school types, particularly middle schools, where key stage 2 assessments take place part-way through the pupils' time in their school and where there is no statutory assessment at the point when their pupils start and leave the school."<sup>40</sup> Unfortunately, when respondents were asked to choose between different ways of solving this issue under the new RBA, there was "no clear consensus on this issue [as] similar proportions of respondents favoured holding schools to account for the particular years that they cover as said that infant and junior schools should be held to account together."<sup>41</sup>

It was also noted that "many respondents emphasised that the key principles of the school accountability system should apply to all school types; for example, that accountability should be transparent, fair and include clear incentives to ensure that all pupils make good progress."<sup>42</sup> Even so, the decision to make KS1 SATs non-statutory in 2023 means that many primary schools that do not teach Year 6 will no longer produce *any* attainment or progress data. Meanwhile, middle and junior schools that generally admit pupils at age 6 or 7 will continue to administer KS2 SATs in Year 6 to measure pupils' attainment but they will also have no progress measures in place once KS1 SATs are withdrawn. The DfE has stated they will work with sector representatives and Ofsted before providing further guidance about the type of information schools could make available ahead of the progress measures being removed for all these primary schools.<sup>32</sup> As it stands, it is unclear how such information will work in practice.

## The high stakes attached to the RBA

When the NFER responded to the DfE's 2017 consultation on primary assessment, they highlighted an important risk generated by the progress measure in the current accountability system: "introducing the progress threshold at KS2 has raised the stakes of the KS1 assessments, creating incentives for teachers to game the system, deflating results to demonstrate greater progress at the end of KS2."<sup>43</sup> They commented that using external marking for KS1 SATs – as they do for KS2 – could solve this problem, but it would "raise the stakes of the tests for both pupils and schools with potential adverse impacts on teaching and learning".<sup>44</sup>

When the Education Select Committee in Parliament investigated primary assessment in 2017, they made a similar observation. Although they “[welcomed] the increased focus on progress in performance measures and the Government’s commitment to introduce an improved baseline measure”,<sup>45</sup> they expressed concern that “the Government has not outlined how it will ensure that any new baseline measure will not be subject to the same ‘gaming’ that Key Stage 1 results were.”<sup>46</sup> In their written evidence to the Committee, Education Datalab – a group of academics and statisticians – reported that “the replacement of the Key Stage 1 externally marked test with teacher assessment in 2003 led to primary schools depressing their scores, knowing it would be used as a baseline for Key Stage Two value-added measures.”<sup>47</sup> Professor Rob Coe, Director of the Centre for Evaluation and Monitoring at Durham University, told the Committee that “the incentive is to want to do badly on a baseline if it is progress you are going to be measured on.”<sup>48</sup> This incentive has essentially been generated by the DfE’s desire to use an assessment of pupils (the RBA) as the basis for an assessment of schools (the progress measure). As will be discussed in the next chapter on SATs, using scores from a single assessment for multiple purposes runs the risk that the assessment does not end up serving any purpose particularly well.

To their credit, DfE officials have already recognised another possible unintended consequence of the RBA: if schools were given the raw scores achieved by each pupil, there would be “the potential for streaming or labelling of the children based on their scores”<sup>49</sup> i.e. a school might label a child as a ‘high performer’ or ‘low performer’ based on their performance in the RBA, and this label would then stick with them for the next seven years. To counteract this, the DfE has chosen not to share individual pupil scores with schools to mitigate the risk that they might attempt to artificially boost their progress measure outcomes (although schools will still receive a narrative statement that describes each pupil’s performance on the RBA).

## The impact of the RBA on pupils and teachers

Despite the findings from the NFER that many teachers were content with the baseline assessment during their pilot study, a report commissioned last year by the National Education Union (NEU) and produced by researchers at the UCL Institute of Education found that only 20 per cent of teachers who responded to their survey believed that an RBA provided an accurate picture of children’s current attainment and 84 per cent stated that an RBA was an unreliable or inaccurate way to measure children’s progress over seven years of primary school.<sup>50</sup> In addition, 83 per cent of respondents said their workload had increased due to the RBA and some teachers also reported staffing constraints and pressures in completing the RBA (e.g. having to bring in supply teachers to cover lessons).<sup>51</sup>

Such findings do not appear to undermine the validity of the RBA as designed by the NFER for reception pupils, as they relate more closely to the way that the RBA will be used as a progress measure rather than the RBA itself. Furthermore, the DfE has acknowledged that the RBA will require more time and effort in its early stages as teachers and leaders become familiar with it. The STA will also conduct 'quality monitoring visits' to schools to check whether they are delivering the RBA in line with government guidance.<sup>52</sup>

The NAHT remains supportive of the RBA on the basis that it will replace KS1 SATs (which are far more time-consuming) as one of the main measures of primary school performance. An NAHT spokeswoman told *The Guardian* in 2019 that "a measure of progress from the start to the end of primary school is preferable and actually places a greater value on the important work and progress made in those critical first few years." She added that "our members have told us that having this one-to-one time with each child is valuable for staff and that it is enjoyed and valued by the children, too."<sup>53</sup> A local branch secretary of the NAHT also commented that when they contacted 300 members for feedback on the RBA, "nobody mentioned the children getting upset [as] the feedback was mainly around the cost of providing cover for the adults conducting the assessments, the narrowness of the questions and, in rural areas, problems accessing the internet."<sup>54</sup> Clearly, there is a diversity of views regarding the value of the RBA as a starting point for measuring pupil progress in future.

## 4. The impact of SATs on teachers and schools

The Children, Schools and Families Committee in Parliament (a previous incarnation of the Education Select Committee) investigated the testing and assessment system in 2008. The then Department for Education and Skills (DfES) told the Committee that, until the introduction of end-of-Key-Stage tests, there were no “objective and consistent performance measures which gave the public confidence about expected standards in primary schools”.<sup>55</sup> Mick Brookes, then General Secretary of the NAHT, said that “nobody in our association wants to return to the 1970s when you did not know what the school up the road was doing, let alone a school at the other end of the country.”<sup>56</sup> The Committee underlined the fact that “educators accept that accountability of schools is a necessary feature of a modern education system and that national testing has an important part to play”,<sup>57</sup> adding that:

*“We consider that the weight of evidence in favour of the need for a system of national testing is persuasive and we are content that the principle of national testing is sound. Appropriate testing can help to ensure that teachers focus on achievement and often that has meant excellent teaching, which is very welcome.”*<sup>58</sup>

Even so, the Committee noted that “national tests are used for a wide variety of purposes at a number of different levels: national, local, school and individual.”<sup>59</sup> Dr Ken Boston, then head of the Qualifications and Curriculum Authority (part of the DfES at the time) agreed that problems quickly arise “when, having achieved a test which is fit for one purpose, it is then used for other purposes.”<sup>60</sup> When it was put to Dr Boston that the Key Stage tests were being used for 14 different purposes, he responded by saying that 14 purposes was “stretching it too far [...because] when you put all of these functions on one test, there is the risk that you do not perform any of those functions as perfectly as you might.”<sup>61</sup> The Committee also received evidence from David Hargreaves, Chris Gerry and Tim Oates who argued that a “one test instrument cannot serve all the Government’s stated purposes of testing because they conflict to a certain extent.”<sup>62</sup> The Committee’s final report illustrated how the use of single assessments for multiple purposes was a cause for concern:

*“The conflicts between the different purposes are not, perhaps, inherent, but arise because of the manner in which people change their behaviour when high-stakes are attached to the outcomes of the tests. Many other [witnesses] have raised similar points, claiming that two purposes in particular, school accountability on the one hand and promoting learning and pupil progress on the other, are often incompatible within the present testing system.”*<sup>63</sup>

This conflict caused by using test outcomes to simultaneously judge pupils and schools was a theme that ran throughout the Committee's report, leading them to devote a considerable portion of their inquiry to understanding the consequences of a high-stakes testing system:

*"Teachers and headteachers need their pupils to do well in tests in order to demonstrate that they are effective teachers, to win promotion and, perhaps, financial reward. Perceived failure of teachers in this respect may lead to demoralisation, being passed over for promotion, difficulty in finding employment in other schools or leaving the profession altogether. Schools are held accountable according to the test results achieved by their pupils and a poor set of results may result in public humiliation in published performance tables, being perceived as a "failing school", interventions by Ofsted and even closure in extreme cases."*<sup>64</sup>

More specifically, there were two effects of SATs that attracted the Committee's attention: 'teaching to the test', and the narrowing of the curriculum.

### **'Teaching to the test'**

The Committee defined 'teaching to the test' as follows:

*"...teachers drilling their pupils in a subject on which they will face a test or examination. In extreme cases, a high proportion of teaching time will be given over to test preparation. The focus of lessons will be narrow, with teachers coaching their pupils on examination technique, question spotting, going over sample questions similar to those likely to be set in the test and generally focussing teaching of the substance of a subject in a way best calculated to maximise marks in the test."*<sup>65</sup>

The DfES felt it was acceptable for children to have an 'appreciation' of how to present their answers in a test but this should nevertheless be "wholly integrated into the classroom experience" and "there should be no cramming ...as the Key Stage tests approach".<sup>66</sup> They also stated that "the teacher who prepares pupils for a test without developing the deeper understanding or more extended experience required may be fortunate enough to enjoy some short-term success, but will not be likely to maintain that performance over time."<sup>67</sup> The Committee retorted that the Government seemed "rather out of touch with what appears to be happening in classrooms", adding that "it has been argued by a great number of witnesses to this inquiry that the high-stakes attached to national testing increases pressure to teach to the test."<sup>68</sup>

ASCL told the Committee that "teachers have been criticised for teaching to the test but, if the system is geared to constantly monitoring progress and judging teachers and institutions by outcomes, it is hardly surprising that the focus is on ensuring that students produce the best

results.”<sup>69</sup> Moreover, the NAHT referred to a survey which indicated that, in Year 6, for four months of the school year teachers were spending nearly half their teaching time preparing pupils for KS2 tests.<sup>70</sup> The Committee’s overall conclusion left little room for doubt about their views on this matter:

*“We received substantial evidence that teaching to the test ...is widespread. [Ministers] ...fail to accept the extent to which teaching to the test exists and the damage it can do to a child’s learning. We have no doubt that teachers generally have the very best intentions in terms of providing the best education they can for their pupils. However, the way that many teachers have responded to the Government’s approach to accountability has meant that test results are pursued at the expense of a rounded education for children.”<sup>71</sup>*

Not long after the Committee published their report, the Coalition Government released a White Paper titled ‘The Importance of Teaching’ in 2010, which made the following observation:

*“...we are concerned that especially in year six, there is excessive test preparation – with some children practising test questions for many weeks in advance of the tests. This is poor practice, and not even an effective way to do well in the tests, as compared to teaching well across a broad curriculum for the same period.”<sup>72</sup>*

Almost a decade later, the Education Select Committee heard evidence that ‘teaching to the test’ could distort the results of Key Stage tests to the point where they would “become severely inflated in being far larger than true gains in students’ learning”.<sup>73</sup> The Committee also cited concerns about “the pressure that schools are under to achieve results at Key Stage 2” and that “many teachers reported ‘teaching to the test’ ...as a result of statutory assessment and accountability.”<sup>74</sup> Evidently, little progress has been made in addressing this issue, and the pressure to ‘teach to the test’ identified by two separate Select Committee inquiries remains prevalent to this day.

## **Narrowing of the curriculum**

The 2008 Children, Schools and Families Committee recognised that ‘narrowing of the curriculum’ is “strongly related to teaching to the test”, and has two elements:

*First, there is evidence that the overall curriculum is narrowed so that the majority of time and resources is directed at those subjects which will be tested and other subjects in the broader curriculum, such as sport, art and music, are neglected. Second, within those subjects which are tested, the taught curriculum is narrowed to focus on those areas which are most likely to be tested (‘narrow learning’) and on the manner in which a component of the curriculum is likely to be tested (‘shallow learning’).”<sup>75</sup>*

A survey submitted to the Committee found that in 90 per cent of primary schools, national testing had led to pupils being offered a narrower curriculum.<sup>76</sup> In response to these findings, the Government said it makes “no apology for the focus on the core subjects of English, maths and science” as they believed that mastery of these disciplines is the key to future success. The Committee noted that the evidence they received suggested that “mastery of the examination is given priority over mastery of the subject and that time taken to prepare for these examinations is taken from the broader curriculum”.<sup>77</sup>

When the Education Select Committee returned to this topic in 2017, they cited evidence from Ofsted who had recently found that most primary schools were spending four hours or more a week teaching English and maths, yet around two thirds of schools spent only one to two hours per week teaching science, and around a fifth spent less than one hour.<sup>78</sup> Shortly after Amanda Spielman was appointed as the new Chief Inspector at Ofsted in 2017, she expressed her concern that the focus on tests such as SATs was coming at the expense of “rich and full knowledge” in a wide range of subjects.<sup>79</sup> The following year, Ms Spielman accepted that Ofsted inspectors had previously put too much weight on test results when grading schools, which added to the pressure to “deliver test scores above all else”.<sup>80</sup> Nevertheless, Ofsted inspectors had seen “curriculum narrowing, especially in upper Key Stage 2, with lessons disproportionately focused on English and mathematics” and that “sometimes, this manifested as intensive, even obsessive, test preparation for Key Stage 2 SATs that in some cases started at Christmas in Year 6.”<sup>81</sup>

At the launch of Ofsted’s new inspection framework in 2019, Ms Spielman gave an example of one primary school surveyed by Ofsted where children had practised SATs test papers in English and maths every week for the past two years, while at other primary schools practice papers were taken from the age of eight.<sup>82</sup> In response to such cases, Ofsted’s current inspection framework was “intended to restore curriculum – largely ‘missing in action’ from inspection for more than a decade – to its proper place, as an important component of the quality of education.”<sup>83</sup>

## The role of performance tables

Performance tables are published annually to show pupil attainment and progress in different primary schools. The 2008 Children, Schools and Families Committee noted that tables “allow for instant comparison of a school’s results against both the local authority average and the national average.”<sup>84</sup> The Government told the 2008 Committee that performance tables are “an important source of public accountability for schools and colleges”,<sup>85</sup> with a minister arguing that if such data was not made publicly available there would be “an outcry that we were hiding things”.<sup>86</sup>

A government-commissioned review of primary testing by Lord Bew ('the Bew Review') in 2011 was adamant that "external school-level accountability is important in driving up attainment and pupils' progress" but "much of the evidence and feedback we have received shows that the main concern lies with the way in which school accountability data is used".<sup>87</sup> The 2017 Education Select Committee inquiry into primary assessment made a similar observation regarding how SATs results are used by government:

*"Many of the negative effects of assessment are in fact caused by the use of results in the accountability system rather than the assessment system itself. Key Stage 2 results are used to hold schools to account at a system level, to parents, by Ofsted, and results are linked to teachers' pay and performance. We recognise the importance of holding schools to account but this high-stakes system does not improve teaching and learning at primary school alone."*<sup>88</sup>

By publishing performance tables for all primary schools in England that include SATs results, school leaders and teachers will understandably feel compelled to concentrate on these assessments. What's more, SATs results are updated on an annual basis, meaning that primary schools have less scope for building new curricula or trying new approaches to teaching and learning in case their published results do not show an immediate return. The NAHT expressed several concerns to the 2008 Children, Schools and Families Committee, stating that "there are very good schools that work against the odds to produce higher educational qualifications than they have ever had in their areas, but they are disabused of that excellent work on an annual basis."<sup>89</sup> In retrospect, the conclusion reached by the 2008 Committee on the culture of performance tables was prescient:

*"The use of performance tables for school accountability means that a school's standing in the performance tables is a matter of significant importance to that school, directly or indirectly affecting the morale of pupils and teachers; the attitudes of parents; the school's standing in the local community and within the wider local authority; the resources allocated to it; and perhaps even the school's very survival. ...Witnesses have commented that the use of performance tables as accountability measures ...has encouraged a "risk-averse culture."*<sup>90</sup>

In terms of solutions, the Committee wanted to see schools given credit for a wider range of activities beyond test results by encouraging SAT results to be read in conjunction with a school's Ofsted report. This chimes with recent efforts by Amanda Spielman to ensure that Ofsted focuses its inspections on the quality of a school's curriculum to a greater extent than in the past. More recently, the 2017 Education Select Committee called for a change to the way that SATs results are published "to help lower the stakes associated with them" – namely, by "publishing a rolling three year average of Key Stage 2 results instead of results from a single

cohort” (albeit with cohort level data still available to schools for their own internal monitoring).<sup>91</sup> Likewise, the Bew Review called for “the introduction of three-year rolling averages ...which we believe would take into account the volatility of results of individual cohorts and provide a sense of achievement over time.”, although this was intended to complement rather than replace annual test data.<sup>92</sup>

The above proposals have merit within discussions over the future of primary assessment, as it is essential that the publicly available data about each primary school is a fair and reasonable reflection of the contribution they have made to their pupils’ education. That said, almost every committee and review that has investigated the role of performance tables in recent years expressed serious misgivings about whether they do indeed offer a fair reflection of what a school has contributed, particularly when primary schools operate in such different contexts and sometimes with very different cohorts of pupils. This is compounded by the continued use of SATs results for a wide variety of purposes (measuring pupils, schools and national performance levels, to name but three) even though this problem was highlighted well over a decade ago by the 2008 Children, Schools and Families Committee. Thus, finding more accurate and reliable ways to measure pupil progress as well as being more discerning about how test results are used must both be prioritised within any new assessment and accountability regime.

## 5. The impact of SATs on pupils

Critics of SATs have frequently cited their supposed negative effects on pupils. Even before the Children, Schools and Families Committee produced their major report on testing and assessment in 2008, the Cambridge Primary Review (CPR) – an independent inquiry into the future of primary education in England – had found evidence that pupils were indeed affected by SATs to some degree. When researchers spoke to pupils, it was clear that “for pupils in years 2 and 6 the notion of SATs looms large in pupils’ minds.”<sup>93</sup> The researchers noticed that “some pupils feel that their learning is almost entirely focused on achieving good grades in SATs” and that they “find both the tests, and the preparation for them difficult”.<sup>94</sup>

The CPR cited previous studies that had shown “pupils felt a sense of unease about what SATs might reveal about themselves as learners, with some pupils indicating far-reaching consequences in which good SATs results are linked to positive life prospects and poor results linked to future failure and hardship.”<sup>95</sup> Given the structure of the accountability system for primary schools, it was unsurprising that “the pressure on pupils at Key Stage 2 tends to be far greater ...than at Key Stage 1” and “overall, children seem only too aware that whilst ‘trying’ is worthy, ‘achieving’ is actually the required outcome.”<sup>96</sup> The researchers also discovered that “the majority of children are aware that SATs results constitute some sort of ‘official’ judgement of them” and “some pupils are also aware of teachers’ own sense of pressure from SATs, as the results are also used in assessing teachers.”<sup>97</sup>

Likewise, the evidence heard by the Children, Schools and Families Committee in 2008 appeared to be leaning in one direction when it came to the impact of testing:

*“Many witnesses argued that testing is stressful for children. Moreover, repeated testing has a negative effect on children, leading to demotivation, reduced learning potential and lower educational outcomes. Testing has even been linked with children’s health, including mental health, problems and lower self-esteem. ...Witnesses have expressed concern that the levels of accountability in schools are resulting in the disillusionment of children. Children not reaching the target standard at a given stage have the impression that they have ‘failed’ whilst they may, in fact, have made perfectly acceptable progress. Whilst some children undoubtedly find tests interesting, challenging and even enjoyable, others do not do their best under test conditions and become very distressed.”<sup>98</sup>*

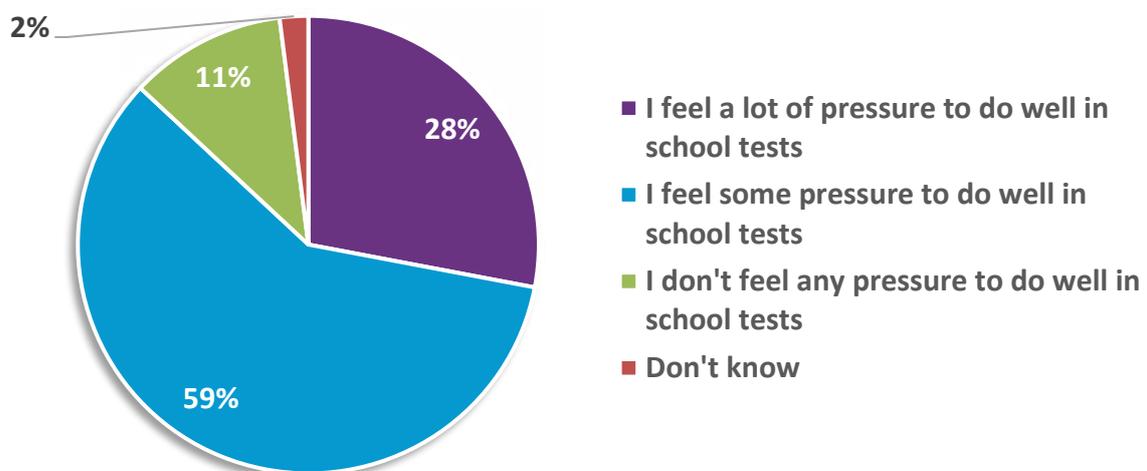
Rather than relying on witnesses, the 2017 Education Select Committee inquiry into primary assessment spoke directly to pupils instead. The Committee held an informal event with Year 6 pupils and teachers from West Denton Primary School during their visit to Parliament to

understand how primary assessment affects pupils and how they felt about their upcoming SATs tests. The full summary of the event was as follows:

- *“In general, the pupils were positive about taking SATs. They felt that SATs were a good opportunity to demonstrate what you knew and getting good SATs results would help you later in life.*
- *“Some pupils told us they could get nervous or anxious about taking the tests, and that feeling nervous during the test might affect how well they did.*
- *“Some pupils pointed out that focusing on tests meant that they didn’t have as much time to do other things, like art or PE. Some suggested that there should be more tests in these subjects to make things fairer.*
- *“The children suggested some improvements to the SATs. Many pupils thought they should be given more time for each test, and that it was time pressure that made them most nervous. Others thought that the writing assessment was unfair. In particular, they felt that pupils who struggled with spelling or handwriting were unfairly penalised in the writing assessment, even if their creativity or composition was of a high standard.”<sup>99</sup>*

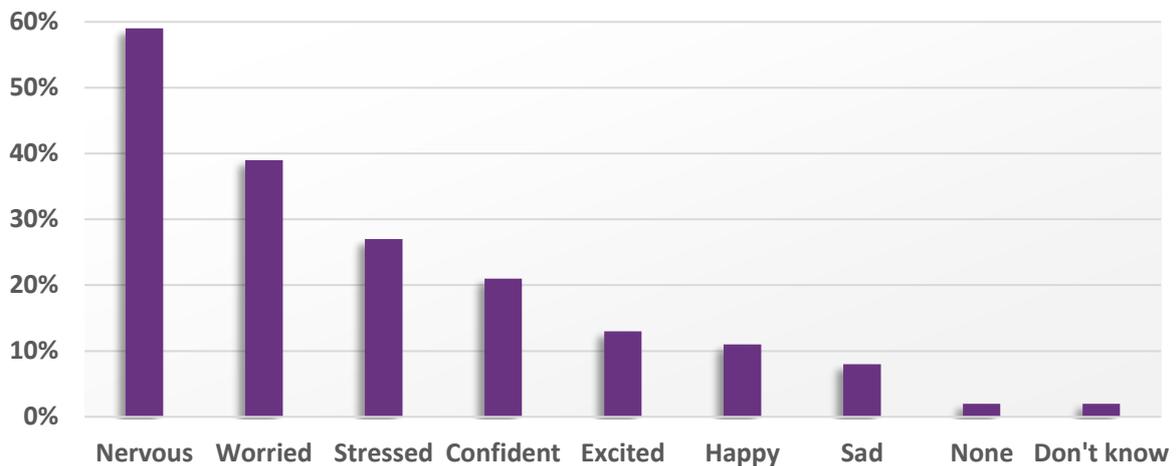
A survey by the polling firm ComRes in 2016 of 750 10 and 11-year-olds in English schools has provided one of the most comprehensive datasets on what pupils really think about testing in primary schools.<sup>100</sup> When ComRes asked pupils how they felt about school tests, almost 60 per cent said they feel ‘some pressure’ to do well, with 28 per cent saying they felt ‘a lot’ of pressure and 11 per cent not feeling any pressure at all (Figure 1).

**Figure 1: “Which of these options best describes how you feel about school tests?” (ComRes, 2016)**



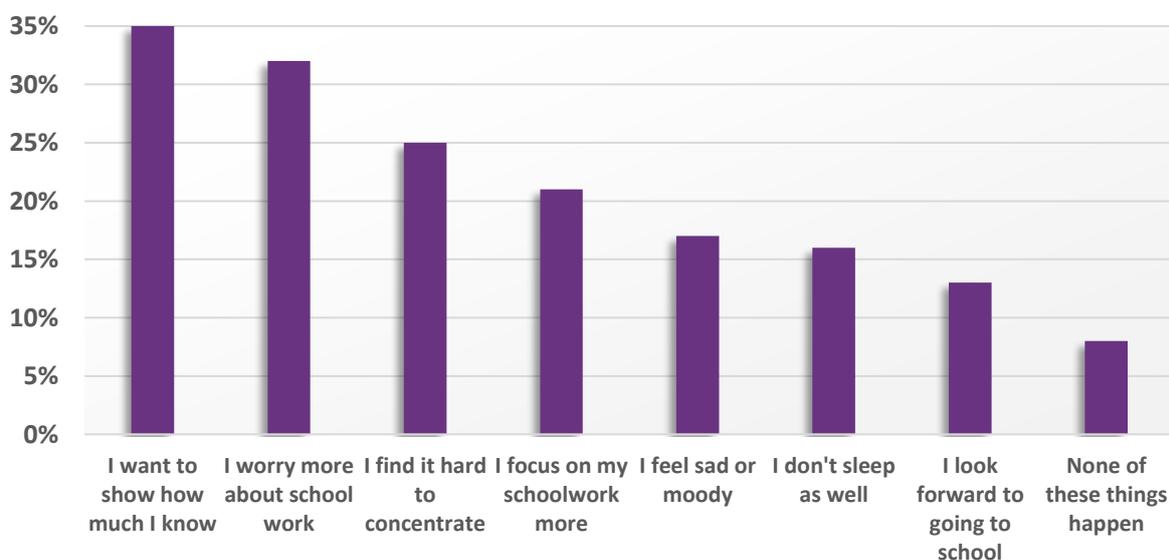
Pupils were also asked how they felt when taking tests at school and were given a list of words to choose from (Figure 2). The most common choices were ‘nervous’ (59 per cent), ‘worried’ (39 per cent) and ‘stressed’ (27 per cent). Even so, a notable (albeit smaller) proportion of pupils reported positive emotions such as being ‘confident’, ‘excited’ and ‘happy’.

**Figure 2: “How do you feel when you take tests at school?” (ComRes, 2016)**



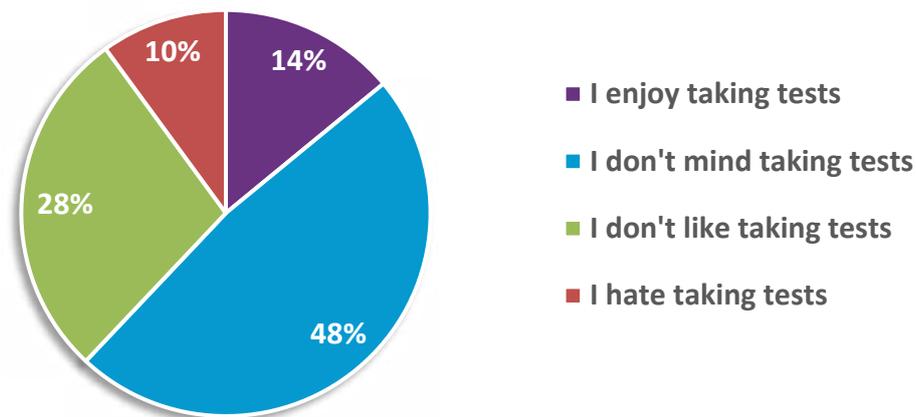
When ComRes asked pupils about their reactions to tests at school, the most common response was for pupils to report that “I want to show how much I know” (35 per cent) and a small proportion of pupils even said they ‘look forward to going to school’ for tests. On the other hand, almost a third pupils said that tests made them ‘worry more about school work’ (32 per cent) and a quarter said that tests made it hard for them to concentrate (Figure 3).

**Figure 3: “When you have tests at school, do any of these things happen?” (ComRes, 2016)**



While it could be argued that these initial findings about the impact of SATs are concerning, other questions posed in the same survey suggested that the situation was more nuanced than it might appear. When pupils were asked what they think about tests at school, 62 per cent said they either ‘enjoy’ them or don’t mind taking them (Figure 4), despite these same pupils saying in response to other questions that the tests can make them nervous or worried. This indicates that primary school pupils, like many adults, can understand the value and importance of tests even if they do not always relish them.

**Figure 4: “What do you think about tests at school?” (ComRes, 2016)**



A separate study published this year that compared data on 2,500 children living in England - where pupils sit the tests - and Northern Ireland, Scotland and Wales - where they do not – also found little difference in wellbeing and happiness levels regardless of whether pupils sat the tests, although the findings were based on data going back to 2012.<sup>101</sup>

### The views of parents on primary school assessment

In 2020, the ‘More Than A Score’ campaign published results from a YouGov survey they commissioned of over 2,000 adults with children aged 3 to 13. The findings suggested that there was some discomfort among parents of older pupils regarding the current role of SATs:

- 66 per cent agreed that their child was under pressure to get good marks
- 57 per cent said their child was anxious or nervous about taking SATs
- 61 per cent said their child spent most of their time studying English and maths in preparation for SATs
- Only 29 per cent believed their child was enjoying school at this time
- 41 per cent said SATs had a negative effect on their child’s love of learning <sup>102</sup>

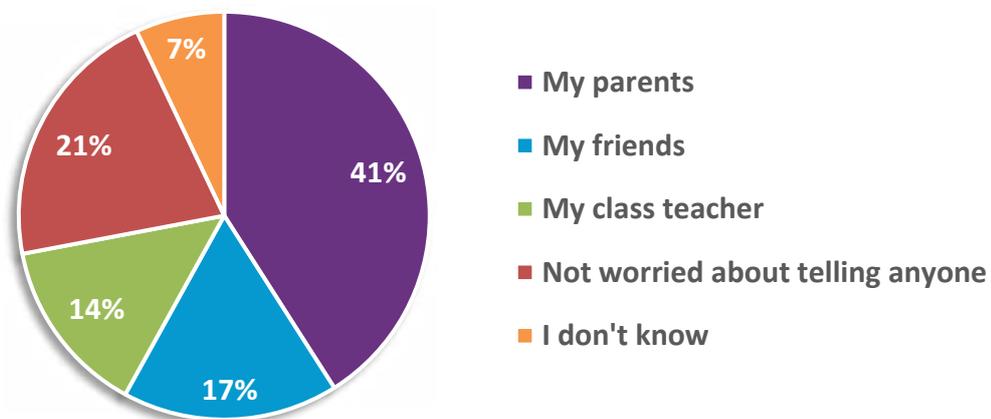
Even before this survey was conducted, Amanda Spielman had previously highlighted a range of practices that had been reported to Ofsted by parents:

*“Generally, primary school parents said that preparing for tests was cutting into their child’s learning time. Around half of the parents who responded to our questionnaire ...believed that test preparation had reduced the teaching time available for the other foundation subjects or for reading for pleasure. Furthermore, a small proportion of parents suggested that, in their child’s school, the focus on past papers, booster sessions and test-related homework was too high.”*<sup>103</sup>

The issue of ‘booster sessions’ has become increasingly prevalent in recent years. It was reported in February 2020 that more primary schools are holding half-term and Easter holiday revision classes for pupils as young as six to prepare them for SATs. This use of holiday ‘booster sessions’ was condemned by the DfE and teaching unions, with one union leader describing them as “an extraordinarily bad idea” with no positive impact.<sup>104</sup> Research by the UCL Institute of Education in 2019 identified the same problem regarding the increasing use of Easter and half-term revision classes as well as ‘additional hours’ on Saturdays from January onwards in some schools. Parents could also be affected by the ‘near universal’ practice of asking children to come to school early during SATs week.<sup>105</sup>

Even if parents are unhappy with the status quo on primary testing, there is some evidence that parents might unintentionally be contributing to the problem. When ComRes asked pupils who they would be most worried about knowing they did badly in a test, ‘my parents’ was named by 41 per cent of 10 and 11-year-olds – a bigger proportion than ‘my friends’ and ‘my class teacher’ combined (Figure 5). This indicates that some parents may be adding to any stress or anxiety that their children are experiencing, and that changes in parental behaviour could therefore help to reduce these feelings among primary school pupils.

**Figure 5: “If you did badly in a test at school, who would you be most worried about knowing the result?” (ComRes, 2016)**



## What is the source of any stress and anxiety among pupils?

The CPR research described earlier in this chapter indicated that pupils were being affected by SATs, yet their research highlighted an important caveat to discussions on this issue: the effect of SATs on individual pupils varied considerably and depended, at least in part, on the actions of teachers and head teachers. For instance, the CPR cited previous research showing that “where schools have created a secure, non-threatening environment, high attainers begin to feel more confident and even exhilarated during the test period. However, under pressure, other pupils become demotivated and dysfunctional as the difficulty of the SATs challenges overwhelm them.”<sup>106</sup> In 2019, Amanda Spielman made a similar observation, as she suggested that primary school pupils should not be told they are taking exams to prevent feelings of anxiety. Although she remains in favour of formal tests in primary schools, the Ofsted Chief Inspector was keen to assert that “good primary schools manage to run key stage tests often with children not even knowing that they're being tested”.<sup>107</sup> She added that, while testing happens in every education system in the world, “it only becomes a big deal for young children if people make it so for them.”<sup>108</sup>

Echoing these sentiments, Sir Steve Lancashire – CEO and founder of the REACh2 academy trusts, one of the biggest groups of primary schools in the country – recently stated that “the reality is that in good schools, children are not stressed [...and] often do not even know they are being tested.”<sup>109</sup> His verdict was that “where children are feeling so distressed, you have to ask what on earth is the school doing to make them feel that way? The first stop for any parent whose child feels this way is a frank discussion with the head.”<sup>110</sup> To illustrate the point, REACh2 conducted a survey of 1,100 pupils from more than 50 of its schools to ascertain their views on being tested. The survey showed that 67 per cent of pupils said they did not mind taking tests while 16 per cent said they enjoyed tests and 16 per cent did not like taking them.<sup>111</sup> Alongside these findings, Sir Steve remarked that “across our schools we’ve done our best to make sure our pupils knew they didn’t have to worry about the SATs, and these results show that this was a success.”<sup>112</sup>

While individual surveys of this nature should not be viewed as definitive evidence, there is good reason to question the notion that SATs themselves are a direct source of stress and anxiety for pupils. It appears that, while tests of this nature do generate some pressure on pupils (as they would do for adults), schools and even parents may be unintentionally contributing to pupils’ concerns. Designing an assessment and accountability system in such a way that it causes undue levels of worry or stress among pupils should be avoided, not least because it may prevent pupils from performing at the level that they are capable of achieving. Thus, the pressure generated by primary school assessments should be proportionate and reasonable within the context of determining the level of progress that pupils are making.

## 6. The phonics check and multiplication check

### The phonics check (2012 onwards)

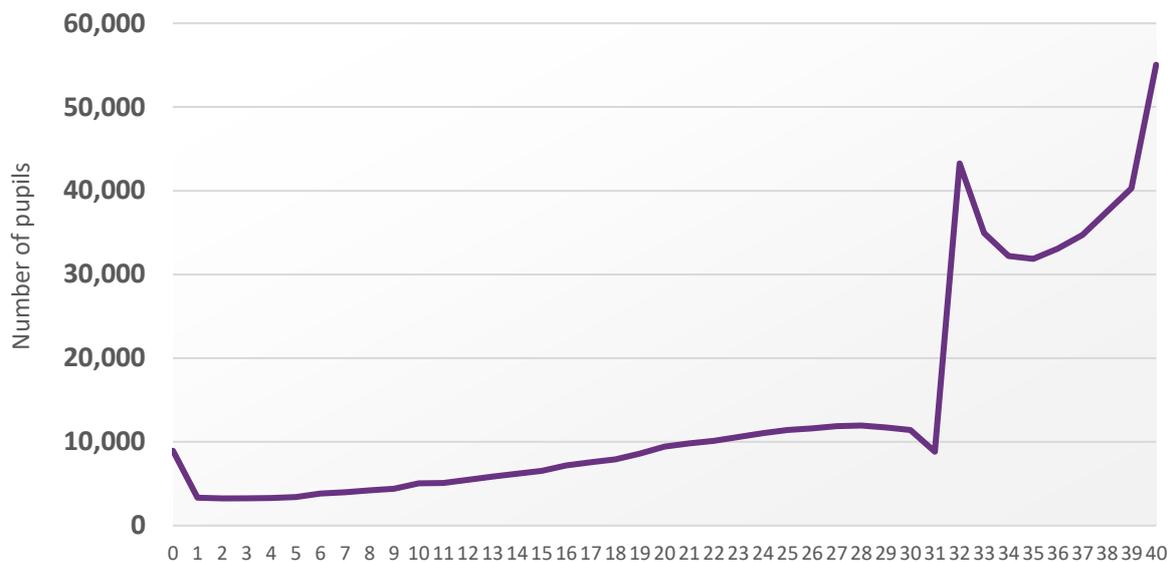
The 'phonics check' tests whether Year 1 pupils can decode (i.e. understand) letters and sounds to an appropriate level, which will allow them to read many short words. The check lasts for about 5-10 minutes, during which time pupils must read a list of 40 words to their teacher (20 real words and 20 non-words). The assessment framework underpinning the phonics check states that it will serve the following five purposes:

- Individual children's results will be made available to parents, so that parents are kept informed about their child's progress in developing word-reading skills;
- Schools can use their results to track their performance over time (although they will not be published in performance tables);
- School-level results are made available to Ofsted to inform their inspections;
- National results will be reported so that the government can track standards over time;
- National and local authority results will also allow schools to benchmark the pupils' performance.<sup>113</sup>

There is an obvious tension between some of these purposes. For example, requiring schools to use the phonics check to monitor pupils' reading skills does not sit comfortably alongside allowing Ofsted to judge the quality of schools based on the results from the same phonics check. Schools will inevitably be concerned about the consequences of low scores from their pupils if Ofsted consults the scores during their inspections, which could undermine the validity (i.e. accuracy and usefulness) of the test results. On a related note, using this same data to judge national standards as well as individual schools will become less valid if schools are incentivised to maximise their pupils' scores rather than reporting them objectively.

Far from being idle concerns, the risks posed by using the phonics check for too many purposes have been amply demonstrated by the results reported since 2012. Pupils are scored against a national standard ('pass threshold'), which has been set at 32 marks ever since the test was introduced. When the check was first delivered in 2012, this threshold of 32 marks was "communicated to schools in advance of the screening check being administered so that schools could immediately put in place extra support for pupils who had not met the required standard."<sup>114</sup> Figure 6 (overleaf) shows the distribution of the phonics test scores in 2012, which reveals – in the words of the DfE – "a spike in the distribution at a score of 32".<sup>115</sup>

**Figure 6: the number of Year 1 pupils (aged 6) in England scoring each possible mark on the first phonics check in 2012**

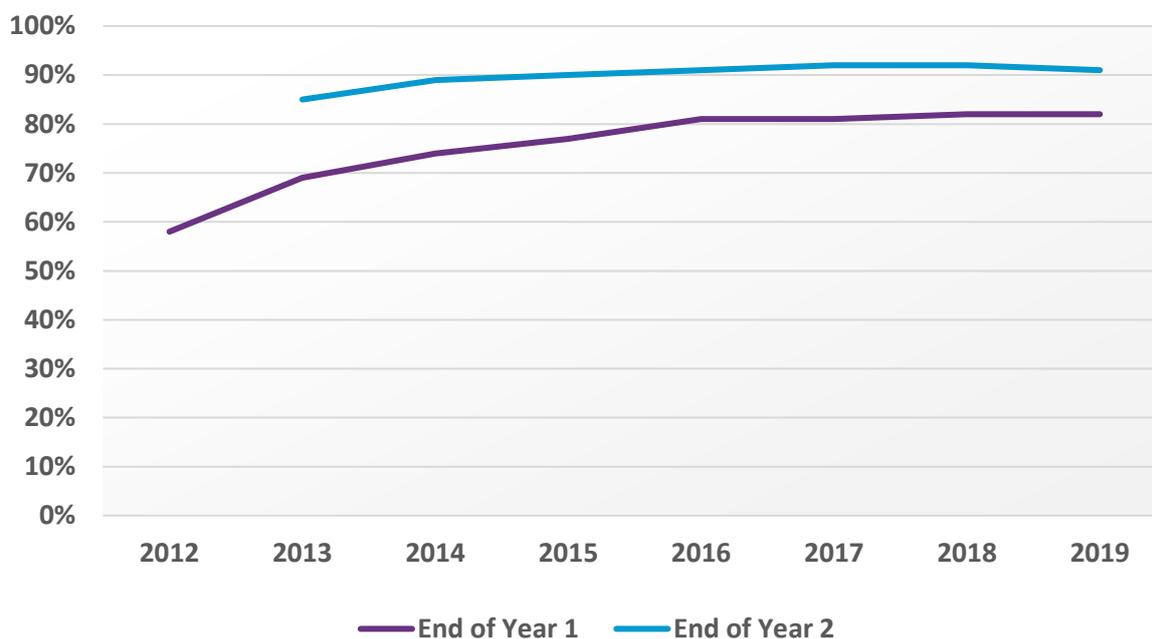


The opening page of the DfE document that described the results of the first phonics check referred to the percentage of children who met the “expected standard” of phonic decoding,<sup>116</sup> which suggests that the pass mark represents the score that the Government was hoping or anticipating pupils might reach. However, on the following page, the DfE referred to the pass mark as the “required standard”,<sup>117</sup> suggesting that schools must ensure their pupils have reached this score as a minimum or potentially face some kind of consequences. This interchangeable use of the phrases ‘expected standard’ and ‘required standard’ continued throughout the DfE document despite them having very different meanings for schools. Even when describing the graph of results shown in Figure 6, the DfE noted that “58 per cent of all pupils eligible for the screening check met the expected standard which was 32 marks out of 40”, yet in the very next sentence they added that “girls outperformed boys with 62 per cent meeting the required standard compared to 54 per cent of boys.”<sup>118</sup> This lack of clarity serves to emphasise how, despite there being no performance tables for schools in terms of their results on the phonics check, the accountability system is almost intentionally placing considerable pressure on schools to ensure their pupils reach a certain standard.

The national results on the phonics check in 2013 showed a similar pattern to those in 2012, but the situation changed in 2014 when the DfE chose not to publish the pass mark before the phonics check was completed (even though it ultimately remained at 32 marks). As can be seen in Figure 7, this meant that the ‘spike’ in results at a score of 32 was less pronounced from 2014 onwards, presumably because schools were now aware that the pass mark could theoretically be moved above or below 32. It is also noteworthy that the number of pupils achieving the highest possible scores has also increased sharply since the first two phonics checks were delivered.



**Figure 8: the proportion of pupils reaching the expected standard in the phonics check by the end of Year 1 and Year 2**



With any new test or assessment, there is likely to be an increase in scores over the first few years of delivery, as those administering the test or assessment become more familiar with it and thus better at preparing pupils for it. This is not necessarily a problem from the DfE’s perspective because a greater focus on the teaching of phonics to help pupils learn to read is essentially what ministers are trying to achieve. Nonetheless, it is uncertain how much of the improvement in scores since 2012 is the result of genuine improvements in pupils’ ability to read as opposed to other factors.

The Children, Schools and Families Committee came to a similar conclusion in 2008, several years before the phonics check even existed. When investigating the dramatic improvements in KS2 SATs scores recorded in the years preceding the Committee’s inquiry, they heard that “an apparent improvement in standards of performance has less to do with an improvement in underlying achievement and more to do with familiarity amongst teachers and students with test requirements.”<sup>122</sup> The Committee’s final report noted that research by Professor Peter Tymms from Durham University “has led him to the conclusion that the substantial improvements suggested by the test scores were illusory”.<sup>123</sup> No research has been conducted into the improvements in phonics check scores at a national level so it is not possible to determine the extent to which the same phenomenon applies in this instance.

## The multiplication check (starting in 2022)

The new MTC, which is being introduced next summer, involves all Year 4 pupils completing an on-screen test to assess whether they can recall multiplication tables up to  $12 \times 12$ . Pupils have six seconds to answer each of the 25 questions. As the MTC has not yet been delivered in schools, there is no research on which to base judgements regarding its validity and effectiveness. That said, the purpose and design of the test are worth exploring at this stage.

The assessment framework for the MTC states that its purpose is almost identical to the phonics check, as the data will be used in the following ways:

- School-level results and individual pupil results will be made available to schools to “allow them to provide additional support to pupils who require it” but will not be published in performance tables;
- School-level results will also be available to Ofsted;
- National results will be reported by the DfE “to track standards over time”;
- National and local authority results will also be reported by the DfE “to allow schools to benchmark the performance of their pupils”.<sup>124</sup>

Seeing as this chapter has already outlined a range of issues caused by using the phonics check results for too many purposes, it is concerning that the results of the MTC will be used in a similar manner. For instance, it is reasonable to predict that there will be a noticeable rise in performance over the first few years of the MTC to reflect the increasing familiarity among teachers and pupils with the new assessment – as noted on the previous page regarding the phonics check. Second, the fact that Ofsted will consult each school’s results on the MTC draws this assessment into the high-stakes accountability system, meaning that teachers and school leaders will place considerable emphasis on how well their pupils perform. This could reduce the validity of the MTC, particularly in relation to the DfE’s attempts to track national standards. Paul Whiteman, general secretary of the NAHT, recently said that “if the government truly meant this to be a low-stakes check to inform teaching they wouldn’t be using the data from it at all for accountability purposes.”<sup>125</sup>

The decision not to introduce a ‘pass mark’ for the MTC (unlike the phonics check) means there is less likely to be the same clustering of results on and around a specific score. This should create a wider distribution of scores from 0 to 25 on the test, which will make it more informative for schools and policymakers. In addition, the use of an on-screen assessment reduces the possibility that pupils’ performance on the MTC will be influenced by the actions of teachers, whereas the phonics check has always been susceptible to teachers (perhaps inadvertently) providing some form of assistance to pupils.

The MTC is intended to reflect the updated National Curriculum, which states that “by the end of year 4, pupils should have memorised their multiplication tables up to and including the 12 multiplication table and show precision and fluency in their work.”<sup>126</sup> The DfE’s consultation on reforming primary assessment in 2017 explained why they felt this new assessment was necessary:

*“Knowledge and recall of multiplication tables is essential for the study of mathematics and for everyday life. Mastering multiplication is an important foundation for further learning in a number of aspects of mathematics, including division, algebra, fractions and proportional reasoning. Evidence shows that automatic retrieval of basic mathematical facts, such as multiplication tables, is critical to children’s effectiveness in solving more complex mathematical problems.”*<sup>127</sup>

The solitary supporting reference for this last sentence was an article from 2009 by Daniel T. Willingham, a professor of cognitive psychology at the University of Virginia. In the article, he cited a report by the National Mathematics Advisory Panel in the USA, which emphasised the importance of learning ‘factual knowledge’ in mathematics:

*“Factual knowledge refers to having ready in memory the answers to a relatively small set of problems of addition, subtraction, multiplication, and division. The answers must be well learned so that when a simple arithmetic problem is encountered (e.g.,  $2 + 2$ ), the answer is not calculated but simply retrieved from memory. Moreover, retrieval must be automatic (i.e., rapid and virtually attention free). This automatic retrieval of basic math facts is critical to solving complex problems because complex problems have simpler problems embedded in them.”*<sup>128</sup>

On that basis, Professor Willingham urged teachers to focus on helping students get to ‘automaticity’, which he said requires some memorisation and ample practice. Furthermore, he recommended that teachers “ensure that students have memorized basic math facts, such as the multiplication table up to  $12 \times 12$ ”.<sup>129</sup>

Given this evidence, there is nothing wrong with seeking to improve students’ retrieval of basic multiplication ‘facts’. However, what is striking about the MTC is that it overlooks the wider call from Professor Willingham to ensure that students can recall all “basic math facts” – not just multiplication tables. The new MTC ignores addition, subtraction and division and it is not clear why the DfE considers fluency in addition and subtraction to be less important than multiplication in this context, especially when the National Curriculum suggests that, regarding addition and subtraction, pupils should practise “mental methods ...with increasingly large numbers to aid fluency”.<sup>130</sup>

Even when discussing multiplication and division, the National Curriculum includes various objectives that the MTC ignores. For example, it states that “pupils should be taught to recall multiplication and division facts for multiplication tables up to  $12 \times 12$ ” and even suggests that “pupils continue to practise recalling and using multiplication tables and related division facts to aid fluency”,<sup>131</sup> yet division facts are excluded from the MTC. Pupils should also be taught to “use place value, known and derived facts to multiply and divide mentally, including: multiplying by 0 and 1; dividing by 1; [and] multiplying together three numbers”, but all these skills are disregarded by the MTC as well.

The reception from stakeholders to the introduction of the MTC has not been overly positive thus far. NAHT general secretary Paul Whiteman has recommended “that the test is scrapped”, deeming it “totally unnecessary” and adding to workload “to no benefit to pupils or teachers”.<sup>132</sup> Similarly, Mary Bousted, joint general secretary of the NEU, stated that they were opposed to the “unnecessary and expensive test” that “simply reflects a ministerial obsession”.<sup>133</sup> By assessing only one specific type of calculation for one specific age group within primary schools, the MTC is certainly an awkward addition to the assessment regime despite its potential cognitive value. What’s more, the movement towards online on-screen testing is a powerful innovation yet, as it stands, this is not utilised for any other part of the National Curriculum in mathematics or other subjects. A more coordinated and coherent approach to both mathematics and online testing should therefore be a prime consideration when determining the future of primary school assessment.

## 7. Alternative forms of assessment

### Teacher assessment

A frequently cited proposal is to replace national high-stakes tests in primary schools with assessments delivered by teachers instead. The 2008 Children, Schools and Families Committee reported that “many witnesses have called for greater emphasis on teacher assessment in order to enhance both the validity and the reliability of the testing system [...and] provide a more rounded view of children’s achievements”, particularly as witnesses “criticised the reliance on a ‘snapshot’ examination at a single point in time.”<sup>134</sup> Similarly, the Bew Review in 2011 found that over 80 per cent of respondents to an online call for evidence wanted to see teacher assessment replace externally-marked testing.<sup>135</sup> Many respondents argued that teacher assessment being used for school-level accountability “would demonstrate trust in teachers’ professional skills, remove incentives for inappropriate test preparation or narrowing of the curriculum, and avoid judging schools through a ‘snapshot’ of data from one week of tests.”<sup>136</sup>

The research evidence on this matter suggests that, far from promoting better judgements on pupils, teacher assessments are likely to be less valid and reliable than external tests:

- Harlen (2004) found that “there is bias in teachers’ assessment relating to student characteristics, including behaviour (for young children), gender and special educational needs” and “overall academic achievement and verbal ability may influence judgement when assessing specific skills”.<sup>137</sup>
- Burgess and Greaves (2009) found that teacher assessments “might be severely detrimental to the recorded achievements of children from poor families, and for children from some ethnic minorities”, while “external testing in some way protects pupils from typically low-attaining groups from subconscious assumptions.”<sup>138</sup>
- Using data for almost 5,000 pupils from the Millennium Cohort Study, research by Campbell (2015) “demonstrates biases in teachers’ average ratings of sample pupils’ reading and maths ‘ability and attainment’” which corresponded to key demographic characteristics such as income-level, gender, special educational needs and ethnicity.<sup>139</sup>
- In 2011, a summary of research in the USA on teacher assessment found that “teacher classroom summative assessment, that is, teacher grading practices, have historically and currently emphasised the lack of validity and reliability of these judgements”.<sup>140</sup>

- A 2016 meta-analysis of research findings on the existence of bias in grading student work in Australia found that “statistically significant” rates of bias “can occur ...when graders are aware of irrelevant information about the students” including their racial / ethnic background, ‘education-related deficiencies’ and poor prior performance.<sup>141</sup>

The Bew Review recognised there were “clear risks that summative teacher assessment will not be sufficiently reliable in a technical sense – i.e. that judgements will not be made consistently by teachers across the country”.<sup>142</sup> The Review also highlighted that there would be a “risk that teachers would come under increasing pressure to make generous assessments for some or all of their pupils.”<sup>143</sup> Some teaching unions suggested that these issues could be solved by a greater degree of ‘teacher moderation’ of each other’s judgements. However, the NFER told the Bew Review that “if the purpose [of summative teacher assessment] is about making judgements on teachers or institutions, there needs to be such stringent moderation practices that the assessment is no longer teacher assessment or it becomes unaffordable”.<sup>144</sup> In the end, the Bew Review concluded that “the evidence ...does not suggest to us that moderation would address the considerable risks around reliability of moving to a system based entirely on teacher assessment.”<sup>145</sup>

What’s more, the 2017 Education Select Committee inquiry into primary assessment “heard a wealth of evidence of the disadvantages of using teacher assessment within a high-stakes accountability system.”<sup>146</sup> For example, Tim Oates from Cambridge Assessment said “we have to be very realistic in terms of the level of dependability that we can yield from teacher assessment and whether it is always fair to expect teachers to assess with a level of consistency that we expect when we use the data for particular purposes.”<sup>147</sup> One primary school teacher even told the Committee that “teacher assessment for accountability, or in any high-stakes, or for any reliability purposes, is impossible [...] because you are basically judging yourself.”<sup>148</sup>

Even the studies that have suggested teacher assessment scores correlate with externally-marked tests<sup>149</sup> fail to account for the role of bias within the assessments, which is a problem because “it is perfectly possible for two separate scores to be very highly correlated, but for one of them still to be significantly biased” at the individual pupil level.<sup>150</sup> Moreover, these correlational studies typically use historical data on teacher assessments that are generated in a low-stakes environment, yet – as this report has repeatedly emphasised – the results of teacher-led assessments could change dramatically once the stakes are increased. This means that historical correlations between teacher assessments and external tests are unlikely to be replicated in a high-stakes accountability system.

To be clear, none of the biases detected by research studies are intentional on the part of teachers, nor are such biases restricted to educational contexts. As Tammy Campbell, author of the 2015 study cited earlier, has said: “I want to stress that this isn’t something unique to

teachers. It's human nature. Humans use stereotypes as a cognitive shortcut and we're all prone to it."<sup>151</sup> Nonetheless, the research evidence does not appear to support a concerted movement away from external testing in primary schools towards teacher assessments. The significant workload burden that teacher assessment generates should also not be underestimated. This was recently emphasised by the creation of 'teacher-assessed grades' to award grades to GCSE and A-level pupils in 2021, with numerous reports emerging of the excessive workloads generated by using teacher assessment in a high-stakes environment.<sup>152</sup> Improvements on this year's model would certainly be feasible in terms of reducing workload and improving standardisation, but such measures would still require extensive piloting to monitor the impact on reliability and validity – providing another reason why using teacher assessment for high-stakes testing will remain problematic for a long time to come.

## Comparative judgement

Despite the evidence cautioning against an over-reliance on teacher assessment, the Bew Review maintained that "the statutory assessment system should include both external testing and teacher assessment".<sup>153</sup> More specifically, "writing composition should be subject only to summative teacher assessment" because it would "encourage pupils to develop and demonstrate a broad range of writing skills over the course of Year 6 ... across a range of genres and writing styles, and that this approach should give children more opportunities than the current system does to write in a wide range of ways."<sup>154</sup> The Review proposed that the "teacher assessment of writing composition should be subject to external moderation", although some elements of writing such as spelling, punctuation, grammar and vocabulary still "lend themselves to externally-marked testing."<sup>155</sup>

Based on past experiences with SATs, there are several concerns with incorporating writing into the KS2 statutory assessments. First, assessing large quantities of writing assignments from a whole cohort of pupils has obvious workload implications for teachers. Second, the use of fairly rigid frameworks based on a set of 'The pupil can...' statements for assessing KS2 writing means it can be difficult for teachers to judge pupils' writing in a holistic manner. Third, as noted in the previous section on teacher assessment, the incentive to judge pupils leniently is a powerful one when each school's KS2 writing results are publicly available. In 2016, an investigation found that some areas of the country had large discrepancies – in some cases, as high as 15 percentage points – between the proportion of pupils achieving the 'expected standard' in reading (based on externally-marked SATs) and writing (based on teacher assessments). Joanna Hall from Ofsted told the Education Select Committee in 2017 that the Government should "have the debate" about whether to "remove teacher assessment entirely" from writing in primary schools, such was the variability in teachers' judgements.<sup>156</sup>

One approach to assessment that aims to tackle these issues is known as ‘comparative judgement’. This involves teachers being asked to judge two pieces of writing from different pupils against one another and say which represents the ‘better writing’ until all the pieces of writing are ranked, at which point the work is graded using an algorithm. Dr Becky Allen told the same Education Select Committee that, for open-ended tasks such as writing composition that do not suit prescriptive criterion-based mark schemes, “comparative judgement is such a compelling way for us to judge the standard of writing of 11-year-olds.”<sup>157</sup> Tim Oates added that it was “well worth exploring” the use of comparative judgements.<sup>158</sup>

A report in 2019 by Ofqual – the examination regulator for GCSEs and A-levels – highlighted several advantages to the approach taken by comparative judgement:

*“When making judgements, examiners and teachers often vary in their adherence to mark schemes or assessment criteria, and often make relative, as opposed to absolute, evaluations of pupils’ work. ...Comparative judgement takes advantage of this fact, building on the idea that it is easier to make relative judgements than absolute judgements, thus potentially improving reliability in judgements. ...Other advantages include the fact that very little training is needed for assessors compared to other methods, and this approach is able to control for any individual differences in severity/leniency in assessors. ...These factors increase the potential for reliability, and indeed good levels of reliability have been reported for assessments of writing using this method.”*<sup>159</sup>

However, the same report identified some drawbacks. For example, instructions are not usually given to teachers for how to differentiate between levels of performance (unlike the current ‘frameworks’ for KS2 writing). This can make it “difficult ...to know how assessors are making their judgements, and therefore how closely those judgements may or may not reflect the desired depth and breadth of the construct under consideration”.<sup>160</sup> There are also limited opportunities for feedback to pupils or to fellow teachers / assessors due to the simple comparison and ranking process, while ‘multi-dimensional assessment’ (e.g. producing separate scores for technical and compositional writing skills) is not possible without multiple rounds of judging,<sup>161</sup> which would increase the burden on teachers. Nevertheless, there appear to be considerable benefits to using comparative judgement for summative assessments of writing compared to the current system of moderated teacher judgements.

## Computer-based testing

Although this report has documented the raft of changes made to primary school assessments since the early 1990s, the assessment system in England has in some respects hardly changed at all. Tests have almost invariably been completed with a pen and paper, followed by a

marking process delivered by teachers or external examiners. Other countries have long since moved beyond this antiquated approach through the introduction of computer testing at various ages:

- **Norway:** after triggering its move to online testing in the early 2000's, Norway was the first country in Europe to implement 'e-assessment' on a national scale. Standardised online assessments are used to test all pupils in reading, mathematics and some parts of English language. These tests are usually held in the autumn after pupils have started the 5<sup>th</sup> grade (age 10), 8<sup>th</sup> grade (age 13) and 9<sup>th</sup> grade (age 14).<sup>162</sup>
- **Denmark:** since 2010, all national tests for primary and junior secondary students have been completed online and are 'adaptive' i.e. the difficulty of the questions adjusts during the test to match the proficiency of the pupil.<sup>163</sup> Pupils must take a reading test every second year from the second grade (age 8), a maths test in grades 3 (age 9), 6 (age 12) and 8 (age 14) and subject-specific tests across grades 7 and 8 in English, Geography, Physics/Chemistry and Biology. Two additional voluntary tests for Danish as a second language are available in grades 5 and 7.<sup>164</sup>
- **Australia:** The National Assessment Program – Literacy and Numeracy ('NAPLAN') is an annual assessment in May for students in Years 3 (ages 8-9), 5 (10-11), 7 (12-13) and 9 (14-15). The NAPLAN is made up of tests in reading, writing, language (spelling, grammar and punctuation) and numeracy. All schools are expected to have transitioned from the paper-based version of the NAPLAN tests to the new adaptive computer-based assessments ('NAPLAN Online') by 2022.<sup>165</sup>
- **Wales:** online 'personalised' (adaptive) assessments in numeracy and reading have recently replaced paper-based national tests for all learners in Years 2 to 9. Although there is no 'test window' for the assessments and there are no stakes attached to the results, pupils are required to take them at least once during the academic year and they can complete the assessments individually, in small groups or as a class.<sup>166</sup>

Adaptive tests have evidently become a trusted assessment tool for primary-aged pupils. Unlike a regular standardised assessment in which every learner is faced with the same questions in the same order, adaptive tests select questions based on the learner's response to the previous question. If a learner has answered a question incorrectly, their next question will be easier; if a learner answers a question correctly, their next question will be more difficult. This provides a personalised assessment for each pupil by automatically adjusting the degree of challenge for them regardless of their current ability level, thereby motivating every learner to succeed as well as accurately capturing their attainment.

As far back as 2011, the Bew Review commented that "some respondents have pointed to the potential of ...computer adaptive testing" and "there are several arguments in favour of

adaptive testing, which is already in use in primary schools in other countries". These benefits include the instant generation of results as well as the reduced time needed to complete adaptive tests relative to SATs.<sup>167</sup> However, the Review noted some potential limitations, such as unequal access to IT facilities and the need for some IT literacy on the part of pupils. The Review's overall conclusion was that "we believe the potential of computer adaptive testing should be explored further, including the relative suitability of the system for assessing specific subjects, with a view to exploring the possibility of introducing in the long term."<sup>168</sup>

More recently, when the DfE asked how to reduce the administrative workload generated by statutory tests as part of their 2017 consultation on reforming primary assessment, "a number of respondents suggested that increased use of online testing could reduce burdens."<sup>169</sup> Although it was noted that "moving to online testing would require significant lead-in times and resource", the DfE stated that "we will look to build on the development of [the Multiplication Tables Check] as we consider how technology could be utilised to reduce burdens associated with assessment in the future."<sup>170</sup>

## Sampling

The 2008 Children, Schools and Families Committee observed that, while the 'full-cohort testing' used for SATs makes it easy to track the performance of individual pupils, it is an unusual method of monitoring national standards. Many other countries such as Australia, Canada, France, New Zealand and Spain make use of 'sample testing' ('sampling') to judge the performance of their education system, which involves assessing a sample of pupils rather than the entire cohort.<sup>171</sup>

The Committee heard of numerous potential benefits of employing sampling to track national standards and the impact of government policy over time. These included reducing the stakes of the testing due to the anonymity of schools and pupils, being able to use a range of assessment methods and testing the same questions each year to increase validity.<sup>172</sup> As the Committee highlighted, sampling is a well-established technique and is used in international comparison studies such as PISA and TIMMS. It was also used in the UK from the mid-1970s and through the 1980s by the Assessment of Performance Unit within the DfE through the 'light' sampling of schools and pupils within schools.<sup>173</sup>

Some witnesses to the Committee mentioned possible drawbacks of sample testing. For example, the NFER said that "low-stakes assessment may not motivate pupils to try hard and show what they can really do, resulting in a potential underestimate of ability [...and] there may be practical difficulties with a system relying on voluntary participation of schools and pupils."<sup>174</sup> The Committee were also conscious that "if [individual school] accountability is to

be decoupled from national tests designed to measure pupil attainment, different tests or inspections will be required".<sup>175</sup> Even so, the Committee recommended that "the purpose of national monitoring of the education system, particularly for policy formation, is best served by sample testing to measure standards over time."<sup>176</sup>

What's more, a number of respondents to the Bew Review in 2011 suggested that sampling "would provide a more efficient way of monitoring national standards [as] they argue that a 'low-stakes' sample test could be administered securely, allowing the same questions to be re-used year after year, providing a very robust measure of standards over time."<sup>177</sup> However, the Review stated that "we do not feel it would be cost-effective to introduce a system of sample testing in addition to whole-cohort National Curriculum testing."<sup>178</sup> The 2017 Education Select Committee also heard evidence of the value of sampling, potentially as a replacement to school performance tables, although they felt that "in order to hold individual schools to account for the performance of pupils at Key Stage 2, statutory testing is arguably the best method".<sup>179</sup>

The DfE already uses a biennial sampling test for KS2 Science to provide a national picture of science attainment. Schools are notified by the STA of their involvement in April, with the test sat in June (the most recent test was scheduled for 2020 but was subsequently cancelled due to COVID-19). A representative sample of approximately 9,500 pupils are randomly selected, based on five pupils from 1,900 schools. Schools that are selected have a statutory obligation to participate. The three test papers are overseen in each school by external administrators, including the delivery of the tests and the marking of scripts. The sampling test results are only reported as national data, meaning that no individual school or pupil is identified and the results are not used for school accountability.<sup>180</sup>

In summary, there is broad agreement regarding the merits of 'sampling' to monitor national standards, but this must be done in a cost-effective manner and not be seen as a replacement for school-level accountability. Given that any new primary assessment system will need to analyse whether national standards are improving or declining over time, sampling is clearly a suitable candidate for reforming primary school assessment.

## 8. Recommendations

The opening chapter of this report set out five objectives that should underpin any upcoming reforms to primary school assessment in England. Below is a summary of what the analysis in this report has discovered in terms of how the current assessment and accountability regime fares against each objective:

### **1. Promoting high standards for pupils in all year groups in terms of their progress and attainment**

National testing can be a powerful mechanism for improving standards and the principle of using tests to collect and disseminate information about pupils and schools is widely accepted. That said, this report has uncovered several issues with the existing high stakes testing regime that may not be promoting (and could even be negating) the push for higher standards. This includes the prevalence of ‘teaching to the test’ and the narrowing of the curriculum – both of which are unwelcome side-effects of the high-stakes accountability system that could artificially increase test scores without improving standards. Furthermore, the shift to a new RBA-KS2 progress measure will mean that any pupil outside of Year 6 will not have any formal records of their performance that are available for parents, schools or the DfE.

### **2. Reflecting the contribution that schools make to their pupils’ education in a fairer and more proportionate manner**

Many commentators agree that it is fairer on schools to measure progress rather than attainment, as progress measures are a better reflection of the contribution that each primary school makes to their pupils’ education. It is also generally agreed that age 4 is a more appropriate ‘baseline’ for measuring progress than age 7, given that most pupils start primary school at this age. However, there are concerns about the fairness and proportionality of the new RBA, particularly the time lag created by using the RBA to present ‘progress’ data to parents several years after the event. What’s more, the degree of pupil mobility throughout primary education will mean that many schools will not have any attainment or progress measures in place from 2023 onwards, which will make it very difficult to judge the contribution that any school makes to a pupil’s education over the course of several years.

### **3. Ensuring that the assessment system supports high-quality teaching and learning rather than encouraging excessive test preparation**

The prevalence of ‘teaching to the test’ – driven by the high-stakes accountability system – can result in non-examined subjects being squeezed out of the curriculum for weeks, if

not months, in advance of statutory tests because schools understandably want to prepare their students for assessments on a specific skill / aptitude at a specific time. This means that precious teaching time is lost because schools are being encouraged to prepare pupils (sometimes an entire year group) for tests instead of focusing on high-quality teaching across a broad and balanced curriculum. Although the accountability system has undoubtedly contributed to these problems, the current assessment regime (particularly the advent of new tests like the MTC) could make the problem worse.

#### **4. Providing more accurate information on the performance of pupils and schools as well as performance at a national level**

The full-cohort testing provided by SATs can help assess standards across the country and over time. The issue is that, with the imminent demise of KS1 as statutory assessments, SATs results at age 11 will be the only publicly available measure of pupil attainment as well as school standards. Several other countries utilise testing much more frequently (typically every 2-3 years) to capture more timely data on pupil and school performance. In addition, other countries monitor national standards separately from pupil and schools, whereas SATs are used to monitor pupils, schools and the national picture too. Meanwhile, the phonics check and the MTC were designed to record national standards, but they are one-off assessments at different points that focus on a limited skill. What's more, these two assessments are being used for too many purposes, particularly their dual role as a component of Ofsted inspections as well as a measure of national standards – thereby reducing their validity.

#### **5. Reducing the assessment burden on teachers and schools**

The decision to make KS1 SATs non-statutory in 2023 was a sensible step by the DfE to reduce the burdens on schools. Similarly, the new RBA is designed to be a light-touch assessment for reception pupils that should not use up a large amount of lesson time. Nevertheless, several drivers of teacher workload remain in place. This includes the way that KS2 SATs in Year 6 promote excessive test preparation, introducing yet another statutory test in the form of the MTC and the impact of high-stakes assessment and accountability on teachers' and leaders' morale and motivation. Without significant changes to the assessment regime in primary schools, there are unlikely to be any tangible improvements in these areas.

It is clear, then, that change is needed if these five objectives for a reformed primary assessment system are to be met in future. As described earlier in this report, the DfE has already set in motion a package of changes that span the next few academic years. For instance, the DfE has previously announced (albeit before the pandemic) that they would bring the delivery of the RBA in-house “to keep costs down” when their current contract with the NFER expires.<sup>181</sup> The DfE is also planning to make KS1 tests non-statutory in 2023 as well

as introduce the RBA and the MTC in this academic year. Consequently, the recommendations in this report will outline a new vision for primary assessment starting from 2026/27, as this should allow sufficient time to develop, pilot and roll out the proposals.

## A new approach to assessment

### RECOMMENDATION 1

The current assessment system – including the full range of Key Stage 1 and Key Stage 2 SATs as well as the ‘reception baseline assessment’ and multiplication tables check – should be scrapped by 2026.

Even though it is reasonable to posit that the current system of KS1 and KS2 SATs has helped to improve primary school standards over the past thirty years, they have now outlived their useful purpose. Schools will understandably ‘teach to the test’ and narrow the curriculum in response to these assessments and their associated accountability metrics. The burdens that SATs generate in terms of the amount of time that schools must dedicate to them, particularly the extent of test preparation, is another undesirable feature of the current assessment and accountability system. In addition, the fact that KS2 SATs will soon be the only standardised national assessment across seven years of primary education demonstrates how the current approach does not provide the information that parents, schools and policymakers need about how well primary education is being delivered across all the different year groups. Thus, the first step in building a new approach to primary assessment is to reject the use of SATs as a method of measuring or improving standards.

Despite the solid rationale for the Government wanting to focus on progress rather than attainment, this report has documented several flaws with the way that the proposed RBA will be used. To be clear, the evidence shows that the RBA has been designed in a valid and reliable way as a test of the cognitive ability of reception-aged children. That said, it is unlikely that the scores produced by the RBA will subsequently generate fair and reasonable judgements in terms of pupil progress, particularly when pupil mobility is a constant feature of primary education and some schools will only teach pupils for a small portion of time from the ages of 4 to 11. Furthermore, the opportunities for ‘gaming’ the RBA have been highlighted by several academics and primary experts, thus potentially undermining its use as a starting point for measuring pupils’ progress. Meanwhile, the MTC will only test an incredibly specific skill at a single point in time across the whole of primary education, meaning that it will tell parents, schools and policymakers very little about the overall progress of each pupil in mathematics or their likely future trajectory.

## RECOMMENDATION 2

From 2026, assessments for reading, numeracy and spelling, punctuation and grammar (SPG) should be delivered through online adaptive tests. These tests automatically adjust the difficulty of the questions to match a pupil's performance, with a correct answer leading to a harder subsequent question and an incorrect answer leading to an easier question.

If the DfE's goal is to measure pupil progress and school standards in a valid and reliable manner at the same time as minimising the burdens on teachers and removing the need for excessive test preparation, there is now a compelling case for introducing adaptive tests in primary schools across England.

Other countries have led the way in showing how these tests can be administered quickly through secure online platforms. Although this report has called for the DfE to scrap their proposed MTC, the delivery model that this test will use – an online test that can be completed on a desktop computer, laptop or tablet – is precisely the foundation on which these new adaptive tests should be built. The three adaptive tests will be on:

- Reading
- Numeracy
- Spelling, punctuation and grammar (SPG)

## RECOMMENDATION 3

Pupils will take the new adaptive tests approximately once every two years. As a result, the tests will provide regular updates on how pupils are performing throughout primary education. Each test will last for around 30 minutes. School leaders will be able to choose how to schedule these tests during the academic year.

### **What types of questions will be included in the adaptive tests?**

The majority of questions in the adaptive tests will be multiple choice or require short answers from pupils. Even so, these questions can still require pupils to carry out a wide range of tasks such as spelling, completing calculations, interpreting short and long texts, inserting words and phrases into sentences, and so on. The computerised nature of the adaptive tests also means that for certain types of questions (e.g. multiplication) a timed element can be introduced so that pupils' speed of response can also be collected along with their accuracy – another reason why the MTC will no longer be required in future as a stand-alone assessment. Based on the online adaptive tests being used in Wales, each individual assessment is likely to last for 30-40 minutes depending on its design<sup>182</sup> with no additional work for teachers and school leaders in terms of marking due to the digitised assessment model.

### **When, and how often, will pupils sit these adaptive tests?**

One of the most concerning features of the upcoming RBA-to-SATs progress measure is that it will not produce sufficient data to judge pupil progress across different types of primary school. For example, if a school only teaches a pupil from the ages 4 to 7, it would be manifestly unfair for that school to be held to account on the basis of what happens to pupils after they leave to attend a different school from the ages of 7 to 11. In addition, there are over 300 primary schools in England that only teach pupils from the ages of 5 to 7 rather than 4 to 7, further narrowing the window for any form of national testing.

To address this, it is vital that the results of the new adaptive tests reflect the contribution made by schools to a pupil's progress in a fair and proportionate way. Furthermore, this report has found considerable evidence to suggest that the nature of SATs – one-off high stakes tests for specific year groups – encourages 'teaching to the test' and the narrowing of the curriculum. Their high-stakes nature also explains why some schools dedicate a considerable period (weeks, if not months) to preparing pupils for SATs, knowing that the performance of their pupils will be publicly reported.

Below is a proposed outline for how the new adaptive tests should be delivered in primary schools. A consultation process will be needed before any proposals are finalised, so this next section is intended to act as a starting point for discussions rather than a definitive solution:

- All primary pupils in England from Year 1 (age 5) to Year 6 (age 11) will sit each of the three tests – reading, numeracy and SPG – approximately once every two academic years (similar to the frequency in countries such as Denmark, Australia and Norway).
- Schools will be given at least 72 hours' notice by the STA that some of their pupils need to be tested to ensure that schools can excuse the pupils from normal lessons.
- On the day of the tests, the STA will electronically send schools a list of pupils from different year groups who need to be tested. Schools must then make the necessary arrangements for the pupils to take the tests in a quiet area away from regular lessons.
- As the tests will be carried out online, multiple pupils can be tested at the same time – even on different tests. Schools equipped with computer rooms could use this facility for concurrent testing, although the tests will be suitable for laptops and tablets as well.
- Although most pupils will only complete each adaptive test every other year, some pupils may be required to sit one of the tests in consecutive years or even sit the same test more than once in a single year.

In terms of how many days of testing will be required during the academic year, this report proposes that school leaders should be allowed to choose whichever format minimises (and potentially eliminates) disruption to their teaching schedules. For example, all the testing for an entire academic year could be carried out in just 2-3 days if an average-sized primary school of 280 pupils chose to have the equivalent of several classes sit tests over a single day, which would free up the rest of the year for regular classroom activities. On the other hand, a school leader may prefer to spread the short tests out over the academic year so that just a handful of pupils would take a test each week, meaning that the tests would essentially run in the background and cause no more disruption than, say, some pupils having a music lesson or being absent from school. Giving headteachers the choice of which approach suits them best would make the tests fit around their teaching schedule rather than the other way around. This would be operationalised by each school informing the STA in September of the maximum number of pupils that the headteacher will allow to be tested on any given day (higher numbers would mean fewer testing days over the academic year).

There are several logistical issues that would need to be discussed in advance of these new tests being rolled out (e.g. how to deal with pupils who miss their tests due to absence / illness). Nevertheless, the goal is to weave these new adaptive tests into the fabric of primary schools to the extent that the tests are viewed as nothing more than another aspect of school life rather than being seen as a huge obstacle to effective teaching and learning and delivering a broad and balanced curriculum (which the current assessment and accountability regime undoubtedly creates). Downgrading the visibility of these tests will also allow schools to focus on high-quality teaching and learning for all pupils throughout the year rather than worrying about which tests are coming up that month or that year (a culture that SATs have engendered).

What's more, this new system of pupils taking online tests approximately every two years will mean that data is collected from all the different types of primary school, and pupil progress can then be calculated on a fair and equivalent basis irrespective of how many year groups are taught within an individual school. That said, there are no tests delivered in reception under this new approach, as it is sensible to give schools time to help pupils acclimatise to their new environment as well as practise completing the online adaptive tests (using practice materials produced by the STA) before the actual testing commences the following year. This should mean that pupils will be unphased by the online tests when they are asked to complete them during the rest of their primary education.

## RECOMMENDATION 4

To encourage pupils to develop their creative writing skills throughout primary school, pupils' writing will be assessed in Year 2 (age 6-7) and Year 6 (age 10-11) through 'comparative judgement' exercises across all primary schools.

With reading, numeracy and SPG now being assessed through online adaptive tests, writing will also need some form of standardised assessment. This report has explained why the reliance on teacher assessment for KS1 and KS2 writing adds to teachers' workload while not necessarily producing valid and reliable scores for pupils. It is therefore proposed that, in future, writing is assessed through a national comparative judgement exercise. Unlike the new adaptive tests, though, the writing task will be sat by a whole year group at the same time – as it is now for SATs – to ensure that pupils are judged on the same creative writing task.

In line with the new adaptive tests, the following proposal for how comparative judgement could operate at a national level is intended to form the basis of a consultation and discussions between stakeholders on this matter:

- **In May each year, pupils in Year 2 and Year 6 will sit a writing assessment.** This will be done separately for each year group to prevent any excessive burdens on schools. The STA will provide schools with blank answer booklets for their pupils in advance of the assessments.
- **The assessments will be a creative writing task on a topic set by the STA** so that the work produced by pupils across the country can be compared.
- **In line with other tests, the writing assessments will be completed under controlled conditions** to ensure that the writing from each pupil is genuinely their work.
- **Once pupils have completed the assessment, their work will be uploaded into an STA portal** so that pupils' scripts can be assessed by each primary school and distributed electronically to other schools.
- **At this point, teachers will start comparing the scripts from their pupils and making judgement on which of two scripts is 'better' than the other,** and these judgements will be used to create a ranked list of their pupils' scripts.
- **Teachers will also be asked to compare scripts from two pupils from other schools** as these judgements will provide the 'anchor' that allows national comparisons between schools.

The most significant advantage of this new approach to writing is that it will avoid several concerns related to teacher assessment in its current form, particularly the issue of potential bias towards different groups of pupils. Consequently, the use of comparative judgement should represent a step forward for assessment in primary schools in terms of the validity of the results.

## RECOMMENDATION 5

To support parents' understanding of how their child is progressing through primary education, they should be provided with a report at the end of Years 2, 4 and 6 that shows their child's most recent results on the new adaptive tests and writing assessments.

Primary schools already work hard to keep parents informed about what level a child has reached at school through, among other things, end-of-term reports and parents' evenings. One of the benefits of the new adaptive testing at a national level is that it will offer a more comprehensive picture of how each pupil is developing throughout their time in primary education. In line with other countries such as Denmark and Wales, each pupil's results should be made available to parents so that they can discuss their child's progress with the teachers and school leaders.

To make this process as simple and accessible as possible, the STA should automatically send schools a full list of pupils' current levels of attainment and progress on all three tests relative to their peers across all primary schools. As pupils will be completing each of the new tests approximately every two years, it would make sense to present pupil-level information to parents at the end of Year 2, Year 4 and Year 6.

Rather than presenting pupils' raw scores on each test, their latest performance in terms of attainment and progress should be described in a more qualitative manner. As noted at the start of this report, the current system of teacher assessment typically describes pupils as 'working towards the expected standard', 'working at the expected standard' or 'working at a greater depth within the expected standard'. It would be worth consulting widely with parents and other stakeholders to see if this approach would be useful and meaningful in the context of the new adaptive tests and writing assessments, particularly when forming the basis for conversations about what a pupil must do to improve in future. Alternative options include a scale (e.g. 'below average' to 'above average') or a model based around the level of additional support that a pupil may require in each subject in light of their most recent results (e.g. 'no extra support needed', 'some extra support needed', 'much greater support needed').

### RECOMMENDATION 6

To help benchmark their performance and identify areas for improvement, primary schools should be provided with an annual 'profile' of results for each year group to show how they are performing on the adaptive tests and writing assessments. The profile will also include national averages and local authority averages as well as the results achieved at 'similar schools' around the country.

As noted earlier in this report, the DfE has decided not to give schools the results for each pupil on the RBA as they are worried that this might incentivise schools to 'label' or 'stream' pupils for the next seven years of their primary education based on their performance in the RBA. This is a reasonable position for ministers to take, and it is sensible for the new suite of assessments (reading, numeracy, SPG and writing) to follow suit.

This report recommends that, rather than schools receiving individual scores for pupils after they complete any of the tests, the school should instead be given an annual 'profile' of results for every year group that shows the overall distribution of scores. As with the reports to parents described in the previous recommendation, the format of the profiles for each year group should be kept simple – for example, using a 7-point scale for the percentage of pupils who, at that point in time, are 'well above average' down to 'well below average'. The results profile for each year group could also be reported in the same way for the writing assessments.

These 'profiles' are designed to capture the current level of performance across a whole year group to inform discussions within schools about improving teaching and learning. Providing class-level results may inadvertently encourage schools to use the profiles to judge individual teachers even when the cohort size may be too small to draw any meaningful conclusions on this matter, so it is better to focus on year-group profiles instead.

The results profile for the separate year groups will be presented alongside the national average and local authority average for each adaptive test / writing assessment. In addition, a comparison will be included in relation to the results achieved at 'similar schools' i.e. those schools that have a similar profile and intake (e.g. number of pupils on Free School Meals, number of pupils with English as an additional language, overall size, levels of prior attainment). This information will help schools to benchmark themselves in terms of the performance of their pupils, with the aim of informing their internal discussions about teaching and learning.

Teachers and school leaders will still be responsible for using their own formative and summative assessments as well as the new adaptive test 'profiles' to decide which pupils need

additional support or interventions in any subject or curriculum domain. Even so, the adaptive tests will provide a wealth of new information for schools across all year groups – a significant improvement on the infrequent summative data that SATs generate for teachers and leaders, which offers them little chance to improve the quality of teaching and learning. There would also be an opportunity for government to consider using the results profiles at each school as a trigger for additional central support or funding.

## RECOMMENDATION 7

There will be eight headline measures of accountability for primary schools:

- Pupils' average **attainment** and **progress** in reading
- Pupils' average **attainment** and **progress** in mathematics
- Pupils' average **attainment** and **progress** in spelling, punctuation and grammar
- Pupils' average **attainment** and **progress** in writing

All eight measures will be reported annually for every primary school using descriptors from 'well above average' to 'well below average'.

It is envisaged that the current accountability measures (described in Chapter 2) will largely remain in place, albeit with some adjustments to reflect the new scores being produced by adaptive tests and comparative judgement. In particular, the new adaptive tests offer three advantages over the existing SATs in terms of monitoring the performance of pupils:

- The new tests will be collecting data on a rolling basis throughout a pupil's time in primary education as opposed to waiting years until a pupil takes their SATs to calculate their attainment and progress as well as the overall level of attainment and progress within each primary school.
- By recording pupil performance across multiple year groups simultaneously instead of relying on results from a single year group, the accountability system will provide a more accurate measure of how a school is performing at present.
- The adaptive tests will be able to calculate the attainment and progress of pupils in relation to their exact age (in years and months) instead of treating all pupils in a year group the same way, thereby creating more finely tuned judgements on performance.

To take advantage of these features of the new adaptive tests, **attainment** and **progress** will still be the headline measures of accountability but they will be calculated in a different way to reflect the more frequent collection of performance data.

## Reading, numeracy and SPG

The adaptive tests will track the performance of pupils over time as every pupil will produce a new score whenever they take any of the tests (approximately every two years) during their primary education. This will allow the Government to make two calculations:

1. **Attainment:** the raw scores achieved by pupils in their most recent sitting of each adaptive test will be compared to the raw scores achieved by other pupils across the country who sat the test at *the same age* (in years and months). By comparing the scores of pupils in every school to the scores of pupils of identical ages in other schools, a school's level of 'attainment' in each of the three adaptive tests can be calculated.
2. **Progress:** the progress that pupils make between their different sittings of the tests throughout primary school will be measured by the three adaptive tests i.e. how much did their scores improve in the *exact length of time* (in years and months) since they last sat the same test. The average progress that pupils at a school make between test sittings can then be compared to the progress made between tests by pupils at other schools. By comparing the progress of pupils at different schools, an aggregated score of the average 'progress' made by pupils at each school can then be calculated.

As a starting point for discussions, a scoring system from 'well above average' to 'well below average' would represent a simple and accessible metric for reporting a school's performance on each of the three adaptive tests. It would be feasible to include the raw scores for attainment alongside these descriptors if stakeholders felt that this was a useful addition for parents. All these metrics should be updated on an annual basis.

## Writing

Comparative judgement may operate differently to the traditional marking of written tests, but it will nevertheless produce a national dataset based on the performance of tens of thousands of pupils and the judgements of thousands of teachers. This will allow comparisons to be made between schools as well as within schools. Attainment will be calculated using the average score of Year 2 and Year 6 pupils on the writing tasks and it will be reported using the same scale of 'well above average' to 'well below average'. Progress will be calculated using the average improvement in writing scores from Year 2 to Year 6. The Year 2 writing task will effectively provide the baseline for the progress score for each pupil up to Year 6, while a school's overall progress will be measured by aggregating the progress scores for pupils at each school and comparing this to the progress made by pupils in other schools.

## RECOMMENDATION 8

The phonics check should continue in Year 1 (age 5-6) but the concept of a 'pass threshold' should be removed. In addition, Ofsted should no longer be provided with each school's scores on the phonics check. These changes will reduce the likelihood of the results being distorted by having this assessment form part of the high-stakes accountability system.

Ministers are evidently strong supporters of the phonics check in terms of how it focuses schools' attention on ensuring that pupils are taught how to read in the most appropriate manner. This report shares the same ambition for all children to learn how to read as early as possible at primary school. In this context, using the phonics check to gauge national standards of reading skills is a worthwhile goal. However, the publication of a 'pass threshold' appears to be acting as a proxy for 'passing the phonics check' in the minds of many teachers, which explains the clear and obvious distortions in pupils' scores that were documented earlier in this report. In addition, giving Ofsted access to each school's phonics check results puts yet more pressure on teachers to make sure that pupils reach this arbitrary pass mark.

If the DfE wishes to use the phonics check to measure national standards in reading, this cannot be achieved in a valid and reliable way when a 'pass threshold' and Ofsted inspections are influencing the results of the phonics check in such dramatic fashion. It is therefore recommended that the phonics check for Year 2 pupils continues, but the notion of a 'pass threshold' should be abandoned and Ofsted should no longer be given access to pupils' scores on the check when they inspect schools (although Ofsted inspectors will still be entitled to inspect how well phonics and reading more broadly are being taught).

## RECOMMENDATION 9

To ensure that national standards are measured separately from the performance of pupils and schools, the national standards for reading, numeracy and spelling, punctuation and grammar should be judged in future using 'sample testing'. This will involve placing a small number of identical (or very similar questions) into the adaptive tests every year so that standards can be monitored over time.

The previous chapter noted how 'sampling' or 'sample testing' (i.e. testing a small group of pupils rather than an entire cohort) has often been touted as a replacement for SATs when it comes to monitoring national standards. However, past Select Committees and government reviews have concluded that decoupling the measurement of pupils from the measurement of national standards may involve developing a whole new set of tests in addition to the existing statutory tests – which would not sit comfortably alongside this report's goal of reducing the workload burdens on teachers and schools. In addition, this report has raised

concerns about the number of uses that the scores from a single assessment (particularly SATs) are currently used for, so simply stretching pupils' scores from the new adaptive tests across too many purposes would also not represent an obvious improvement.

Instead, this report recommends that the DfE should use this new vehicle of computerised testing to monitor national standards as well as judge individual pupils but to monitor national standards separately from recording each pupil's performance level. The simplest way to achieve this is by inserting identical (or near identical) questions into the adaptive tests sat by a pre-determined number of randomly selected pupils around the country, regardless of how well they are performing during their test. Through this mechanism, the DfE would be able to ask the same questions to a sample of pupils year-on-year, which would make it easy to measure national standards as well as track any changes in standards over time. No additional tests would be required for pupils or schools, and the national results could be published alongside school performance tables every year.

## 9. Areas for further consideration

### Reintroducing science tests

Last year, Ofsted's annual report claimed that the decision to drop science tests in primary schools might explain the country's "stagnant" performance in this subject in international rankings. The report also said that some primary schools have prioritised English and maths over other subjects, and that the quality and quantity of science teaching has reduced as a result. Amanda Spielman added that "we know from the DfE's sample test that key stage 2 science achievement has plummeted since these control levers were removed."<sup>183</sup> The proposed introduction of adaptive tests could be a sensible moment to re-evaluate the lack of science tests in primary schools. Adding new adaptive science tests alongside the suite of other new tests would be perfectly feasible, given the extensive period of development for these new tests in between now and 2026/27. New tests may arguably add to the burden on teachers and leaders so any decisions on this matter should again be consulted on widely. Even so, science is already tested through national sampling every two years, so there is a clear argument in favour of re-formalising the testing of science in future.

### Passing data from primary to secondary schools

While SATs results are typically passed from state-funded primary schools to secondary schools, secondary schools are not required to use this information. In fact, many secondary schools draw on their own assessments of new pupils' abilities and aptitudes when deciding how to group them. As noted earlier in this report, the DfE has decided not to give schools the results for each pupil on the new RBA as they are worried that this might incentivise schools to 'label' or 'stream' pupils for the rest of their time at primary school. The same problem applies to secondary schools, as some institutions use the results of SATs to produce 'flight paths' that predict a pupil's future progress in their GCSEs based on their past performance in SATs. Indeed, some secondary schools will put pupils into sets at the start of Year 7 based on their SATs results, knowing that the progress of secondary pupils will be measured by government from SATs up to their GCSE results in five years' time. This suggests that 'labelling' remains a problem for older age groups as much as for primary pupils.

In future, the Government could simply hand secondary schools the results of pupils' most recent adaptive tests and writing assessments, in the same way that secondary schools receive the SATs results for their Year 7 intake. This would allow secondaries to decide whether they want to use the results of the new suite of primary assessments to support new pupils. That said, it is possible that some primary pupils will not have sat one or more of the adaptive tests

in their final year, meaning that their previous scores on the adaptive tests may be slightly out-of-date. On that basis, there is a case to argue that the results of the new adaptive tests and writing assessments should not be shared with secondary schools as these may unduly influence their perceptions of – and judgements on – pupils when they first arrive (which could lead to ‘labelling’). Instead, secondary schools would be left to decide how best to assess their new intake with whatever tasks and assessments they deem necessary to produce a holistic judgement on the aptitudes of each pupil.

There are pros and cons to both approaches described above, so the Government should carefully consider which option would most closely match their aims for the new primary assessment system in terms of supporting the transition into secondary education as well as constructing a reliable progress measure for secondary schools.

### **The age at which pupils should begin taking the adaptive tests**

This report has proposed that the new adaptive tests should begin in Year 1, as this will remove any need to formally test pupils in their reception year – thus marking a departure from the use of the RBA with 4-year-olds. However, given that KS1 SATs are currently sat in Year 2 and the adaptive testing introduced in Wales only begins in Year 2, it is potentially worth considering an additional year of no testing under this new system. This would reduce the overall number of tests that a pupil would sit throughout primary education, but at the expense of delaying the ‘baseline’ for their subsequent performance on later tests (a problem that the RBA was supposed to overcome by being delivered at the start of primary education). Delaying the adaptive tests until Year 2 will also limit the information available on pupil attainment and progress in any primary school that only teaches up to Year 2, which would negate one of the most important benefits of the new adaptive tests – namely, that all schools would be treated equally by the new system. As a result, this report recommended that the new shorter adaptive tests should begin in Year 1, although it may be worth consulting with stakeholders to further interrogate this proposal – particularly around the level of support that would need to be available to younger pupils.

### **Moving beyond multiple-choice questions**

This report has proposed that the new adaptive tests will largely use questions that are multiple-choice or require short answers, as these are well-suited to this form of assessment. However, the capabilities of various forms of online testing are progressing at pace, so it is important to consider whether these new tests might also include a wider range of question types that would be better suited to testing a broader range of skills. This could form part of

the extensive piloting that this report envisages (over the course of several years) to see whether different types of questions would be suitable in the context of adaptive tests lasting around 30-40 minutes.

## The veracity of primary school tests

The STA define maladministration as “any act that could jeopardise the integrity, security or confidentiality of the national curriculum assessments and could lead to results that do not reflect the unaided abilities of pupils”.<sup>184</sup> This includes test papers being incorrectly opened, pupils cheating, over-aiding of pupils by test administrators, changes being made to a pupil’s script or the ‘inflation’ / ‘deflation’ of teacher assessment judgements. According to the STA’s latest report, almost 800 maladministration investigations were launched in 2018 – a rise of almost 200 compared to 2017, with over 80 per cent of those relating to KS2.<sup>185</sup> The main source for allegations were schools themselves (48 per cent of investigations) and the main type of allegation was a test administrator over-aiding pupils (33 per cent).<sup>186</sup> In total, 122 schools had their KS2 pupils’ results amended or annulled as a result of maladministration.<sup>187</sup>

Under the current testing arrangements, the number of pupils given extra time to complete SATs has almost doubled since 2015, with at least 14 per cent of Year 6 pupils being given additional time to finish the tests.<sup>188</sup> An independent survey in 2019 of over 1,200 teachers found that 20 per cent of staff involved in SATs reported having seen something that constitutes ‘maladministration’ and over 10 per cent of pupils are being given a ‘reader’ or a ‘scribe’ (or both) to help them complete SATs papers<sup>189</sup> - the latter being technically permissible but such practices are becoming more widespread without any obvious educational justification. Dr Becky Allen, who oversaw the aforementioned survey, commented that “we have a major problem if 1-in-10 eleven year-olds in mainstream education are not able to complete a maths assessment independently, since we’ll expect most of them to do maths unaided in secondary school lessons just a few months later.”<sup>190</sup>

When less than 1 per cent of primary schools are having their results amended or annulled, and serious maladministration such as pupils’ test papers being amended after completion is very rare, it does not seem necessary to introduce new measures to counteract these issues. It would be unwise to assume that online adaptive tests and comparative judgement would stop all forms of maladministration in any case. Nevertheless, the ability to influence the results of the new adaptive tests would be more restricted than for SATs given their digitised nature and reduced marking load for teachers at the same time as giving a more valid and reliable indication of a pupil’s current level of knowledge and understanding.

## Conclusion

*“The benefits brought about by [National Curriculum testing], compared to the time before the accountability of the National Curriculum, have been immense. The aspirations and expectations of pupils and their teachers have been raised. For parents, the benefits have been much better information not only about the progress their own child is making but also about the performance of the school their child attends. And for the education system as a whole, standards of achievement have been put in the spotlight, teachers’ efforts have been directed to make a difference and performance has improved.”<sup>191</sup>*

The above declaration, made by the then DfES to the Children, Schools and Families Committee in 2008, shows how deeply the commitment to national testing has run in government circles for many years. Successive Committees in Parliament have largely agreed that, in principle, the use of national tests can help raise standards. Nevertheless, the pervasiveness of the problems caused by the high stakes testing system in England, such as ‘teaching to the test’ and the narrowing of the curriculum, are as visible as they are disconcerting. This cannot and should not be ignored any longer.

The new approach outlined in this report aims to directly address the weaknesses of the existing assessment and accountability arrangements while still drawing on its various strengths:

- **Promoting high standards for pupils in all year groups in terms of their progress and attainment:** The new adaptive tests and comparative judgement exercises will maintain the Government’s focus on aiming for high standards in all schools, but it does so in a way that reduces the need to ‘teach to the test’ or narrow the curriculum. The regular, shorter tests sat by pupils will each represent just one point on their journey through primary education, meaning that the stakes of each test will be far lower than SATs. This should produce more accurate and useful data on the progress and attainment of pupils.
- **Reflecting the contribution that schools make to their pupils’ education in a fairer and more proportionate manner:** as it stands, the new progress measure that compares RBA scores with SATs results in Year 6 will not give a fair reflection of the work done by all schools due to the inability of this progress measure to track pupils throughout primary education. The new adaptive tests will build a more accurate and comprehensive picture of how pupils are progressing within each school, which will benefit parents, teachers, school leaders and policymakers because these tests will capture the progress made across all year groups in primary school as well as providing more up-to-date information for parents and addressing the problem of pupil mobility.

- **Ensuring that the assessment system supports high-quality teaching and learning rather than encouraging excessive test preparation:** by moving away from large disruptive assessments like SATs that can dominate an entire year of schooling, and by giving school leaders flexibility over how the new adaptive tests operate in their schools, it is hoped that much more room can be created for teaching and learning instead of preparing pupils for tests. Passing the results of the tests to parents and schools as well as providing schools with ‘profiles’ of each year group will also support leaders and teachers as they seek to identify any areas of underperformance or provide additional support to their pupils.
- **Providing more accurate information on the performance of pupils and schools as well as performance at a national level:** having regular shorter tests throughout primary education, coupled with the incorporation of ‘sample testing’ within the same assessment system, will give teachers, school leaders and policymakers a fuller picture of how standards are changing over time at a school, local authority and national level. This will identify schools that need additional support more rapidly than the current system allows. What’s more, preventing Ofsted from accessing a school’s phonic check results along with the removal of the ‘pass threshold’ will improve the validity of the phonics check in monitoring national standards in reading.
- **Reducing the assessment burden on teachers and schools:** by replacing SATs and the RBA as well as the upcoming MTC with smaller and shorter tests every couple of years for each pupil, it is hoped that teachers and leaders will see their workload reduced. In addition, the computerised nature of the new adaptive tests will eliminate the need for teachers to spend time marking their pupils’ work as the results will be compiled, distributed and analysed electronically. Giving school leaders the chance to shape when the new adaptive tests are delivered in their school will also mean that future national assessments are something that is done *with* schools instead of being done *to* schools.

There is no such thing as a ‘perfect’ assessment system for primary schools, and this report does not claim to have produced one. Nevertheless, the new approach described in this report would represent a genuine step forward in terms of how, why and when we assess what primary school pupils know and understand. As the title of this report suggests, ‘making progress’ in policy terms should be the goal for politicians and policymakers as much as it is for primary pupils. On that basis, the proposals in this report offer a positive and ambitious vision for the future of primary assessment and accountability system in England.

## References

- 
- <sup>1</sup> Professor P.J. Black, *National Curriculum: Task Group on Assessment and Testing* (London: Department of Education and Science and the Welsh Office, 1988), 7.
- <sup>2</sup> Department of Education and Science/Welsh Office, *The National Curriculum: A Consultation Document* (London: Her Majesty's Stationery Office, 1987), 35.
- <sup>3</sup> Black, *National Curriculum: Task Group on Assessment and Testing*, 62.
- <sup>4</sup> *Ibid.*, 65.
- <sup>5</sup> Anne West et al., 'The Standard Assessment Tasks and the Boycott at Key Stage 1: Teachers' and Headteachers' Views in Six Inner-city Schools', *Research Papers in Education* 9, no. 3 (1994): 321–37.
- <sup>6</sup> Ron Dearing, *The National Curriculum and Its Assessment: Final Report* (London: Her Majesty's Stationery Office, 1994), 13.
- <sup>7</sup> BBC News, 'School Tests Overhaul to Ease Stress', *BBC News Online*, 20 May 2003.
- <sup>8</sup> Tom Richmond and Sam Freedman, *Rising Marks, Falling Standards: An Investigation into Literacy, Numeracy and Science in Primary and Secondary Schools* (London: Policy Exchange, 2009), 52.
- <sup>9</sup> Polly Curtis, 'Science Sats to Be Scrapped but Maths and English Tests Expected to Continue', *The Guardian*, 6 May 2009.
- <sup>10</sup> Angela Harrison, 'Sats Boycott Hits Tens of Thousands of Pupils', *BBC News Online*, 10 May 2010.
- <sup>11</sup> The Conservative Party, *Invitation to Join the Government of Britain: The Conservative Manifesto 2010* (London: The Conservative Party, 2010), 52.
- <sup>12</sup> *Ibid.*
- <sup>13</sup> Jessica Shepherd, 'Sats Figures Show Reading Standards Drop for 11-Year-Olds', *The Guardian*, 3 August 2010.
- <sup>14</sup> Department for Education, '2010 to 2015 Government Policy: School and College Funding and Accountability', Webpage, 2015.
- <sup>15</sup> Laura McInerney, 'Nick Gibb: We Are Taking the Teacher Shortage Seriously', *Schools Week*, 1 October 2015.
- <sup>16</sup> National Foundation for Educational Research, *Achievement of 15-Year-Olds in England: PISA 2018 Results - Executive Summary* (London: Department for Education, 2019), 2.
- <sup>17</sup> Oxford University Centre for Educational Assessment, *Progress in International Reading Literacy Study (PIRLS): National Report for England* (London: Department for Education, 2017), 12.
- <sup>18</sup> The Labour Party, *It's Time for Real Change* (London: The Labour Party, 2019), 39.
- <sup>19</sup> Liberal Democrats, 'Plan for Britain's Future', Webpage, 2019.
- <sup>20</sup> Department for Education, *Primary Assessment and Accountability under the New National Curriculum* (London: Department for Education, 2013), 4.
- <sup>21</sup> *Ibid.*
- <sup>22</sup> *Ibid.*
- <sup>23</sup> *Ibid.*, 5.
- <sup>24</sup> Department for Education, *Reforming Assessment and Accountability for Primary Schools: Government Response to Consultation on Primary School Assessment and Accountability* (London: Department for Education, 2014), 6.
- <sup>25</sup> Department for Education, 'Reception Baseline Comparability Study Published', Webpage, 7 April 2016.
- <sup>26</sup> Department for Education and the Standards and Testing Agency, *Primary Assessment in England: Government Consultation* (London: Department for Education, 2017), 16.
- <sup>27</sup> Department for Education, *Primary Assessment in England: Government Consultation Response* (London: Department for Education, 2017), 13.
- <sup>28</sup> Helen Ward, 'Plans to Broaden Baseline beyond Maths and Literacy Dropped', *Times Educational Supplement*, 27 February 2019.

- 
- <sup>29</sup> Standards and Testing Agency, *Reception Baseline Assessment Validity: Report on the RBA Pilot* (London: Department for Education, 2020), 24.
- <sup>30</sup> *Ibid.*, 15.
- <sup>31</sup> Schools Week reporter, 'DfE Claims Trial Shows Baseline IS Accurate, and Confirms September Roll-Out', *Schools Week*, 26 February 2020.
- <sup>32</sup> Department for Education, *Primary Assessment in England: Government Consultation Response*, 14.
- <sup>33</sup> Standards and Testing Agency, *Reception Baseline Assessment Validity: Report on the RBA Pilot*, 17–18.
- <sup>34</sup> Harvey Goldstein et al., *Baseline without Basis: The Validity and Utility of the Proposed Reception Baseline Assessment in England* (London: British Educational Research Association, 2018), 28.
- <sup>35</sup> Harvey Goldstein et al., *A Baseline without Basis: The Validity and Utility of the Proposed Reception Baseline Assessment in England* (British Educational Research Association, n.d.), 29.
- <sup>36</sup> *Ibid.*, 16.
- <sup>37</sup> *Ibid.*
- <sup>38</sup> Goldstein et al., *Baseline without Basis: The Validity and Utility of the Proposed Reception Baseline Assessment in England*, 3.
- <sup>39</sup> *Ibid.*, 2.
- <sup>40</sup> Department for Education, *Primary Assessment in England: Government Consultation Response*, 22.
- <sup>41</sup> *Ibid.*
- <sup>42</sup> *Ibid.*
- <sup>43</sup> National Foundation for Educational Research, *Department for Education Primary Assessment in England Government Consultation: NFER Response* (London: NFER, 2017), 13.
- <sup>44</sup> *Ibid.*
- <sup>45</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, HC-682 (London: Her Majesty's Stationery Office, 2017), 21.
- <sup>46</sup> *Ibid.*, 20.
- <sup>47</sup> *Ibid.*, 19.
- <sup>48</sup> House of Commons Education Committee, 'Oral Evidence: Primary Assessment - HC 682', Webpage, 18 January 2017.
- <sup>49</sup> Standards and Testing Agency, *Reception Baseline Assessment Validity: Report on the RBA Pilot*, 33.
- <sup>50</sup> Guy Roberts-Holmes et al., *Research into the 2019 Pilot of Reception Baseline Assessment (RBA)* (London: UCL Institute of Education, 2020), 6.
- <sup>51</sup> *Ibid.*, 7.
- <sup>52</sup> Standards and Testing Agency, *RBA Quality Monitoring Visits Guidance: June 2021* (London: DfE, 2021).
- <sup>53</sup> Liz Lightfoot, 'Fears That New Tests Will Damage Four-Year-Olds "Unfounded"', *The Guardian*, 5 November 2019.
- <sup>54</sup> *Ibid.*
- <sup>55</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08 - Volume I*, HC169-I (London: Her Majesty's Stationery Office, 2008), 10.
- <sup>56</sup> *Ibid.*, 12.
- <sup>57</sup> *Ibid.*
- <sup>58</sup> *Ibid.*, 13.
- <sup>59</sup> *Ibid.*, 17.
- <sup>60</sup> *Ibid.*
- <sup>61</sup> *Ibid.*, 18.
- <sup>62</sup> *Ibid.*

- 
- <sup>63</sup> Ibid.
- <sup>64</sup> Ibid., 41.
- <sup>65</sup> Ibid., 42.
- <sup>66</sup> Ibid., 43.
- <sup>67</sup> Ibid.
- <sup>68</sup> Ibid.
- <sup>69</sup> Ibid., 44.
- <sup>70</sup> Ibid.
- <sup>71</sup> Ibid., 49.
- <sup>72</sup> Department for Education, *The Importance of Teaching* (London: Her Majesty's Stationery Office, 2010), 48.
- <sup>73</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, 15.
- <sup>74</sup> Ibid., 16.
- <sup>75</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08 - Volume I*, 49–50.
- <sup>76</sup> Ibid., 50.
- <sup>77</sup> Ibid.
- <sup>78</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, 15.
- <sup>79</sup> Katherine Sellgren, 'Teaching to the Test Gives "Hollow Understanding"', *BBC News Online*, 11 October 2017.
- <sup>80</sup> Katherine Sellgren, 'Ofsted Admits Adding to "teach-to-the-Test" Mentality', *BBC News Online*, 18 September 2018.
- <sup>81</sup> Ibid.
- <sup>82</sup> Sian Griffiths, 'School Inspectors Will Look beyond Sats Test Results, Says Ofsted Chief Amanda Spielman', *The Sunday Times*, 12 May 2019.
- <sup>83</sup> Ofsted, 'Education Inspection Framework 2019: Inspecting the Substance of Education', Webpage, 29 July 2019.
- <sup>84</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08*, HC 169-I (London: The Stationery Office, 2008), 33.
- <sup>85</sup> Ibid., 39.
- <sup>86</sup> Ibid., 33.
- <sup>87</sup> Lord Bew, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report* (London: DfE, 2011), 9.
- <sup>88</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, 19.
- <sup>89</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08*, 34.
- <sup>90</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08 - Volume I*, 39.
- <sup>91</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, 19.
- <sup>92</sup> Bew, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report*, 12.
- <sup>93</sup> Carol Robinson and Michael Fielding, *Children and Their Primary Schools: Pupils' Voices (Primary Review Research Survey 5/3)* (Cambridge: University of Cambridge Faculty of Education, 2007), 16.
- <sup>94</sup> Ibid.
- <sup>95</sup> Ibid.
- <sup>96</sup> Ibid.
- <sup>97</sup> Ibid., 17.
- <sup>98</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08 - Volume I*, 55.

- 
- <sup>99</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, 27.
- <sup>100</sup> ComRes, *BBC Newsround – SATs Survey* (London: ComRes, 2016).
- <sup>101</sup> John Jerrim, 'National Tests and the Wellbeing of Primary School Pupils: New Evidence from the UK', *Assessment in Education: Principles, Policy & Practice*, 23 September 2021, 1–38.
- <sup>102</sup> More Than A Score, *Too Many Tests for No Good Reason: What Parents Really Think about Standardised Testing in Primary Schools* (London: More Than A Score, 2020).
- <sup>103</sup> Ofsted, 'HMCI's Commentary: Recent Primary and Secondary Curriculum Research', Webpage, 11 October 2017.
- <sup>104</sup> Richard Adams, 'Primary Schools in England Hold Half-Term Sats Revision Classes', *The Guardian*, 13 February 2020.
- <sup>105</sup> Dr Alice Bradbury, *Pressure, Anxiety and Collateral Damage: The Headteachers' Verdict on SATs* (London: UCL Institute of Education, 2019), 23.
- <sup>106</sup> Carol Robinson and Michael Fielding, *Children and Their Primary Schools: Pupils' Voices (Primary Review Research Survey 5/3)* (Cambridge: University of Cambridge Faculty of Education, 2007), 16.
- <sup>107</sup> Eleanor Busby, 'Children Should Not Know They Are Doing Sats Exams to Prevent Anxiety, Ofsted Chief Suggests', *The Independent*, 14 May 2019.
- <sup>108</sup> *Ibid.*
- <sup>109</sup> Steve Lancashire, 'Children Doing Tests Should Not Cause Such Hysteria - in Good Schools, Pupils Aren't Stressed', *The Daily Telegraph*, 11 May 2016.
- <sup>110</sup> *Ibid.*
- <sup>111</sup> Will Hazell, 'Exclusive: Primary Pupils "pretty Relaxed" about Sats Tests, School Chain Claims', *The i Paper*, 6 September 2019.
- <sup>112</sup> *Ibid.*
- <sup>113</sup> Standards and Testing Agency, *Assessment Framework for the Development of the Year 1 Phonics Screening Check: For Test Developers* (London: Department for Education, 2017), 4.
- <sup>114</sup> Department for Education, *Phonics Screening Check and National Curriculum Assessments at Key Stage 1 in England (2011/12)* (London: DfE, 2012), 3.
- <sup>115</sup> *Ibid.*
- <sup>116</sup> *Ibid.*, 1.
- <sup>117</sup> *Ibid.*, 2.
- <sup>118</sup> *Ibid.*, 4.
- <sup>119</sup> Ofsted, *School Inspection Handbook: Handbook for Inspecting Schools in England under Section 5 of the Education Act 2005* (Manchester: Ofsted, 2019), 87.
- <sup>120</sup> *Ibid.*
- <sup>121</sup> Ofsted, 'School Inspection Handbook', Webpage, 2021.
- <sup>122</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08 - Volume I, HC169-I* (London: Her Majesty's Stationery Office, 2008), 58.
- <sup>123</sup> *Ibid.*, 59.
- <sup>124</sup> Standards and Testing Agency, *Multiplication Tables Check Assessment Framework* (London: Department for Education, 2018), 4.
- <sup>125</sup> Martin George, 'Times-Tables Check "an Enormous Waste of Money"', *Times Educational Supplement*, 13 November 2018.
- <sup>126</sup> Department for Education, *Mathematics Programmes of Study: Key Stages 1 and 2 (National Curriculum in England)* (London: DfE, 2013), 17.
- <sup>127</sup> Department for Education and the Standards and Testing Agency, *Primary Assessment in England: Government Consultation* (London: Department for Education, 2017), 24.

- 
- <sup>128</sup> Daniel T. Willingham, 'Is It True That Some People Just Cannot Do Math?', *American Educator*, 2009, 16.
- <sup>129</sup> *Ibid.*, 19.
- <sup>130</sup> Department for Education, *Mathematics Programmes of Study: Key Stages 1 and 2 (National Curriculum in England)*, 25.
- <sup>131</sup> *Ibid.*
- <sup>132</sup> Mark Smulian, 'Heads Call for Times Tables Test to Be Scrapped', *Times Educational Supplement*, 8 September 2019.
- <sup>133</sup> Eleanor Bubby, 'Government Trials Times Tables Test for Eight-Year-Olds despite Opposition from Teacher Unions and Parents', *The Independent*, 2018.
- <sup>134</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08 - Volume I*, 24.
- <sup>135</sup> Lord Bew, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report* (London: DfE, 2011), 53.
- <sup>136</sup> Department for Education, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report*, 2011, 48.
- <sup>137</sup> Bew, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report*, 49.
- <sup>138</sup> *Ibid.*
- <sup>139</sup> Tammy Campbell, 'Stereotyped at Seven? Biases in Teacher Judgement of Pupils' Ability and Attainment', *Journal of Social Policy* 44, no. 3 (2015): 517–47.
- <sup>140</sup> Susan M. Brookhart, 'The Use of Teacher Judgement for Summative Assessment in the USA', *Assessment in Education: Principles, Policy & Practice* 20, no. 1 (2013): 69–90.
- <sup>141</sup> John M. Malouff and Einar B. Thorsteinsson, 'Bias in Grading: A Meta-Analysis of Experimental Research Findings', *Australian Journal of Education* 60, no. 3 (2016): 245–56.
- <sup>142</sup> Bew, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report*, 50.
- <sup>143</sup> *Ibid.*
- <sup>144</sup> *Ibid.*, 51.
- <sup>145</sup> *Ibid.*
- <sup>146</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, HC-682 (London: Her Majesty's Stationery Office, 2017), 10.
- <sup>147</sup> *Ibid.*
- <sup>148</sup> *Ibid.*, 11.
- <sup>149</sup> Kaili Rimpfeld et al., 'Teacher Assessments during Compulsory Education Are as Reliable, Stable and Heritable as Standardized Test Scores', *Journal of Child Psychology and Psychiatry* 60, no. 12 (2019): 1278–88.
- <sup>150</sup> Daisy Christodoulou, 'Even Reliable Assessments Can Be Biased', Webpage, 2019.
- <sup>151</sup> Richard Adams, 'Children from Poorer Families Perceived by Teachers as Less Able, Says Study', *The Guardian*, 2015.
- <sup>152</sup> Louise Lewis, 'My Life Right Now? Marking, Marking and More Marking', *Times Educational Supplement*, 20 May 2021; Anonymous, 'Why GCSEs 2021 Are Another Shambles Waiting to Happen', *Times Educational Supplement*, 16 March 2021.
- <sup>153</sup> Bew, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report*, 13.
- <sup>154</sup> *Ibid.*, 14.
- <sup>155</sup> *Ibid.*
- <sup>156</sup> Freddie Whittaker, 'Consider Scrapping KS2 Teacher-Marked Writing Tests, Says Ofsted Director', *Schools Week*, 18 January 2017.
- <sup>157</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, 11.
- <sup>158</sup> Whittaker, 'Consider Scrapping KS2 Teacher-Marked Writing Tests, Says Ofsted Director'.

- 
- <sup>159</sup> Ofqual, *A Review of Approaches to Assessing Writing at the End of Primary Education: Drawing upon Historical and International Practices* (Coventry: Ofqual, 2019), 30.
- <sup>160</sup> *Ibid.*, 28.
- <sup>161</sup> *Ibid.*
- <sup>162</sup> Leire Aranbarri, 'Digitise the Norwegian National Tests at Scale', Webpage, 14 January 2019.
- <sup>163</sup> Louise V. Beuchert and Anne B. Nandrup, *The Danish National Tests at a Glance* (Aarhus: Aarhus University, 2017).
- <sup>164</sup> Eurydice, 'Denmark: Assessment in Single Structure Education', Webpage, 2020.
- <sup>165</sup> Australian Curriculum, Assessment and Reporting Authority, 'NAPLAN Online', Webpage, 2021.
- <sup>166</sup> Hwb, 'Reading and Numeracy Assessments', Webpage, 2019.
- <sup>167</sup> Bew, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report*, 74.
- <sup>168</sup> *Ibid.*
- <sup>169</sup> Department for Education, *Primary Assessment in England: Government Consultation Response* (London: Department for Education, 2017), 29.
- <sup>170</sup> *Ibid.*
- <sup>171</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08 - Volume I*, 89–90.
- <sup>172</sup> *Ibid.*, 62.
- <sup>173</sup> *Ibid.*
- <sup>174</sup> *Ibid.*
- <sup>175</sup> *Ibid.*
- <sup>176</sup> *Ibid.*, 66.
- <sup>177</sup> Bew, *Independent Review of Key Stage 2 Testing, Assessment and Accountability: Final Report*, 56.
- <sup>178</sup> *Ibid.*
- <sup>179</sup> House of Commons Education Committee, *Primary Assessment: Eleventh Report of Session 2016–17*, 21.
- <sup>180</sup> Standards and Testing Agency, *Key Stage 2 Assessment and Reporting Arrangements* (London: Department for Education, 2019), 47.
- <sup>181</sup> Amy Gibbons, 'Exclusive: DfE to Bring Baseline Assessment in-House', *Times Educational Supplement*, 18 October 2019.
- <sup>182</sup> Welsh Government, 'Personalised Assessments – Frequently Asked Questions (FAQs) for Parents and Carers 2020 to 2021', Webpage, 2021.
- <sup>183</sup> John Roberts, 'Ofsted Links Weak Science Scores to Lack of KS2 Testing', *Times Educational Supplement*, 21 January 2020.
- <sup>184</sup> Standards and Testing Agency, *2018 Maladministration Report: National Curriculum Assessments at Key Stage 1 and Key Stage 2* (London: Department for Education, 2020), 3.
- <sup>185</sup> *Ibid.*, 5.
- <sup>186</sup> *Ibid.*
- <sup>187</sup> *Ibid.*, 11.
- <sup>188</sup> Helen Ward, 'Exclusive: Doubling in Number of Pupils given Extra Time for Sats', *Times Educational Supplement*, 4 September 2018.
- <sup>189</sup> Allen, Rebecca, 'Nobody Knows Which Schools Are Good', Webpage, 22 September 2019.
- <sup>190</sup> *Ibid.*
- <sup>191</sup> House of Commons Children, Schools and Families Committee, *Testing and Assessment: Third Report of Session 2007–08 - Volume I*, 10.