

Review of Key Stage 2 testing, assessment and accountability

Progress Report

Contents

Foreword by Lord Bew	3
Membership of the Review Panel.....	4
Abbreviations	4
Remit of the Review and Outline of the Evidence Gathering Process.....	5
Introduction	7
Overview of the current system.....	9
Purposes of statutory assessment.....	9
Benefits and deficits of the current system	10
Accountability	14
The role of summative teacher assessment judgements	20
Specific groups of pupils and specific types of school.....	26
Specific groups of pupils.....	26
Specific types of primary schools	28
Specific aspects of current statutory tests	29
Writing tests.....	29
Science Assessment	30
Validity and reliability of Key Stage 2 tests	31
The use of levels	32
Timing of tests	32
Alternative approaches to statutory assessment.....	34
Transition to Secondary School	36
Annex A – Stakeholders who have submitted evidence to the Review.....	37

Foreword by Lord Bew

It is clear that testing, assessment and accountability at Key Stage 2 have a huge influence on primary school education. Ensuring they are right is extremely important for pupils, parents, teachers, head teachers and the nation as a whole.

It is therefore unsurprising that ways of reforming this part of the system are as hotly debated as any other area of education policy. The strength of feeling shows the importance of finding the right way forward.

The panel and I feel the weight of responsibility of this Review, and are committed to bringing about meaningful, fair and lasting improvements. We realise how hard teachers and head teachers across England work and we want to support and empower them to give every child the best possible education.

Change is clearly needed, and we are grateful for the many suggestions we have heard about how the system could be improved. However, we have been struck by the complexity of the challenge we face. Each proposal has arguments against, risks and uncertainties, and it is becoming increasingly clear, as many will have already concluded for themselves, that there will not be a single set of solutions which can command universal support.

This Progress Report outlines the evidence and ranges of opinion that we have heard so far. We will continue to review all of the evidence and will develop, between now and the Final Report this summer, both short term and long term recommendations to address the problems of the current system. This report does not include any recommendations. They will be outlined in the Final Report when we have fully considered all the evidence and opinion submitted.

I would like to take this opportunity to thank the several thousand stakeholders who have submitted evidence to the Review. We greatly appreciate the 4000 responses to the online consultation, the 50 stakeholders who have spoken to us in formal evidence sessions, and the many written submissions. The review has generated an invaluable amount of evidence and feedback from a wide range of stakeholders about Key Stage 2 testing, assessment and accountability, which we will ensure we take into account and learn from as we move towards our recommendations.

Lord Paul Bew

Membership of the Review Panel

Lord Bew is a cross-bench peer, Professor of Irish Politics at Queen's University in Belfast, and a Member of the Royal Irish Academy (MRIA). He was a historical adviser to the Saville Inquiry from 1998 to 2001.

Membership of the panel in full is:

- Lord Bew – Chair.
- Helen Clegg OBE – Head teacher, Shiremoor Primary School in North Tyneside. National Leader of Education.
- Sally Coates – Principal, Burlington Danes Academy in West London.
- Kate Dethridge – Head teacher, Churchend Primary School in Reading. National Leader of Education.
- Lubna Khan – Head teacher, Berrymede Junior School in Ealing. Local Leader of Education.
- Ruth Miskin – founder, Read-Write Inc. and former primary head teacher.
- Miriam Rosen – former Executive Director, Ofsted.
- Tim Sherriff – Head teacher, Westfield Community School in Wigan. Local Leader of Education.
- Greg Wallace – Executive Principal of Best Start Federation in Hackney. National Leader of Education.

Representatives of Ofsted and Ofqual act as observers.

Abbreviations

AAIA	Association for Achievement and Improvement through Assessment
ACME	Advisory Council for Maths Education
ASCL	Association of School and College Leaders
ATL	Association of Teachers and Lecturers
CIEA	Chartered Institute of Education Assessors
NAHT	National Association of Head Teachers
NASUWT	National Association of Schoolmasters/Union of Women Teachers
NCPTA	National Confederation of Parent Teacher Associations
NFER	National Foundation for Education Research
NGA	National Governors' Association
NUT	National Union of Teachers
OECD	Organisation for Economic Cooperation and Development
QCDA	Qualifications and Curriculum Development Agency
SCORE	Science Community Representing Education

Remit of the Review and Outline of the Evidence Gathering Process

This review has been framed by a particular remit, which it may be helpful to revisit at this stage so that we can be clear about what is in our scope.

There are two broad positions that the Secretary of State has asked us to adhere to throughout. Firstly, we are mindful that the Organisation for Economic Cooperation and Development (OECD) concludes that external accountability is a key driver of improvement in education and particularly important for the least advantaged, and so the Government views a system of objectively measuring pupil progress and holding schools to account as vital. Secondly, the Government has made it clear that it wants schools and teachers to be free to set their own direction, trusted to exercise their professional discretion and be accountable for the progress of the children in their care. The Secretary of State has therefore been clear that school autonomy must be accompanied by robust accountability.

Within those parameters this Review is designed to cover the following key issues:

- how best to ensure that the system of assessment in primary schools can improve standards of attainment and progress of pupils, and help narrow gaps;
- how best to ensure that schools are properly and fairly accountable to pupils, parents and the taxpayer for the achievement and progress of every child, on the basis of objective and accurate assessments; and that this reflects the true performance of the school;
- how to avoid, as far as possible, the risk of perverse incentives, over-rehearsal and reduced focus on productive learning;
- how to ensure that parents have good quality information on the progress of their children and the success of schools;
- how to ensure that performance information is used and interpreted appropriately within the accountability system by other agencies, increasing transparency and preserving accountability to parents, pupils and the taxpayer, while avoiding the risk of crude and narrow judgements being made;
- how to ensure that tests are rigorous, and as valid and reliable as possible, within an overall system of assessment (including teacher assessment) which provides the best possible picture of every child's progress;
- how best to ensure that the assessment system allows us to make comparisons with education systems internationally; and
- how to make administration of the system as simple and cost-effective as possible, with minimal bureaucracy.

We expect that this review will make recommendations which the National Curriculum Review will wish to consider as it moves towards its final conclusions.

Given the scale and complexity of this review we have endeavoured to gather as much evidence and feedback as possible, in an open, transparent and outward facing way. A 12-week

call for evidence has invited all interested parties to contribute. As a result we have received over 4000 online responses, taken oral evidence from 50 stakeholders, and many written submissions have been sent in. The stakeholders from whom we have received evidence are listed in Annex A.

Introduction

Concerns about ‘teaching to the test’ and narrowing of the curriculum, rather than being issues which have emerged recently, have been discussed now for over a century. At the turn of the 20th century teachers were not only worried about these issues, but felt it was ‘daft’ that they had not been resolved in the 19th century¹. Exactly 100 years ago, England’s Board of Education set up the Consultative Committee on Examinations in Secondary Schools, noting that “*examinations as ends in themselves have occupied too much of the thoughts of parents and teachers. Their very convenience and success led to their undue multiplication and to their occupying too large a place in the system of national education*”.

In the 21st century we are still debating these issues as fiercely as ever. The volume of online consultation responses and the range of opinion expressed reflects the level of concern that exists today. Such strength of feeling is to be expected given that the impact of assessment on education is so significant.

We realise that every assessment process will have some degree of unintended consequence. Given the importance of getting testing, assessment and accountability right, the feedback that too many schools feel they must drill children for tests and spend too much time on test preparation in Year 6 at the expense of productive teaching and learning is deeply worrying. There are also significant concerns about a focus on children at the borderline of levels (especially levels 3 and 4) at the expense of other pupils, which is clearly undesirable. In addition, many heads and teachers have raised concerns about the way results from statutory assessment currently contribute to the school accountability system.

We have set out to review the current system externally and make recommendations about how we should tackle the problems. Many contributors recognise positive features of the current system which we need to be careful to protect, such as the proven impact of strong accountability on driving up achievement and progress, and the value parents place on test results in providing information about how their child and their child’s school are performing².

This report summarises the evidence which the review has received, both from the wide range of individuals and organisations who have presented to us directly and from the online call for evidence. This call for evidence received nearly four thousand responses, of which 61% came from primary heads and 23% from primary teachers. Parents contributed 4% of responses, governors 3%, secondary teachers and heads 2% and local authorities 1%. A selection of views from the call for evidence are included in this document.

In view of the strength of feeling and the wide range of views held, it is inevitable that some of the evidence we have been presented with is conflicting or contradictory. This progress report does not seek to reconcile or evaluate such differences of opinion, or to give particular weight to the views of teachers or academics. It simply sets out our understanding of the evidence we have seen.

We have heard many references to how other countries are reforming their approaches to testing, assessment and accountability, which England may be able to learn from. We should seek to learn from other countries’ approaches. However, as every system functions in its own unique context, it is unlikely that the whole answer lies in any one example.

¹ TES magazine, ‘Test century – a lasting obsession with exams’, 10 December 2010.

² Discussed in the section on ‘Accountability’ below.

We are very conscious that recommendations of this review will need to be considered in relation to the National Curriculum, which is currently also under review. Statutory assessment and future programmes of study will need to be developed so that they complement each other, particularly given the significant concern that statutory assessment at the end of Key Stage 2 currently narrows the primary school curriculum.

Before we move on to covering the evidence received so far, it is worth saying at this stage that **statutory tests are only part of the assessment landscape**. We realise that assessment routinely takes place in the classroom as teachers watch, listen and interact with pupils, and testing can only ever be a snapshot of performance at one particular moment. Evidence highlights the importance of 'Assessment for Learning'³, and we are aware that schools have plenty of expertise about how best to use it. We want the assessment that happens in the classroom on a day to day basis to be led by schools and individual teachers, and we do not want to dictate how this should be done. It is therefore not the role of this Review to consider ongoing non-statutory assessment practices in schools. However, this report will look at the role of summative teacher assessment at the end of Key Stage 2 and the arguments that have been made for it to have a greater role in the accountability system.

³ Black, P., Harrison, C., Lee, C., Marshall, B., and Wiliam, D., *Working Inside the Black Box*, Kings College London, (2002).

Overview of the current system

The system of national curriculum testing in England has undergone significant change since its introduction in the early 1990s. Of the original National Curriculum Tests at Key Stages 1, 2 and 3, only Key Stage 2 tests remain a statutory requirement. These tests are frequently referred to as “SATs”, reflecting the ‘standard assessment tasks’ originally intended to accompany the 1989 National Curriculum; while these were replaced with written tests, the name has persisted in popular use. For consistency, this review refers to National Curriculum Tests.

In England, statutory assessment is made in relation to national eight-level scales (level descriptions) for each subject. Pupils are expected to attain level 4 at the end of Key Stage 2. In 2010 20% of pupils did not achieve this level in English and 21% did not achieve this level in maths, while 33% of pupils achieved level 5 in English and 34% achieved level 5 in maths. While statutory assessment at the end of Key Stage 2 is measured in whole levels, many schools use sub-levels to provide more fine-grained tracking of their pupils.

At present, all children at the end of their final year in Key Stage 2 (Year 6) sit statutory externally set and marked National Curriculum Tests in English (reading and writing) and maths. A sample of schools (5%) also sits an externally set and marked science test. In addition, schools are required to make statutory teacher assessment judgements in English (reading, writing and speaking & listening), maths and science.

Individual pupils’ test results and teacher assessment judgements are reported to parents and carers, while schools’ results are published in national Achievement and Attainment Tables. These tables also include information about school size, numbers and percentages of pupils with Special Educational Needs, year on year comparisons, contextual value added scores, progress measures based on percentages making expected progress, average points score results and absence records. The Achievement and Attainment Tables are frequently used by the media to create ‘league tables’ ranking schools based on their test results.

A wide range of school performance data is captured in RAISEonline, a data-management system only accessible to the school, its local authority and Ofsted. As well as accountability to school managers, this data is used by local authorities (for example through monitoring by School Improvement Partners) and Ofsted’s inspections. The Government has announced its intention to reform the landscape of school accountability from September 2011. For example, schools will no longer be required to set performance targets.

Purposes of statutory assessment

Almost all respondents have at some point questioned the purposes of statutory assessment. There seems to be widespread concern that there are too many purposes, which can often conflict with one another. 71% of online consultation respondents believe strongly that the current system does not achieve effectively what they perceive to be the most important purpose.

Dr. Newton’s work in this area is perhaps the most well known⁴. He has identified at least 16 purposes for which statutory assessment has been widely used. He cautioned that the current

⁴ Newton, P.E., ‘Clarifying the purposes of educational assessment’, in *Assessment in Education: Principles, Policy & Practice*, 14 (2), 149-170 (2007); Newton, P.E., *Evaluating assessment systems, Paper 1: Submission to the*

test results are better suited for some of these purposes than for others. He suggested that assessments should be designed with a clear purpose in mind and that there should be caution in making inferences which the data was not intended to support.

Many other respondents share a similar view. The House of Commons Children, Schools and Families Select Committee's 2008 report on Testing and Assessment⁵ recommended that the use of national testing for multiple purposes – including measuring pupil attainment, school and teacher accountability, and national monitoring – should be stopped.

The Expert Group on Assessment⁶ identified four key purposes for the assessment system as a whole (encompassing both statutory summative assessment and ongoing formative assessment):

- “to optimise the effectiveness of pupils’ learning and teachers’ teaching;
- “to hold individual schools accountable for their performance;
- “to provide parents with information about their child’s progress; and
- “to provide reliable information about national standards over time.”

One of the key early decisions for this Review will be to define clearly the purposes of statutory assessment in primary education and what we want it to achieve in the future. The system can then be designed to serve these purposes.

Benefits and deficits of the current system

We have heard a great deal of evidence about the positive and negative aspects of the current testing, assessment and accountability systems. There is considerable disagreement and great strength of feeling. There is clearly support for parts of the current system, but significant concern that its balance and overall impact leads to negative consequences.

There is strong evidence that external school-level accountability can be important in driving up achievement and progress⁷. Professor William suggests that the effect of external assessment regimes is equivalent to an increase in the rate of learning of approximately 20%⁸, though he cautions that this may be counteracted by the effects of perverse incentives in the system. Professor Sammons notes that the combination of policy attention, resources and considerable support and professional development, plus the three related components of national assessment, national inspection and a national curriculum, has had a positive impact on raising standards and reducing the number of poorly performing schools in England. She believes this effect has been stronger in England than in Wales and Scotland, which have given less attention to accountability⁹.

Tim Oates observes that “assessment can drive learning in a beneficial manner”¹⁰, recognising the importance of statutory assessment as one of a number of ‘control factors’ which affect

Education and Skills Select Committee Inquiry into Testing and Assessment, Qualifications and Curriculum Authority (2007).

⁵ House of Commons Children, Schools and Families Select Committee, *Testing and Assessment: Third Report of Session 2007-8*, HC 169-I (2008).

⁶ *Report of the Expert Group on Assessment*, DCSF (2009).

⁷ Discussed further in the section on Accountability.

⁸ William, D., ‘Assessing achievement at the end of KS2’, (2010).

⁹ Sammons, P., ‘Zero tolerance of failure and New Labour approaches to school improvement in England’ in *Oxford Review of Education* 34:6 (2008).

¹⁰ Oates, T, *Could do better – Using international comparisons to define the National Curriculum in England*, (2010).

classroom practice. He outlines some key benefits of the National Curriculum and its associated assessment, including raising attainment¹¹, progression¹² and teachers' expectations¹³; improving transfer through a common structure¹⁴; and identifying issues such as the Key Stage 3 'dip'¹⁵.

Good academic outcomes do not have to come at the expense of narrowing the curriculum. There is evidence¹⁶ to suggest that very good primary schools are successful both in promoting pupils' academic progress, achieving benchmark attainment levels for most children at the end of Key Stage 2, and in ensuring high quality educational experiences across the curriculum.

We have heard strong arguments that the information generated by the current system informs policy development, is useful in determining the support needed by schools, and helps determine trends in children's learning. Some feedback suggests that many pupils enjoy the Key Stage 2 tests because they are able to show what they can do, and many parents value the information the system currently provides. There is also an argument that the current system focuses attention on Year 6, which can make up for poorer rates of progress earlier in Key Stage 2.

When asked to identify aspects of the current system which worked well and should be retained, respondents to the online call for evidence identified the following:

- 7% of online consultation respondents said that the current system provided national standards for comparison and consistency, noting that standard tests allowed schools to be compared nationally; allowed schools to identify the needs of children including high achievers and those children who were underachieving; enabled schools to maintain their accountability to public funding; and that external tests did not allow schools to inflate their results.
- 6% of online consultation respondents felt that externally marked tests which changed every year were important and valuable. Respondents also believed that external tests allowed schools to check how valid their own teacher assessments were. It was noted that testing still remained a more time-efficient method of assessing than continuous teacher assessment.

Despite these benefits, few would disagree that the current system of testing, assessment and accountability has led to undesired consequences. In particular, the increased focus on National

¹¹ Chitty, C., *Educational Policy in Britain* (2004); Colwill, A. & Peacey, N., 'Planning, teaching and assessing the curriculum for pupils with learning difficulties: curriculum guidelines to support the revised National Curriculum' in *British Journal of Special Education* 28.3, 120-122 (2003).

¹² Chitty, C. op cit; Tymms, P., 'Are standards rising in English primary schools?', in *British Educational Research Journal* 30(4), 477-494 (2004); Whetton et al, *Standards in English Primary Education: the International Evidence*, Cambridge Primary Review Research Survey 4/2, (2007).

¹³ Barber, M., *Crossing the bridge*, AAIA (2002); Hopkins, D., *School improvement for real* (2001); Tabberer, R., *Primary Education: expectations and provision*, NFER (1997).

¹⁴ Dobson, J. & Pooley, C.E., *Mobility, Equality Diversity: a study of pupil mobility in the secondary school system*, University College London (2004); Strand, S., 'Pupil Mobility, Attainment and Progress During Key Stage 1: a study in cautious interpretation' in *British Educational Research Journal* 28.1, 63-78 (2002).

¹⁵ Powell, R., Smith, R., Jones, G., Reakes, A., *Transition from Primary to Secondary School: Current Arrangements and Good Practice in Wales*, NFER (2006); Doddington, C., Flutter, J. & Rudduck, J., 'Exploring and explaining 'dips' in motivation and performance in primary and secondary schooling', in *Research in Education* 61, 29-38 (1999).

¹⁶ Ofsted, *Twenty outstanding primary schools: excelling against the odds*, HMSO (2009).

Curriculum Test results has created a situation where the curriculum in some schools has been narrowed to meet external targets¹⁷.

Ofsted's 2009 Annual Report¹⁸ said that "*in some schools, over-emphasising preparation for the national tests in English, maths and science, especially in Year 6, restricts the time available for activities that can most interest and challenge pupils: speaking and listening in English, using and applying maths and scientific investigation. Pupils' attainment can then become narrowly based*".

The Cambridge Primary Review Team said "*as children move through the primary phase, their statutory entitlement to a broad and balanced education is increasingly but needlessly compromised by a 'standards' agenda which combines high stakes testing and the national strategies' exclusive focus on literacy and numeracy... The most conspicuous casualties are the arts, the humanities, and those kinds of learning in all subjects which require time for talking, problem solving and the extended exploration of ideas*"¹⁹.

When asked to identify aspects of the current system which most need to improve, respondents to the online call for evidence identified a range of criticisms:

- 50% of respondents wanted league tables removed, suggesting that their impact had skewed the whole test regime and yet reflected very little about a school.
- 47% of respondents felt that the current system encouraged schools to 'teach to the test' and narrowed the curriculum. They said that pressure on schools to improve or maintain their position in the league tables meant children were simply taught what they would be tested on, missing out on a rich and varied curriculum in their final year at primary school.
- 36% of respondents noted that the tests (and the pressure for schools to maintain their position in league tables) placed too much stress on children and that many did not perform well in test situations, especially where so much material was tested in a short time.
- 25% of respondents were concerned that the quality of marking of tests was either poor or variable, and felt the test results took too long to be returned.
- 11% of respondents said that, as so much depended on how well the school did in the tests, teachers felt pressurised into doing all they could to make sure the children performed well, sometimes at the expense of a rich and varied curriculum. School staff, especially school leaders, suggested that a single set of bad test results could damage their careers or the reputation of their schools, damaging morale.

We have also noted research evidence which suggests that the National Curriculum and its assessment leads to the marginalisation of certain subjects²⁰, an adverse effect on teaching and learning²¹, the 'cramming effect'²², demotivation and dysfunctional behaviour for some

¹⁷ Earl, L., Watson, N., Levin, B., Leighwood, K., Fullan, M. and Torrance, N, *Watching and Learning 3: Final Report of the External Evaluation of England's National Literacy and Numeracy Strategies*, Ontario Institute for Studies in Education (2003).

¹⁸ Ofsted, *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2008/2009*, HMSO (2009).

¹⁹ Alexander, R. (ed), *Children, their World, their Education: final report and recommendations of the Cambridge Primary Review* (2009).

²⁰ Rawling, E., 'The rises and falls of geography', in *Teaching Geography* 26.3, 108-11 (2001).

²¹ Alexander, R. (ed), *Children, their World, their Education: final report and recommendations of the Cambridge Primary Review* (2009).

²² Rose, J., *Independent Review of the Primary Curriculum: Final report*, DCSF, (2009).

pupils, particularly those that are lower achieving²³, and has perhaps lowered primary school children's enjoyment of maths²⁴. Other evidence argued that 'high stakes' testing may also prompt pupils to concentrate not on how much they had learnt but how well or badly they have done in terms of National Curriculum levels, which can reduce their learning potential and have a detrimental effect on educational outcomes²⁵.

We will now go on to explore the evidence according to the key themes that have emerged during the review so far.

²³ Robinson, C. and Fielding, M., *Children and their Primary Schools: pupils' voices*, Cambridge Primary Review Interim Report 5.3 (2007).

²⁴ Ruddock et al, *Where England Stands in the Trends in International Mathematics and Science Study (TIMSS) 2003: National Report for England*, NFER (2004).

²⁵ Harlen, W. and Crick, R., 'Testing, Motivation and Learning', Nuffield Foundation, University of Bristol (2002).

Accountability

While there is considerable debate about how schools should be held accountable, the evidence submitted to the review overwhelmingly accepts that, as public bodies, schools should be held accountable for the education of their pupils. This is an important point of principle.

There is strong evidence that external school-level accountability is important in driving up achievement and improving pupils' progress. The OECD has concluded that a 'high stakes' accountability system can raise pupil achievement in general and not just in those areas under scrutiny²⁶.

Other research suggests that systems in which local authorities, schools, teachers and, ideally, pupils are held accountable for their academic performance are associated with increased pupil achievement. A recent review in this area²⁷ concluded that the effect of such assessment regimes could be equivalent to an increase in the rate of learning of approximately 20% and suggested that 'high stakes' accountability systems are probably the most cost-effective method for raising achievement yet developed.

Research by Professor Sammons suggests that schools vary in their effectiveness in promoting pupils' academic results²⁸ and that such differences in academic effectiveness of schools tend to matter more for disadvantaged pupils²⁹. She has illustrated the benefits on pupils' attainment and progress of going to a more academically effective primary school³⁰. Given the variation in school effectiveness and its impact on attainment and progress, especially for the most disadvantaged, Professor Sammons suggests that having one externally set and marked summative assessment of the academic effectiveness of primary schools is a very important method of studying the variation in pupils' attainment across the country and between schools. It also provides an independent measure of each pupil's attainment, which is especially important in identifying pupils with low attainment who may need extra support as they move to secondary school.

Some international evidence suggests that 'high stakes' testing improves standards and can ensure a high degree of engagement with the system. For example, Tim Oates observes that in Singapore "*high stakes primary assessment for allocation to secondary schools plays an important role*" alongside other measures designed to focus teaching practice³¹. In this case, 'high stakes' testing encourages teaching to the test, which actually means teaching a broad curriculum.

In his evidence to the review, Sir Michael Barber argued that if schools (and government) are to make decisions based on evidence, they need regular assessment and data which are

²⁶ *PISA 2009 Results IV: What Makes a School Successful? Resources, Policies and Practices*, OECD (2010).

²⁷ William, D., 'Assessing achievement at the end of KS2' (2010).

²⁸ Sammons, P., *School effectiveness and equity: making connections*, CfBT (2007).

²⁹ Sammons, P., 'Equity and Educational Effectiveness', in Peterson, P., Baker, E., McGaw, B. (eds), *International Encyclopedia of Education* vol.5, 51-57, Oxford (2010).

³⁰ Sammons et al, 'Influences on Children's Cognitive and Social Development in Year 6', EPPE, DCSF Research Brief (2008); Sammons et al, 'Children's Cognitive Attainment and Progress in English Primary Schools During Key Stage 2: Investigating the potential continuing influences of pre-school education', *Zeitschrift für Erziehungswissenschaft* 10, 179-198 (2008).

³¹ Oates, T, *Could do better – Using international comparisons to define the National Curriculum in England*, University of Cambridge (2010).

comparable over time. This enables analysis of the performance of different groups of pupils. This therefore requires valid, reliable and consistent data³².

Much of the evidence received suggests that the key concern with the current testing system is that it is ‘high stakes’ for schools. Since National Curriculum Tests are designed to measure not only the achievement of students, but are also used for accountability purposes, there is considerable pressure on schools to perform well. Such pressure is often communicated to pupils or leads to behaviours such as ‘teaching to the test’.

62% of online consultation respondents had concerns about the way in which National Curriculum Test data is used. The following quotations from the online consultation reflect the views of the majority of respondents about accountability.

“There is support in the teaching profession for testing. It is how the crude test data under the current system is used that is the source of concern for the profession.”

“Abolish league tables. By giving so much weighting to the tests their importance is inflated. Summative snapshot judgements don’t tell the whole story and formative assessment is a better indicator of progress.”

“As a teacher at a school in challenging circumstances, I feel that the current system is unfair. I work extremely hard all year for all my pupils, yet the exceptional progress that some of them make during their school years is not taken into account, as they do not meet national expectations and often do not perform at their best under pressure. As a teacher, I dread KS2 tests results arriving, as I know if results are not as good as expected, my teaching immediately comes into question and my hard work is not recognised”.

“All schools have their own character; comparison purely on end of KS2 results does not present a true picture of the school and its achievements. A school is a sum of its parts and needs to be judged on all its attributes, not a snapshot in time.”

A significant number of respondents, including the main teaching unions, have argued strongly against the ‘high stakes’ nature of current statutory assessment. Many respondents feel that schools should be held accountable only for factors directly within their control, in particular the contribution they make to their pupils’ education. NUT, NAHT, and ATL believe that test results are subject to unnecessary and damaging over-interpretation, and undermine pupil learning³³. They argue that England’s accountability system is excessive, confused and overlapping, placing too much onus on schools being accountable to central government at the expense of other stakeholders. The NASUWT believes that it is the uses to which test results are put that causes the problems. NASUWT advocates a developmental and supportive accountability system, where the focus is on improving the quality of teaching and learning and helping both schools and individuals to improve, concentrating on achievements rather than weaknesses.

NGA expressed concern at how responsive the school system has become to performance tables. Many heads report they “have to” teach to the test or else results will drop, in spite of clear evidence that it is a poor strategy. This suggests that the dominance of National Curriculum Tests has become ingrained in school culture, which can damage schools’ ethos. AAIA criticises the current system of accountability for relying primarily on data from the ‘one off’

³² Mourshed, M., Chijioke, C., Barber, M., *How the world’s most improved school systems keep getting better*, McKinsey (2010).

³³ ATL, NAHT, NUT, *Common Ground on Assessment and Accountability in Primary School* (2010).

Key Stage 2 tests. ASCL believes league tables are driving the whole education system, leading to assessment for its own sake rather than to identify a pupil's teaching needs.

Some academic experts have blamed the 'high stakes' nature of the system for the current concerns of narrowing the curriculum. For example, Professor Torrance says "*the more individual student achievement is tied to system accountability the more accountability measures will dominate student experience*"³⁴. Professor Tymms and Dr. Merrill, Lord Sutherland, NUT and SCORE felt strongly that the use of test data to compile league tables has led to excessive test preparation which is constraining for schools and pupils.

Many respondents either argue against publishing school-level performance data which could be used to create league tables, or stop short of suggesting this but strongly highlight some of the perceived problems that league tables cause.

50% of online consultation respondents argued that league tables should be abolished, or suggested that test results should not be used for compiling tables. Online consultation respondents often commented that, even if a school had outstanding teaching practice, a lower place in the tables compared to other local schools could mean fewer parents selecting the schools for their children as they might see it as inferior. The criticism of the published school-level performance data is summed up by the House of Commons' 2010 Education Select Committee report³⁵, "*the Achievement and Attainment Tables present a very narrow view of school performance and there are inherent methodological and statistical problems with the way they are constructed. For instance they are likely to favour independent and selective schools, which have a lower intake of deprived children or of children with Special Educational Needs.*"

Warwick Mansell, Mary James and Dr. Newton have all highlighted the drawbacks of publishing school-level data³⁶, arguing that the public is likely to make inferences beyond the uses for which the current test results are designed. Mansell and James argue that any published assessment results should be accompanied by a 'health warning' "*that they measure only part of a school's business and that these results are influenced by factors beyond the school's control, such as the prior attainment and background characteristics of its pupil population*".

Test results also impact on Ofsted judgements – potentially adding to the 'high stakes' nature of the current system.

15% of those responding to the online consultation have commented on the impact of test results on Ofsted's inspection judgements, some expressing deep concern. NUT, NAHT and ATL believe that Ofsted places an undue emphasis on test results as a proxy for school effectiveness³⁷, despite their limitations. They argue that the inspection system should take greater account of the whole school's achievements. The House of Commons Select Committee's 2010 Report on school accountability³⁸ concluded "*when evaluating academic attainment, we recommend that Ofsted gives less evidential weight to test results and derivative measures and gives more weight to the quality of teaching and learning observed by inspectors*

³⁴ Torrance, H., *Using assessment in education reform* (2009).

³⁵ House of Commons, Children, Schools and Families Committee, *School Accountability: First Report of the Committee, Session 2009-10*, HC 88-1 (2010).

³⁶ Mansell, W., James, M. and Assessment Reform Group, *Assessment in schools: Fit for purpose? A Commentary by the Teaching and Learning Research Programme* (2009).

³⁷ ATL, NAHT, NUT, *Common Ground on Assessment and Accountability in Primary School* (2010).

³⁸ House of Commons, Children, Schools and Families Committee, *School Accountability First Report of Session 2009-10*, HC88-1 (2010).

in the classroom". Ofsted's response to the Committee confirmed that such measures had already been introduced in its 2009 inspection framework³⁹.

There is widespread agreement about the importance of providing good quality information to parents. 57% of online consultation respondents said that it was important that parents were informed of their children's progress in particular.

Research from four surveys⁴⁰ shows that the majority of parents of primary age pupils value the information from external tests. Survey respondents also wanted information about the performance of primary schools to be available to the public, and placed value on National Curriculum Tests in providing information about the how their child and their child's school is performing.

The National Confederation of Parent Teacher Associations (NCPTA) survey in 2008 shows a strong majority of parents (78%) placed a high or medium value on their children taking external tests. Only a minority of parents (21%) said that they are of no use to them.

The DCSF Ipsos MORI survey of 939 parents in England with children of school age in March 2009 found that 75% of parents think information on the performance of primary schools should be available to the general public. 70% of parents placed value on the tests in providing information about how their child's school is performing. 65% of parents placed importance on their child taking part in Key Stage 2 tests. 44% of parents thought that tests should stay as they are. 36% of parents wanted to replace tests, but of those, most felt that some form of testing or assessment should remain.

When consulting between 2008-2009 on the previous Government's School Report Card proposals⁴¹, a DCSF survey found that the majority of parents (91%) and the public (84%) said it was very important to them personally to know how well each school performs. The majority of parents (87%) and the public (82%) agreed a lot or a little that test and exam results are one important measure of a school's performance. In addition to test results, parents and the public would most like information on behaviour at each school (parents 65%/public 56%). The next most popular kinds of information were attendance rates and how well pupils progress during their time in the school.

In contrast, a 2009 NAHT survey⁴² indicated that the majority of parents surveyed by the union did not favour external testing. The survey suggested that parents instead value information such as attendance rates and behaviour at their child's school or prospective schools.

Many respondents have expressed opinion about the information parents find valuable and whether the current system provides it.

59% of online consultation respondents, predominately head teachers and teachers, said that more weight should be given to teacher assessment. They said that, rather than being provided with a snapshot test result, parents would welcome the more comprehensive information that

³⁹ House of Commons, Children, Schools and Families Committee, *School Accountability: Responses from the Government and Ofsted to the First Report of the Committee, Session 2009-10*, HC 486 (2010).

⁴⁰ *The Parent Perspective: SATs and National League Tables*, National Confederation of Parent Teacher Associations (NCPTA) survey October 2008; *Survey of 936 parents in England with children of school age*, DCSF Ipsos MORI March 2009; *21st Century Schools: A World-Class Education for Every Child; A School Report Card: consultation document. Analysis of responses to the consultation documents*, House of Commons (2009).

⁴¹ *21st Century Schools: A World-Class Education for Every Child; A School Report Card: consultation document. Analysis of responses to the consultation documents*, House of Commons (2009).

⁴² NAHT survey of 10,465 parents (January/February 2009).

could be provided from teacher assessment. Online consultation respondents believed that accurate information based on teacher assessment should be given throughout the child's education and not just at the end of Key Stage 2. Many proposed that this information should be given with a clear indication of how individual children were progressing against age-related expectations.

A range of stakeholders have given their views about what information parents find useful. NGA reported that some parents like external accountability and believe Key Stage 2 tests are important. However parents sometimes apply additional pressure on their children to perform well, for example setting additional work to be undertaken at home, contributing to pupils being 'over-rehearsed'. NAHT said that "*parents, politicians and the public have a right to see information about the performance of their children and their schools. We just don't believe the raw data they get from the current league tables is relevant or helpful*".

Some respondents have argued that the information currently provided to parents is confusing and difficult to interpret. Dr. Newton considers that the multiple purposes of testing can cause confusion for parents when trying to make use of the results, in particular when comparing potential schools for their child. ARK does not feel that test results are sufficiently useful in providing information to pupils and parents about individual performance, as the information comes too late for any parent/school discussion to help improve pupil performance, and the measurement is not sufficiently reliable at pupil level. Some of those who responded to the online consultation felt that end of year reports had become too technical and, despite the best efforts of schools, many parents did not fully understand the issues of levels and progress.

Suggested changes to the accountability system have focused around broader accountability measures.

Suggestions for broader measures have included results from other subjects or other year groups alongside Key Stage 2 English and maths. Ofsted inspection reports, pupil attitude surveys and methods of getting feedback from parents have also been mentioned. 66% of online consultation respondents want surveys and feedback from parents to be used. 61% want a focus on Ofsted inspection reports. Many respondents feel these measures would help broaden the curriculum across primary school by reducing the incentive to focus narrowly. For example, NFER argues that broadening schools' accountability measures would have a positive 'backwash' encouraging broader teaching in Year 6 and earlier in Key Stage 2.

A greater emphasis on measures of pupil progress has been proposed, especially by NAHT, who argue that, while straightforward achievement is frequently linked to a pupil's background, the progress they make reflects the contribution made by their school. Two schools may achieve the same attainment results, one from an affluent area and one from a disadvantaged area. In such a case pupils in the disadvantaged area may have made much more progress; progression therefore acts as a proxy measure for a school's overall quality of teaching. NAHT believes that primary schools contribute a great deal of social and personal development (which some schools need to focus much more on than others) which is not acknowledged by the current accountability framework.

Several respondents proposed that published performance data should be based on three or five year rolling averages – which might reduce some of the effect on results of individual pupils or cohorts and provide a sense of achievement over time. This data is already used in RAISEonline but could be made public. NAHT and NGA in particular argue that rolling averages would reduce the risk averse behaviour such as 'teaching to the test' and aversion to innovation that annual results appear to encourage.

Some suggest that summative teacher assessment judgements should continue to be reported alongside test results, as was the case in 2010. However, many respondents felt that tests will always ‘trump’ teacher assessment and carry greater weight. At present the deadline for schools to submit teacher assessment results falls after the return of National Curriculum Test results – though in 2010 37% of schools had submitted their summative teacher assessment ahead of receiving test results on 6 July.

National Sampling has been suggested as an alternative approach for Government to measure national standards.

A number of respondents (including NAHT, ASCL, NUT, ATL, AAIA, the Cambridge Primary Review and SCORE) have suggested introducing a national sampling system to provide Government with information on the performance of schools and pupils. They advocate national sampling as a means of generating robust national statistics alongside employing summative teacher assessment for school-level accountability.

Proposals for national sampling differ in their detail, and would offer varying degrees of information. However, it is clear that tests administered to a carefully-selected small sample of pupils could enable robust inferences about performance of the national cohort. Based on the experience of the Assessment of Performance Unit’s sampling programme in the 1980s, a sample of 10,000 pupils would be sufficient, though robust information on sub-groups of the population would require a larger sample⁴³. Importantly, a test designed to report on the national cohort need only sample one or two pupils per school and may not therefore provide meaningful school-level results.

Some proponents of national sampling question whether the current system gives us an accurate picture of national trends, and therefore suggest that sampling could produce results which are at least as reliable. For example, the Cambridge Primary Review team believe that at a national level, the assumption that aggregating individual pupils’ test results enables trends in attained standards to be identified, is problematic. The team argued that the use of National Curriculum Test results for national monitoring means that we have “*extremely limited information, collected under stressful conditions, which provides little useful data about national levels of performance and even less about how to improve them*”⁴⁴.

⁴³ ‘Submission to the Expert Group: Issues to Consider When Developing a National Monitoring System’, NFER (2009).

⁴⁴ Alexander, R. (ed), *Children, their World, their Education: final report and recommendations of the Cambridge Primary Review* (2009).

The role of summative teacher assessment judgements

Firstly, it is clear that there is no single definition of teacher assessment. The term ‘teacher assessment’ can have very different meanings in different schools, and ‘teacher assessment’ judgements can be based on a wide range of evidence. It is typically an ongoing activity, integral to good teaching and learning and not a tightly-defined process. It is clear that teacher assessment routinely plays two roles in schools. It can be:

- *formative*, whereby teachers identify pupils’ learning needs and what is required to support them;
- *summative*, where teachers reach professional judgements about what a pupil has learned and/or what level they have reached.

There is universal support for the importance and appropriateness of formative teacher assessment for the purpose of supporting teaching and learning of each pupil throughout Key Stage 2. The review has heard much about the value of formative teacher assessment (or ‘assessment for learning’). It is clear that teachers and head teachers see formative assessment as fundamental to teaching and learning, and place it at the heart of each school’s policy and practice. **While we recognise and celebrate the place of formative teacher assessment, it is outside the scope of this review.**

The debate about the emphasis on teacher assessment begins when considering other uses of summative assessment, including how best to ensure schools are properly and fairly accountable to pupils, parents and the taxpayer.

Many respondents advocate a system in which summative teacher assessment judgements are used as the basis for accountability. The majority of online consultation respondents (predominantly primary teachers and heads) supported this approach. 83% think schools’ own teacher assessment judgements should form the main information used to compare schools and hold them accountable. Views become stronger when particular parts of the curriculum are considered. For example, 90% of online consultation respondents think that there are aspects of English (such as speaking and listening) which should only be assessed by teachers.

The following quotations from the online consultation reflect the views of the majority of respondents about teacher assessment and its role.

“Teacher assessments provide a much more accurate assessment of a pupil’s achievement at the end of KS2 based on their progress through KS2 and not just on a single day when tests are administered.”

“Teacher assessment is extremely rigorous in our school and we feel that this should be used to get an overall picture at the end of KS2. It could be administered more like KS1 - the test can happen at any point during the year and this can then identify weaknesses. Teacher assessment has a much greater significance. Teachers can evidence work and then this can be moderated throughout the [local] authority to ensure the same standards are being adhered to.”

However, it should be noted that some online consultation respondents argued in favour of whole cohort external testing and said they would not want to see it replaced by teacher assessment. The following quotations reflect views held by a minority of online consultation respondents.

"I know of children who felt cheated and let down by their schools when SATs were boycotted last year. It is their time to shine. The children want to show what they can do. It's their cup final. All of the children start secondary school with a test result that is fair and honest. Teacher assessments are just teacher marked assessments. I'd prefer to continually assess and have my teacher assessment validated by externally marked tests."

"Don't have teacher assessment, it is time consuming and cannot be judged against a common baseline."

Some respondents, including several teaching unions (NAHT, NUT, ATL, ASCL), set out proposals for how summative teacher assessment judgements could be used as the basis for school accountability. However, they accept that, for this to be robust, it would need to be backed up by moderation, either across groups of schools or test-based (statistical) moderation involving a sample of pupils in each school.

In their joint policy statement⁴⁵, NAHT, NUT and ATL argue that *"teacher assessment, when it is subject to proper external checks and is carried out by well-trained professionals, has clear advantages over national curriculum testing... it is a more valid form of assessment, because a much wider range of pupils' learning, over a much longer time period, can be evaluated by the teacher than is possible through a few short, one-off, tests"*. They accept the potential for errors of judgement (citing the risk of bias on ethnic lines) and propose a process of moderation that is *"teacher-led and locally organised, but accompanied by a resource which supports national standardisation: a national bank of assessment materials from which teachers can choose to draw to check their assessments"*. The joint statement further suggests introducing a system of 'low stakes' sample testing to monitor national standards over time.

In their evidence to the review, NUT and ATL expanded on these proposals. NUT felt that, as long as teachers are empowered to be good assessors, there is no reason why they would be any more inaccurate than the current system with its "proven inaccuracies". However, ATL noted that using summative teacher assessment alone could lead to teachers becoming personally responsible for their school's ranking.

NAHT similarly proposed freeing schools to choose their own approaches to summative teacher assessment. They accepted the need for results to be demonstrably robust, validated by moderation and sample testing of pupils in each school. However, NAHT accepted that, should Ofsted judge a school's assessment system to be inadequate, external tests should be reintroduced until identified improvements had been made. NAHT argued that, as the 2009 HMCI report⁴⁶ shows that 92% of primary schools inspected (and/or interim assessed) are satisfactory or better, the majority of schools could be trusted to have robust assessment. The safeguards provided by inspection and external moderation would identify the schools requiring additional support.

In its evidence to the panel, ASCL recommended a combination of teacher assessment (based on nationally exemplified standards, accompanied by an agreed moderation process) and tests chosen from a bank (but marked by a test agency) which could be administered at any point during Year 6. They stressed the need for robust moderation so that secondary schools had confidence in the results (which provide the baseline to measure progression at GCSE). Schools could have a chartered educational assessor to train staff and ensure that proper

⁴⁵ ATL, NAHT, NUT, *Common Ground on Assessment and Accountability in Primary School* (2010).

⁴⁶ Ofsted, *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2009/10*, HMSO (2010).

moderation processes were carried out. Cross-phase moderation could be used between primary and secondary schools to help improve secondary teachers' trust in the data and their understanding of the primary curriculum.

AAIA likewise supports a model of summative teacher assessment judgements based on statutory assessment tasks and internally-marked tests, modelled on practice at Key Stage 1, which it believes has allowed teachers to develop as effective assessors. Moderation should be enforced through local authorities and an external accrediting body established to ensure consistent judgements are maintained.

However, NASUWT strongly disagrees with this approach, warning that simply substituting externally-moderated teacher assessment for tests would just change the data available without addressing the fundamental issues they identify in the accountability system. Based on the experience of their members in Wales, NASUWT argues that such a move would only to serve to increase the workload on teachers.

Different approaches have been proposed for how teacher assessment could be moderated to inform robust school-level accountability.

Suggested models for moderation fall into two main categories. Firstly, external moderation involves the evidence for teachers' judgements being scrutinised by a different professional. This approach is currently used at Key Stage 1, where local authorities moderate a sample of pupils in 25% of their schools. Suggested external moderators include: teachers from different schools, specialist local authority staff, accredited chartered assessors, or moderation across groups of schools. ASCL and Professor Brown suggested that cross-phase moderation could be used between primary and secondary schools to develop their assessment skills and to enable primary and secondary schools to build a common understanding and respect for assessment judgements. It could also help improve secondary teachers' trust in the data and their understanding of the primary curriculum.

Secondly, test-based moderation has been suggested. Proposed models rely on: using whole cohort tests to moderate summative teacher assessment judgements for each pupil; having a sample of pupils sit tests to gauge the overall accuracy of teacher assessment; or using tests to generate an 'envelope' of results to help frame teacher assessment judgements. Test-based moderation assumes that pupils will perform similarly in tests (which capture a snapshot of achievement) and in teacher assessment (which draws on a much broader base of evidence).

It has been suggested that the reliability of teacher assessment judgements could also be increased by including the use of highly structured and focused assessment tasks conducted in the classroom⁴⁷. Dr. Isaacs suggested that a mixture of externally-set tests together with internally-assessed tasks (with external moderation if necessary) could provide robust moderation. However, if carried too far, heavily prescribed assessment tasks risk compromising teachers' freedom and flexibility to select from a wide range of evidence over a period of time.

Many respondents argued that a greater focus on teacher assessment would need to be matched by additional professional development. Ofsted's 2010 HMCI report⁴⁸ shows that "assessment continues to be an area in which schools need to improve" with only 53% of schools being judged to be good or outstanding. Some respondents have argued strongly for a

⁴⁷ Torrance, H., *Using assessment in education reform* (2009).

⁴⁸ Ofsted, *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2009/10*, HMSO (2010).

greater focus on teachers' professional development to improve teacher assessment, and have suggested particular ways this could be achieved.

For some, placing greater weight on summative teacher assessment judgements would give teachers the opportunity to develop their assessment to a higher standard. 8% of online consultation respondents (overwhelmingly primary teachers or heads) felt that external tests undermined the trust and professionalism of teachers, showing a lack of faith in teachers' professional judgement.

The NUT, NAHT, ASCL, CIEA and Lord Sutherland believe that all schools should have access to an accredited assessor to improve the quality of teacher assessment. ASCL suggested that schools could have a chartered educational assessor who would train staff and ensure that a proper moderation process was being carried out.

Lord Sutherland argued that building confidence and strengthening teacher assessment cannot be achieved only through putting moderation in place; it also requires building teachers' skills and providing support for assessment across schools (the CEIA have suggested that every school would benefit from having a lead assessor). The Cambridge Primary Review advised that high quality CPD was needed to make schools more confident in using a range of methods for assessing pupils more accurately.

However, some research evidence suggests that teacher assessment must be used cautiously and may not provide sufficiently reliable data for school accountability, even with external moderation or testing a sample of pupils.

Professor Wiliam cautioned that, in his view, experience over the last ten years does not suggest that moderation would make summative teacher assessment sufficiently reliable to be used for school-level accountability. However, he suggested that tests could be used to create an 'envelope' of levels which teachers could assign to pupils based on their ongoing assessment (reflecting teachers' accuracy in ranking pupils by ability).

Discussion at a seminar organised in 2009⁴⁹ by NFER raised an important point on the credibility of teacher assessment:

"It was accepted that, by definition, an expert teacher is also an expert assessor. However, teachers themselves may not believe this, nor may other stakeholders. A key step in the wider acceptance of teacher assessments is to develop confidence in the system. There is a general concern about reliability and bias when using teacher assessment with research evidence about the bias in individual cases. As one delegate pointed out, it only needs one well placed article questioning reliability to undermine confidence. Currently, if a test score and a teacher assessed score are provided, many people would give more credence to the test score. How do you start to change this perception? The test and the teacher assessment may be measuring different things and these need to be recognised."

Professor Harlen's 2004 paper⁵⁰ reviewed 30 research studies into summative teacher assessment judgements. One of the findings suggests that a way of making teacher

⁴⁹ Parkes, C., and Maughan, S., *Policy and research seminar on Methods for Ensuring Reliability of Teacher Assessments: Proceedings*, NFER (2009).

⁵⁰ Harlen, W., *A Systematic Review of Evidence of the Impact on Students, Teachers and the Curriculum of the Process of Using Assessment by Teachers for Summative Purposes in Research Evidence* EPPI-Centre, Social Science Research Unit, Institute of Education (2004).

assessment more effective for summative purposes is to make the criteria and processes much tighter and more prescriptive: “*the reliability of portfolio assessment where tasks were not closely specified was low*”⁵¹, although “*the finer specification of criteria, describing progressive levels of competency, has been shown to be capable of supporting reliable teachers’ assessment (TA) while allowing evidence to be used from the full range of classroom work*”⁵². She also cited research evidence that “*teachers’ judgements of the academic performance of young children are influenced by the teachers’ assessment of their behaviour; this adversely affects the assessment of boys compared with girls*”⁵³. There is also some evidence of bias in summative teacher assessment linked to student characteristics, including behaviour (for young children), gender and special educational needs; overall academic achievement and verbal ability may influence judgement when assessing specific skills. This evidence suggests that certain pupils could be disadvantaged by a system which relied more heavily on teacher assessment⁵⁴.

Teacher assessment in Wales provides some evidence of the impact of assessment practices focusing on the individual pupil, based on teacher judgement moderated across local clusters of schools.

Since 2005 there have been no statutory tests or tasks in Welsh schools at Key Stages 2 or 3. Instead, schools are required to make summative teacher assessment judgements for all eligible pupils at the end of Key Stage 2. There is also standardisation and moderation of examples of pupils’ work both within schools and by clusters of schools (including secondary schools). National exemplification and guidance on moderation has been made available since 2008.

Estyn (the Welsh inspectorate) recently published a survey⁵⁵ into the effectiveness of teacher assessment. They found that, “*although many clusters of schools have good systems to transfer assessment data as pupils move from the primary to the secondary school, many secondary departments do not use this information to plan timely intervention for pupils whose progress is too slow*”. In addition, “*Estyn’s evidence from school inspections across Wales consistently indicates that about a quarter of schools inspected each year have shortcomings in aspects of assessment... Estyn’s inspections of schools consistently indicates that assessment is one of the weakest areas of work in schools. There are also weaknesses in the use that secondary teachers make of assessment information from primary schools... Overall, primary and secondary school teachers are becoming more confident about their understanding of the characteristics of pupils’ work that demonstrate the NC level descriptions. Nevertheless, different perceptions as to what constitute appropriate standards in the core subjects at different levels in end-of-key stage 2 and key stage 3 assessments continue to exist*”.

⁵¹ Koretz, D., Stecher, B.M., Klein, S.P., McCaffrey, D., ‘The Vermont Portfolio Assessment Program: findings and implications’, in *Educational Measurement: Issues and Practice* 13, 5-16 (1994); Shapley, K.S. and Bush, M.J., ‘Developing a valid and reliable portfolio assessment in the primary grades: building on practical experience’, in *Applied Measurement in Education* 12, 11-32 (1999).

⁵² Rowe, K.J. and Hill, P.W., (1996) ‘Assessing, recording and reporting students’ educational progress: the case for ‘subject profiles’ in *Assessment in Education* 3, 309-352 (1996).

⁵³ Bennett, R.E., Gottesman, R.L., Rock, D.A., Cerullo, F., ‘Influence of behaviour perceptions and gender on teachers’ judgements of students’ academic skill’, in *Journal of Educational Psychology* 85, 347-356 (1993).

⁵⁴ Burgess, S. and Greaves, E., *Test Scores, Subjective Assessment and Stereotyping of Ethnic Minorities*, Centre for Market and Public Organisation, University of Bristol (2009).

⁵⁵ Estyn, *Evaluation of the arrangements to assure the consistency of teacher assessment in the core subjects at key stage 2 and key stage 3* (2010).

Estyn concludes that *“in KS2, confidence in the system of assessment, and in particular in the consistency of its application across different school cluster groups, is not as reliable as it should be because there is not, as yet, an effective process in place to moderate these judgements externally. In addition, as it currently operates, the process does not actually verify that the levels finally awarded to examples of pupils’ work at the end of key stage 2 are accurate across the full range of individual schools, clusters of schools and local authorities. This means that a level 2, or any given level, will not mean the same in different parts of the country”*. The absence of external moderation of Key Stage 1 teacher assessments in Wales *“make the reliability of value added and benchmark measures between the key stages uncertain”*.

Professor Daugherty (chair of the Daugherty Assessment Review Group which reported in 2004)⁵⁶ observed that the reforms introduced in Wales had smoothing of the transition between Years 5 and 7 (‘crossing the divide’) as their primary purpose. He made it clear that the current system of national assessment in Wales, focusing on the individual pupil and based on summative teacher judgements moderated across local clusters of schools, cannot and should not be expected to supply aggregate data that is robust enough for accountability purposes.

In a speech on 2nd February Leighton Andrews (Minister for Education in Wales) acknowledged systematic failings in Welsh education, and made a commitment to introduce *“a national system for the grading of schools to be operated by all local authorities... all schools will produce an annual profile containing performance information to a common format”* and to introduce *“a national reading test which will be consistent across Wales”*⁵⁷.

⁵⁶ Summarised in Daugherty, R., ‘National Curriculum Assessment in Wales: adaptations and divergence’, in *Educational Research* 51(2), 247-250 (2009).

⁵⁷ ‘Teaching makes a difference’, Reardon Smith Theatre, Cardiff, 2 February 2011.

Specific groups of pupils and specific types of school

Any future approach to testing, assessment and accountability needs to support every pupil. This represents a considerable challenge given the huge diversity of the primary sector in England, with over 17,000 schools.

Specific groups of pupils

We have received evidence about the impact of the system on particular groups of pupils and their specific needs and issues which need to be taken into account.

Feedback suggests that the progress and achievements of pupils with special educational needs is not appropriately recognised or celebrated by the current system.

Respondents, in particular the NUT and NAHT, have expressed their concern that existing progress measures often rely on traditional subject based learning. This often neglects the amount of additional teaching and learning of essential skills and behaviours necessary for pupils with SEN. Some respondents have suggested that different measures based on progress could be reported instead, as children with SEN are often less likely to make two levels of progress or achieve level 4 in core subjects⁵⁸. Proponents of new measures do accept that measuring the progress of children with SEN can be challenging, for example because pupils with life limiting, high-dependency conditions may have breaks in their education, and some pupils with SEN can exhibit challenging behaviour which can restrict their ability to access the curriculum. However, respondents have often noted that some pupils with SEN can, with appropriate support, make expected progress. A number of stakeholders expressed concern at the current inconsistency in processes for giving statements of special educational needs. They felt that different criteria created unfairness in the system, with some schools able to use SEN statements as an excuse for underperformance. There is a risk of perverse incentives in the current system, since accepting a pupil with significant special needs may not be in the interests of a school's performance results.

Pupils who moved school during Key Stages 1 or 2 tend on average to achieve less well in end of Key Stage 2 tests than children who had attended the same school throughout or who had completed a whole key stage in the same school. Research evidence shows that even a small amount of movement can have a significant effect on aggregate test results.

Some respondents have proposed that the system should take account of the transience of a school population, for example excluding the results of pupils joining during Year 6 from the accountability system, rather than assessing them as part of the school where they are on roll as at present. A number of stakeholders argue that schools should be held accountable for factors within their control, and that it is unfair to be measured for the achievement of a pupil who has spent a very limited time in the school.

Testing may have different impacts on low achieving and high achieving pupils. Some research shows that testing can potentially have a negative effect on lower achieving pupils. For instance evidence suggests it can be stressful and demotivating⁵⁹. Testing can reinforce the low

⁵⁸ Evidence from 2009/10 shows that at KS2 the gap based on attainment in English and maths is 53.8 percentage points, which has narrowed from 56.6 percentage points in 2006 (Dobson, J. & Henthorne, K., *Pupil Mobility in Schools: Final Report*, DfEE (2000)).

⁵⁹ Harlen, W., *A Systematic Review of Evidence of the Impact on Students, Teachers and the Curriculum of the Process of Using Assessment by Teachers for Summative Purposes*, EPPI-Centre, Institute of Education, (2004); *Children and their Primary Schools: pupils' voices*, Cambridge Primary Review Interim Report 5.3 (2007).

self-image of lower achieving pupils and lead to high levels of test anxiety. The extensive use of feedback, tracking, targets and extra coaching can reduce the self-esteem of lower attaining children⁶⁰.

Feedback suggests that pupils at the other end of the spectrum can also lose out as high attaining children who have already reached level 5 at the start of Year 6 can waste time on material they have already mastered. Some online consultation respondents have suggested that some of the highest attaining children make little progress in Year 6; as so much of the focus is on pupils at the level 3/level 4 boundary, higher and lower attaining children can receive less attention. A few respondents have argued that the fact that, in 2010, 51% of pupils achieved level 5 in reading and 30% in maths suggests that the tests do not discriminate sufficiently for more able pupils.

In addition, the Centre of Policy Studies argues that the current tests are biased towards children in middle class homes⁶¹. They suggest that the 'psychometric' model of testing unfairly penalises children of poorly-educated parents as it focuses almost exclusively on "*higher-order skills*" and consequently "*lower-order skills simply receive far too little attention in primary school*". They argue that English tests which focus on 'emotional literacy' and 'reading comprehension', rather than (where appropriate) "*the decoding skills and the basic conventions of written English do no favours for children of poorly-educated parents*".

There are significant differences in teacher assessments of pupils from different ethnic groups. Research shows that⁶² on average the achievement of Black Caribbean and Black African pupils is underestimated relative to white pupils, whereas the achievement of Indian, Chinese and mixed white-Asian pupils is overestimated. These findings suggest that replacing tests with teacher assessment might be detrimental to the recorded achievements of children from poor families, and for children from some ethnic minorities. Professor Harlen shows that teacher assessment can be biased by students' characteristics, and that a 'halo' effect can extend ability in some areas of the curriculum to others⁶³. Inaccurate judgements may increase attainment gaps between ethnic groups later in academic life: setting in secondary school may depend on earlier recorded attainment and pupils' motivation may be affected by poor achievement.

Some pupils need longer to complete tests and are therefore potentially disadvantaged by timed tests. At present pupils who meet certain eligibility criteria are allowed up to 25% additional time in National Curriculum Tests. Pupils using the modified large print or Braille papers are automatically entitled to 100% additional time. However, some respondents have expressed concerns that there are a significant number of pupils who would benefit from additional time on the tests.

⁶⁰ Reay, D. and Wiliam, D., "'I'll be a nothing": structure, agency and the construction of identity through assessment' in *British Educational Research Journal* 25(3), 343-354 (1999); Harle W. and Deaken-Crick, R., 'A systematic review of the impact of summative assessment and tests on students' motivation for learning', in *Research Evidence in Education Library* 1, EPPI-Centre, Institute of Education (2002).

⁶¹ Burkard, T. and Cleford, T., 'Replacing KS2 SATs – A submission to Lord Bew's review of Key Stage 2 testing, assessment and accountability', Centre for Policy Studies (2010).

⁶² Burgess, S., Wilson, D. and Worth, J., *A natural experiment in school accountability: the impact of school performance information on pupil progress and sorting*, Centre for Market and Public Organisation/University of Bristol (2010); Chevalier, A. and Gibbons S., 'Assessment and age 16+ education participation', *Research Papers in Education* 23(2), 113-123 (2008).

⁶³ Harlen, W., *A Systematic Review of Evidence of the Impact on Students, Teachers and the Curriculum of the Process of Using Assessment by Teachers for Summative Purposes*, EPPI-Centre, Institute of Education (2004).

22% of respondents to the online consultation wanted more time to be allowed in the reading test. There is clear concern that some pupils, especially those of lower ability, struggle to complete the test paper within the 45 minutes available. In 2004, QCA commissioned a study to investigate the impact of additional time on the Key Stage 2 tests in English reading and science. The results showed that the impact of additional time in science was negligible since most pupils finished within the appropriate time. However, for reading, the scores of both SEN and non-SEN pupils were improved significantly. No changes to timings were made as a result of the research, since the QCA curriculum team believed that sustained reading at speed was part of the construct to be assessed. Respondents have suggested this is a principle that could be revisited in the future.

Tests in a single week make pupils' absence an issue. Pupils are judged on their performance on the day of the test, and absence may mean that able pupils are recorded with no level.

Evidence shows it is consistently the youngest children in each year group who perform less well in school tests and the eldest who perform the best. There is some evidence that summer born children are disadvantaged. Recent DfE research⁶⁴ indicates that the gap between pupils born in September and August (i.e. with nearly a year's age difference) is slightly larger than the gender gap, while the FSM gap is twice as large and the SEN gap four times as large.

Age standardisation in national test outcomes and national assessments has been proposed by some respondents to prevent 'dysfunctional learning identities' from forming in the youngest children. This is widely used at Key Stage 1, though less common at Key Stage 2. Proponents of age standardisation argue it has the potential to unlock significant gains in the education system. However, some respondents have pointed out that there is a risk that age appropriate standards can become an excuse for low expectation and therefore underachievement. QCDA currently publishes tables which allow the marks on individual reading and maths papers to be converted into age-standardised scores.

Specific types of primary schools

As well as particular groups of pupils with specific needs and issues, respondents have pointed to particular types of school that face specific challenges.

It has been observed that small schools can be particularly penalised by the current system as their results are more likely to be volatile. Small schools have only occasionally been referenced in evidence to this review, but those respondents who have referred to them have expressed deep concern. NGA points out that changes in cohorts have a greater impact on small schools, and it is these schools which would therefore benefit most from a change to assessment based on rolling averages rather than single year groups. Currently, very small schools (with fewer than 10 pupils in the Year 6 cohort) are excluded from the Achievement and Attainment Tables.

Some head teachers of junior schools have expressed concerns that they are treated unfairly by the accountability system. They believe they are particularly disadvantaged by the current system because of concerns that infant schools inflate their end of Key Stage 1 teacher assessments, in spite of local authority moderation. This reduces junior schools' ability to demonstrate progress within the accountability system.

⁶⁴ *Month of Birth and Education*, DfE Research Report (2010).

Specific aspects of current statutory tests

There is widespread recognition that Key Stage 2 tests themselves are well developed.

Many respondents have expressed praise for the current tests and explained that their concerns are not about the tests themselves, which “*do what they set out to do*”, but about the wider system. Online consultation respondents often highlighted the high quality of the test materials. NFER for example says that “*the current tests could be considered the best tests of their type and are developed using the most thorough test development approaches available*”. NASUWT noted that the current tests are relatively broad, though they will inevitably struggle to cover the whole range of the curriculum. NAHT felt that tests are most effective where there are ‘right’ and ‘wrong’ answers and so maths and some elements of science lend themselves to written tests. The reading and maths tests have been mentioned infrequently by the respondents to the review, suggesting a relatively high level of satisfaction with them.

Despite the widespread praise for the current tests themselves **there are issues with specific aspects of the current Key Stage 2 tests.**

Feedback suggests that the maths test is generally perceived as the most effective in terms of the quality of questions in the written papers and mental arithmetic test, the coverage of the curriculum, and accuracy of the marking. The majority (62%) of respondents to the online consultation felt that the current maths Key Stage 2 tests were reasonably or very effective. ACME observed that the maths tests are “*reasonably valid and reliable as measures of children’s attainment in number, shape and space and data handling*”. However, ACME notes that the tests do not test the whole curriculum because of their short written nature (omitting the ‘use and application’ of maths), and pupils’ classroom experience is therefore often narrowed. The main criticism which several respondents have highlighted is the lack of opportunity for pupils to tackle “*meatier problems*” within the tests, and therefore to carry out mathematical enquiry or investigation.

The reading tests seem to be reasonably well supported by respondents, although concerns have still been raised. Some respondents have argued that the reading tests are too imprecise a measure of reading ability, allowing a child to attain level 4 but in fact have a reading age or a comprehension score below their chronological age. There is also concern that the test disadvantages children who are slower at processing written information, and therefore often do not finish the tests in the time allowed. 27% of online consultation respondents recommended including a wider choice of topics, as pupils’ interests could influence their comprehension.

Writing tests

The writing tests have generated by far the most concern. Feedback has consistently suggested that the current writing tests are particularly problematic. 43% of online consultation respondents felt the writing National Curriculum Tests are ‘inadequate’, and 33% felt they are ‘not very effective’. 47% of online consultation respondents want the writing test to be replaced with teacher assessment.

35% of online consultation respondents felt that the writing tests did not adequately reward creativity. They said that the writing tests were not particularly inspiring and did not reflect the current practice in classrooms where children were encouraged to take time with their writing and put effort into spelling, punctuation and handwriting. Respondents also noted that many children produced their best work as part of a structured lesson following an inspired discussion or school trip. Some respondents believed that going into a test without knowing what they had

to write about stopped many pupils from performing at their best. 32% of online consultation respondents suggested that there should be more choice of genre as children learned to write across a range of genres and had strengths in different areas. Many teachers feel that the choice of the writing prompt is crucial, since children's best writing reflects experiences they can relate to. However, Professor Wiliam cautioned that less able pupils tend not to choose genres which will allow them to show their full potential. Other respondents have suggested that when National Curriculum Tests allowed a choice of genre, some schools responded by preparing examples of work for pupils to memorise and repeat.

It has been suggested that writing tests could simply be assigned an overall level rather than a specific mark. This would avoid some of the contentiousness of a precise score, recognising that there is inevitably a degree of subjectivity in the marking process. Alternatively, writing could be tested with a number of separate short tests which would lend themselves to online testing (for example covering spelling, grammar, punctuation and handwriting), leaving speaking and listening and extended writing to summative teacher assessment.

Some respondents, including NAHT, argue that writing should not be subject to external marking. They believe that the content of writing tests can disadvantage pupils who have limited experience due to their background, which can lead to unfair questions for some pupils. NAHT proposes that attainment in writing should be teacher assessed and based on a portfolio of work, which would include work in specified, externally set genres and be produced under varying conditions, such as timed assessment.

Science Assessment

Science assessment has generated some debate during the Review. Various suggestions have been received about how it could be assessed in the future. The Wellcome Trust commissioned research in January 2009⁶⁵ which showed that all groups of children preferred tests (though not National Curriculum Tests) rather than summative teacher assessment to determine their progress in science. *“Children’s and parents’ ideas for improving KS2 science assessment suggested that assessment should be designed to motivate science learning by employing a variety of approaches, including non-SATs (sic.) tests, which allow for choice and have a strong emphasis on investigative work”*. The research recommended that *“science assessment at KS2 should be embedded in normal science class work and should include the use of end-of topic, as opposed to end-of-year, testing. It should cover a range of sources of evidence from practical, oral and written work, and should focus on the understanding of science as opposed to knowledge recall”*.

SCORE believes that the Key Stage 2 tests in science, before the sample was introduced, had a particularly detrimental impact on pupil's experience of the primary science curriculum. SCORE argues that the primary science curriculum should *“provide children with authentic experiences of the natural world and opportunities to satisfy their natural curiosity through hands-on enquiry”*, which is difficult to achieve through a 'high stakes' testing and accountability system.

34% of online consultation respondents believe that science needs to be more practical and that assessments should be based on practical science. 18% of online consultation respondents said that removing external tests and using summative teacher assessment was a good idea and saw it as the way forward. Some respondents noted that since National Curriculum Tests in

⁶⁵ Murphy, C., Kerr, K., Lundy, L., McEvoy, L., *Attitudes of Children and Parents to Key Stage 2 Science*, Wellcome Trust (2010).

science were abolished children still learned science but what they were taught was more relevant and more fun. However, respondents felt that since the tests ended, science teaching had lost some impetus and was generally seen to be of lower importance.

Validity and reliability of Key Stage 2 tests

Some respondents have expressed concern about the validity and reliability of the Key Stage 2 tests. All respondents accept that a degree of compromise between validity, reliability and manageability will always be necessary to ensure that the particular assessment best meets the stated purposes. However, some believe the right balance has not yet been struck.

Validity is concerned with *what* a test is assessing: the extent to which a test measures what it is intended to measure and whether the evidence and theory support the intended interpretation of the test outcomes. Reliability is concerned with *how well* a test measures whatever it is measuring: the degree of consistency and the precision of the measurement.

Some academic experts, for example Professor Harlen, argue that 'high stakes' testing means that validity is compromised by the need for high reliability in the results. This inevitably means that what is tested is limited to what can be easily marked, thereby excluding use and application of maths, speaking & listening and practical science⁶⁶.

Some teaching unions have also expressed concern about the validity of tests. The ATL for instance have said that tests do not and cannot cover the full range of what is taught across English and maths. However, for NASUWT and its members the current tests are relatively broad, although they will inevitably struggle to cover the whole range of the curriculum.

Others believe validity to be less of an issue. For example, in relation to the 2009 National Curriculum Tests Ofqual⁶⁷ reported that, although there were some issues with quality assurance and complexity of the mark schemes in English, overall the results were valid for their stated purpose and yielded consistent results across the country, irrespective of who marked the papers.

There has been some debate about the reliability of tests. 25% of online consultation respondents were concerned that the quality of marking was either poor or at best variable. It is generally accepted that tests will always include a degree of measurement error and it is the margin of error for both pupils and schools which needs to be considered. For instance, in evidence to the Select Committee, Black and colleagues suggested that inevitable concerns around reliability were not in themselves an argument against the use of formal tests, but rather a warning that they should be used with an understanding of their limitations⁶⁸.

However, there is mixed opinion from respondents though about the margin of error of the current tests. As an illustration of the range of opinion, Professor Wiliam suggested in 2000 that 32% of pupils could be given the wrong level⁶⁹. More recent work by Dr. Newton suggests a figure closer to 16%. He warns of the complexity, burden and cost risks of increasing test length

⁶⁶ Harlen, W., *The Quality of Learning: assessment alternatives for primary education*, Cambridge Primary Review Research Summary 3/4 (2007).

⁶⁷ *Monitoring report: 2009 Key Stage 2 tests*, Ofqual (2010).

⁶⁸ Black *et al* evidence to select committee 2008, quoted in Alexander, R., (ed) (*Children, their World, their Education: Final Report and Recommendations of the Cambridge Primary Review* (2007)).

⁶⁹ Wiliam, D., 'Reliability, validity, and all that jazz', *Education 3-13*, 29(3) 9-13 (2000).

to address this. Ofqual's recently-published report⁷⁰ indicates a classification accuracy of 85% for the 2010 English National Curriculum Tests, 87% for the science sample and 90% in maths.

The use of levels

The use of levels to measure pupil achievement has been questioned. The tests are designed to assess pupils working between levels 3 to 5 of the National Curriculum.

Many respondents do not believe that National Curriculum levels are an effective measure of attainment. Tim Oates, for example, believes that the current reporting is weak, and a lot of effort is made to assign pupils a level, but this is not broken down to explain its meaning to parents. ARK shares this view and argues that level 4 as a threshold measure is unhelpful as it can cut across a huge ability range. Dr. Newton believes that it is not necessary to award levels in order to make comparisons.

There is some concern about how test items are spread across levels 3, 4 and 5, which could potentially allow a pupil to achieve a level 5 based primarily on answering level 3 and level 4 questions. Many respondents suggest that this is why schools are so interested in getting test papers back and then poring over them to make sense of their results based on the specific questions asked. A number of assessment experts have suggested re-establishing 'criterion referencing', i.e. there would be a series of criteria and to achieve level 4 a pupil would need to achieve a sufficient number of them.

Potential approaches to assessment which would not require the use of levels have been put forward. Some countries simply measure pupils' scores based on national tests, while others use scores adjusted by pupils' ages. Norm-referencing provides an alternative, where pupils are ranked in order of ability within their school early on and again at a later stage; a pupil making average progress would be in the same relative position within their cohort. Using fine grades or raw scores instead would also reduce the apparent gaps between levels.

There is some concern that the current tests are only designed to measure pupils' attainment up to level 5. In 2010, 51% of pupils achieved level 5 in reading. This year QCDA will also offer level 6 tests to schools on an optional basis, based upon the Single Level Tests pilot. However, several respondents (including NFER) suggest that there are difficulties with this approach in that it may not be appropriate to award level 6 to primary school pupils who have not studied elements of the curriculum which are introduced in secondary schools. NFER therefore suggests that a 'high 5' could be considered as an alternative.

Timing of tests

The timing of tests has been debated. Opinion is mixed, particularly amongst online consultation respondents. 31% advocated shortened tests at more than one point. 27% suggested retaining tests in Year 6. 18% recommended tests at the start of secondary education. 30% selected 'other', and related comments suggest that many of these respondents do not consider there is a place for external tests at all.

Some advocates of moving tests to the start of Year 7, including Professor Brown, argue that it would ease pressure in Year 6 and change teachers' practice from drilling short-term learning into laying the foundations for success in October. This could encourage a move away from the

⁷⁰ Opposs, D. and Qingping, H., *The Reliability Programme: Final Report*, Ofqual (2011).

mindset of short-term learning and would make it harder for pupils to be awarded levels where they are not secure.

NFER and ARK are amongst those who feel that a move would be problematic in view of the known 'Year 7 dip' and the difficulties of holding primary schools to account for data from secondary schools. NUT did not think that moving tests to Year 7 would resolve any of the issues with the current system. NFER and Dr. Newton also argue that the current timetable is appropriate for a summative test and there is no compelling reason why this should be changed. ASCL felt it was inappropriate for secondary schools to be responsible for the administration of tests which would be used to hold primary schools accountable.

Some, including ARK, have suggested that tests should happen later in the summer term to maximise teaching time. However, when QCDA consulted in 2009 on moving the test dates to June in 2011, many schools and local authorities raised concerns, including the impact on educational trips and out of classroom learning opportunities, regular summer sports events and secondary transition arrangements.

Professors Baird and Elwood suggest that a national test taken towards the end of Year 5 or at the beginning of Year 6 could act as a measure of how well pupils are doing and what they need to do to improve.

Alternative approaches to statutory assessment

Some respondents have proposed greater use of technology in statutory assessment.

Many respondents have mentioned the potential of technology in statutory assessment. Lord Sutherland recommended that the delivery process for national curriculum tests should be modernised and improved, including the piloting of on-screen marking⁷¹. Pearson, the parent company of the current test operations contractor, Edexcel, have been clear that the increased use of technology would make the process more efficient and create richer data which would be more useful to schools.

However, some respondents have highlighted the practical challenges, including the timing and speed of introduction that need to be considered.

Some respondents have pointed to **the potential of computer adaptive testing**, and outlined the advantages for the English primary sector. The Centre for Policy Studies and Professor Tymms in particular have advocated computer adaptive testing, arguing that it can give a more accurate picture of a pupil's ability or academic achievement. This approach typically uses a multiple choice format, although it can be more complex. Pupils are presented with their own personalised test according to their level of performance. Questions are selected from a bank of questions using item response theory algorithms. Marking then happens instantaneously.

There are several arguments in favour of adaptive testing, which is already in use in primary schools in other countries, and many primary schools in Northern Ireland use adaptive tests. The proponents of adaptive testing point to significant efficiencies, arguing that a test need only be half as long as a National Curriculum Test in order to provide the same degree of precision. Adaptive testing can give both a score for summative assessment and more detailed information for formative assessment.

However, some respondents have pointed out limitations such as security issues, for example ensuring that a particular child sits the test. The availability of computers in some schools is also an issue.

There are examples of online testing in other countries from which we can learn. Tim Oates highlighted a test developed by Massachusetts Institute of Technology (MIT) in physics as a good example of the potential of online testing. Sir Michael Barber similarly highlighted experiments in computer-driven assessment in the US and Hong Kong. In New York the 'Wireless Generation' company provides online testing materials, which they mark and return to schools together with comparisons against schools with similar results.

There has been considerable debate about Single Level Tests

Single Level Tests (SLTs) were 60 minute tests, covering a single national curriculum level (levels 3–6) providing a link between formative teacher assessment and summative testing. They were designed to be age independent (with pupils entries from Year 3 to Year 6), taken at two points each year (June and December). Entry for a test was determined by the teacher judging the pupil to be ready to be entered for a particular level.

⁷¹ Sutherland, S., *The Sutherland Inquiry: An independent Inquiry into the delivery of National Curriculum tests in 2008 A report to Ofqual and the Secretary of State for Children, Schools and Families*, TSO, (2008).

Schools in 10 local authorities participated in a three-year pilot; in the third year only, the maths SLT was used in an accountability context (i.e. as a replacement for the maths National Curriculum Test). The pilot schools reported⁷² that SLTs contributed to a broader and more balanced curriculum across Key Stage 2, particularly in Year 6, and combined use of SLTs and Assessing Pupils' Progress (APP) materials impacted positively on teaching. Tracking of pupils' progress improved as a result, impacting positively on most Key Stage 2 pupils, particularly the more able. When used in an accountability context (in maths only), schools invested more time in preparing pupils for the maths SLT than the reading and writing SLTs. However, less Year 6 time was invested in revising for the maths SLT in June 2010 than for the maths National Curriculum Test in 2009. Head teacher representatives from the pilot commented positively on the use of SLTs in their schools. They felt that SLTs quality assured teacher assessment and increased teachers' confidence. Pupils felt less intimidated by taking a test aimed at their level rather than a test covering several levels, especially as they were aware that in most cases there would be a second opportunity to take the test.

Several other respondents have also expressed views on SLTs. Professor Brown accepted the principle of testing 'when ready', but felt SLTs should not be used in an accountability context due to the risk of pressure being placed on pupils throughout primary education rather than just in Year 6. QCDA felt the SLT approach offered a new way of looking at assessment, but conceded that providing only two test windows each year does not really equate to testing when ready. Dr. Newton cautioned that the concept of SLTs had proved overly complex, and questioned whether an effective model had been reached even after three years of piloting.

Developing a market-based approach to statutory testing and assessment has been suggested.

AQA has suggested that schools should be able to choose from a range of tests, encouraging testing organisations to compete on price and service, bringing new organisations into the market as a means of reducing risk and increasing innovation. It argues that this would substantially reduce the risk of test delivery failure. The involvement of a number of testing organisations could support improvement and innovation, such as introducing computer adaptive testing. As with general qualifications, testing organisations would be encouraged to produce parallel forms of a test (i.e. more than one organisation produces a Key Stage 2 writing test). All forms would need to demonstrate that they assessed the relevant parts of the National Curriculum and that standards were consistent across testing organisations.

ARK suggested that, instead of providing a single national test, schools could be required to administer and report results from at least one of a range of approved tests, producing standardised results. This would stimulate the market of test providers and encourage innovation. GL Assessment proposed distinguishing between the 'internal' tests and assessments which schools choose to purchase and the 'external' assessments which they are statutorily required to conduct. They felt it was appropriate to retain a single form of statutory external assessment and allow schools to determine what form of internal assessment was most appropriate.

⁷² *Evaluation of the Single Level Test (SLT) pilot final report*, PwC for DCSF (2010).

Transition to Secondary School

Feedback suggests that secondary schools make limited use of statutory test data to support transition.

41% of online consultation respondents said that one of the most important purposes of the tests should be to provide information about pupils' achievement at the time of transfer between schools. However, secondary school respondents have expressed concern that Key Stage 2 results are not always a suitable proxy for the attainment of pupils on entry to Year 7. Many secondary schools and head teachers have given evidence about the perceived disparities between the level gained at the end of Year 6 and the pupil's subsequent performance at the start of Year 7. Many secondary schools are concerned that over-preparation in Year 6 means that test results may not give an accurate reflection of a pupil's actual attainment.

Research evidence shows that secondary schools make widespread use of Cognitive Abilities Tests (which predict ability, regardless of previous attainment) and other internal assessments⁷³. The wide range of commercially available tests that are bought in by secondary schools shows the level of demand for additional information. This may suggest that they doubt the usefulness or accuracy of National Curriculum Test results, or that they see the value of triangulating with a different form of assessment.

More cross phase moderation has been proposed as a way of ensuring that more useful information about pupils is passed to secondary schools. Advocates of 'cluster' moderation suggest that secondary schools could make more active use of information and examples of pupils' work supplied by primary schools. Secondary schools could provide a 'moderation hub' for feeder primaries. This would enable Year 6 and Year 7 teachers to collaborate, at a local/regional level, to agree standards of attainment. Year 7 teachers would then be able to understand their new intake better and this would constitute a useful development exercise for all involved.

⁷³ Kirkup, C., Sizmur, J., Sturman, L. and Lewis, K., *Schools' use of data in teaching and learning* NFER (2005).

Annex A – Stakeholders who have submitted evidence to the Review

Online Consultation:

The online call for evidence launched on 25 November 2010 and closed on 17 February 2011, attracting **3,940** responses. Of these, **2,386 (61%)** were primary head teachers; **900 (23%)** were primary school teachers; **156 (4%)** were parents/carers; **124 (3%)** were governors; and **59 (1%)** were pupils. The remaining **8%** of respondents comprised: education professionals; local authorities; secondary school teachers and head teachers; teaching unions/professional associations; and others. We have completed an initial analysis of all the online consultation responses and have summarised the main findings in this progress report. We will continue to study the online consultation responses and will publish a full analysis in the final report in June.

Organisations and stakeholders who have given oral evidence to the panel:

Association for Achievement and Improvement through Assessment (AAIA)
 Advisory Council for Maths Education (ACME)
 Professor Robin Alexander
 ARK Schools
 Association of School and College Leaders (ASCL)
 Association of Teachers and Lecturers (ATL)
 Professor Jo-Anne Baird
 Professor Sir Michael Barber
 Julian Barrell
 Professor Margaret Brown
 Cambridge Primary Review
 Professor Richard Daugherty
 Professor Janette Elwood
 GL Assessment
 Professor Wynne Harlen
 Bill Holledge
 Dr. Tina Isaacs
 Warwick Mansell
 Dr Christine Merrell
 National Governors' Association (NGA)
 National Association of Head Teachers (NAHT)
 National Association of Schoolmasters/ Union of Women Teachers (NASUWT)
 National Foundation for Education Research (NFER)
 National Union of Teachers (NUT)
 Dr. Paul Newton
 Tim Oates
 Ofsted
 Pearson UK
 Qualifications and Curriculum Development Agency (QCDA)
 Steiner Waldorf Fellowship
 Lord Sutherland
 Professor Peter Tymms
 Professor Dylan Wiliam
 Selection of junior, primary and secondary school head teachers

A number of written submissions have also been received by the Review.