

Report of the
ASSESSMENT REVIEW GROUP

Redressing the balance

January 2017



FOREWORD

David Ellison - Chair of the Assessment Review Group

Assessment is at the heart of high quality teaching and learning. It is the means by which teachers evaluate progress and diagnose pupil needs. It provides the opportunity for students to recognise their progress and helps parents engage in their children's educational journey. To put it simply, assessment helps teachers to teach and pupils to learn.

Yet somewhere, somehow, this core purpose has become distorted. Accountability arrangements are based on the results of statutory assessments, making the assessments incredibly high stakes and leading to negative consequences. The quality and accuracy of the assessments declines in proportion to the stakes attached to them; and the narrower the measure, the more dramatic the effect. The fear of 'intervention' - and the devastating consequences of it on schools and teachers - have stretched the system and those working within it to breaking point. We know schools need to be accountable, but they should not operate in fear and uncertainty. The balance has been lost within current arrangements. It is time to redress that balance.

In the spring of last year a consensus emerged within the profession, and beyond, that a fundamental review of statutory assessment was needed. The Assessment Review Group was established in May 2016 to consider the current system and to try to identify an alternative, better vision for the future. This report is the culmination of the discussions that took place within the group over a series of meetings.

The report sets out a series of principles that should underpin any future assessment system. Whilst we have set out a broad vision for what an alternative system could look like, inevitably there remain some complex questions that need further discussion and investigation within and beyond the profession. Our hope is that this report will provide a useful starting point. We certainly do not see it as the end of the debate and we would welcome further discussion and expert opinion in these areas.

It is clear that any attempt to design a national assessment system is likely to prove controversial. Throughout the review process value judgements have come into play. Even within the group itself, there was not always unanimous agreement about each aspect of an alternative system. There is no 'perfect' solution and trade-offs will inevitably be required. We would not want to repeat the mistakes of the past by rushing this process - it is important that we get the new system right even if that means taking a little longer to plan it properly.



Whatever the future holds, the group was clear that benefits for pupils will only be fully realised if we rigorously defend the core purpose of assessment in supporting learning. By being true to this purpose, we stand the best chance of being able to unleash the potential of all pupils within our classrooms.

David Ellison
Chair of the Assessment Review Group

CONTENTS

Foreword	2
Contents	3
Context	4
The Assessment Review Group	5
Six Guiding Principles of Assessment	6
Future Model of assessment	11
Issues for consideration:	14
Start of school assessment	
National assessment banks and sampling	
Alternative models for writing assessment	
Single and cross key stage schools	
Afterword by NAHT deputy general secretary, Nick Brook	23
Membership of the Assessment Review Group (and thanks to contributors)	25
Bibliography	26

CONTEXT

The National Curriculum and associated assessment regime were introduced in England through the Education Reform Act of 1988 (NAHT, 2014). In 1992 the first secondary school performance tables were published by the DfE, followed by the first primary school performance tables in 1996. The problems faced by schools through the use of assessment data for accountability purposes are long in the making.

Early in 2011 the Secretary of State announced a review of the National Curriculum by an expert panel. The panel reported in December 2011 and concluded that attainment targets and level descriptors should not be retained in the revised National Curriculum. In June 2013, the Secretary of State announced that levels had become too abstract, detracted from real feedback to pupils and parents, and that schools have found difficulty in applying them consistently. New tests would be introduced in the summer of 2016.

By spring 2016 it was clear that government-led reform of assessment lacked a clear vision for a stable, proportionate and coherent approach to acknowledging children's achievements and measuring school performance. Its implementation fell short too. Guidance was delayed and obscure; the Key Stage 1 spelling test was published online by mistake; test papers contained errors; the new expected standards were inaccessible to many children and inconsistent Local Authority moderation of writing teacher assessment made comparison of performance between areas meaningless. It had become clear that the system was not working for schools or government.

In October 2016, following prolonged pressure from NAHT, the DfE – under the leadership of a new Secretary of State, Justine Greening – took action to tackle the short and medium term challenges faced by schools and address the criticism levelled at proposed future reforms. The government announced, amongst a range of measures, that they would not introduce a proposed Year 7 resit of SATs, a significant shift in approach which was welcomed by the group. It was also agreed that interventions would not be made on the basis of 2016 data alone. The government committed not to introduce new assessments prior to the 2018/19 academic year, and promised a consultation in the New Year on the long term shape of assessment, to be informed by the independent Assessment Review Group.

In this and other statements the new Secretary of State has signalled a shift from her predecessors and has demonstrated a welcome openness to working with the sector to address system failures. In producing this report we have sought to learn the lessons from past mistakes, not dwell on them, and offer what we hope are constructive proposals to shape deliberations about the future of the primary assessment system.

THE ASSESSMENT REVIEW GROUP

The independent Assessment Review Group was established by NAHT in May 2016, to determine the assessment arrangements that should be in place from the start of school through to the beginning of Key Stage 3, and to promote the development of a system that works for pupils, parents, teachers and school leaders.

The group was chaired by David Ellison, a deputy head teacher and comprised of experienced practitioners, leading academics, and experts on assessment. The group met on eight occasions through the summer and autumn of 2016 and considered evidence, key themes and issues in relation to the current assessment system and procedures, accountability requirements, progress measures, and statutory testing regime.

Expert witnesses were invited to attend specific sessions to provide additional insight and challenge to the emerging proposals. Membership of the group and a list of contributors can be found at the end of the report.

The members of the Assessment Review Group have had many passionate debates across a wide range of issues. The range of views and ideas presented were debated, and potential proposals challenged and dissected. Inevitably, the group did not reach a consensus on every element so this report is a reflection of the majority views of the Assessment Review Group, rather than of every individual member or the organisations they may represent. It should not therefore be assumed that any individual or organisation necessarily endorses the entire contents of this report.

SIX GUIDING PRINCIPLES OF ASSESSMENT

1 Assessment is at the core of good teaching and learning

Ongoing, high quality assessment is at the heart of effective teaching and learning. Assessment enables teachers and pupils to develop a strong understanding of strengths, areas for development and to plan for the next steps in learning. Effective assessment allows teachers to have a positive impact on the learning and progress of all pupils. Information from assessments can be utilised by school leaders to guide school improvement. Such assessment takes a variety of forms, from observation and discussion through to low-stakes in-class tests. There is a wealth of research and writing on this topic, and we are better equipped than ever to continue to develop effective and robust assessment practices in our schools.

High quality assessment is also critical in supporting good transition, especially as pupils move between key stages or schools. It allows teachers to ensure that the learning journey is not interrupted and the learning that has already taken place is built upon. The information that is shared needs to be more than just an overarching label or a number; more detailed and nuanced information about each pupil's learning profile is essential.

We should continue to focus on improving the effectiveness of day to day assessment across all schools. The better our teachers get at assessment, the more effective they are likely to be in the classroom. A strong focus on assessment from initial teacher training through to ongoing CPD for established teachers and school leaders must be a national priority.

2 Statutory assessment should be separated from ongoing assessment that happens in the classroom

Pressure arising from the use of statutory assessment results for high-stakes accountability systems, can lead to a number of negative consequences, such as teachers 'teaching to the test' and neglecting those curriculum areas or topics that are not likely to be formally assessed (Yeh, 2005). Teaching can disproportionately emphasise rote learning of factual knowledge rather than the acquisition of creative skills, problem-solving skills or general reasoning. Ofsted's 2015/16 Annual Report recognised the issue of the curriculum being narrowed, and raised concerns at the small amount of time being spent teaching science and foreign languages in primary schools. In another study carried out by inspectors, the impact of limited time in the curriculum, a lack of separate lessons, and limited opportunity to develop learning at greater depth were all identified as issues for the study of design and technology in primary schools (Ofsted, 2016).

Lord Bew's Independent Review of Key Stage 2 testing, assessment and accountability found "strong evidence... that external school-level accountability is important in driving up standards and pupils' attainment and progress" (Bew, 2011). However, researchers (Volante, 2004), (Stiggins, 1999) have argued that the use of data from national assessments can have a negative impact on pupil performance, particularly when data from high stakes assessment is used to hold schools to account (Bethell, Kellaghan and Ross, 2011). To help reduce this adverse impact it is important to strike an appropriate balance between the use of statutory assessments and the data which these produce, and the ongoing assessment which happens every day in every classroom in the country.

It is all too easy for statutory assessment to become entangled with in-school assessment - particularly when schools are driven to predict and provide data on future performance in statutory assessments. Under these conditions, in-school assessments inevitably take on the form of statutory assessments, in order to produce compatible data, however inappropriate this form may be to support teaching and learning. We should shift away from predictions of future performance and focus more on capturing accurate pictures of current performance of pupils against expected standards for their age. This has a major impact on what data should and shouldn't be asked for.

The core focus of assessment should be on supporting learning, not simply tracking progress. To help maximise the progress children make, we should expect all schools to have highly effective and robust assessment processes in place. These are entirely separate from statutory assessments but should give a clear sense of how children are progressing, and how they can be supported to progress further. Such information should allow teachers and school leaders to identify which children need additional support or challenge and in which specific areas.

It is a reasonable expectation that schools should be able to explain to external agencies, such as Ofsted, how they use assessment to support children's learning, but in line with current expectations there should be no one preferred way of doing this. Ofsted should continue to evaluate what the school uses already and not expect assessment information to be presented in any particular style or format or with any given frequency. To facilitate this, Ofsted, and other external agencies, must have a strong understanding of effective assessment.

3 Data from statutory assessment will never tell you the whole story of school effectiveness

Performance data rarely provides answers as to why something has happened. However, the imbalance in current accountability arrangements is such that, too often, weaker performance of a cohort leads to a presumption that this must be the result of failure on the part of the school. Raw data from statutory assessments should not be used to draw simplistic conclusions about a school's performance or lead to heavy-handed intervention. This misuse is at the heart of many of today's problems with assessment. Results from such assessments are a useful indicator of a need for further investigation and may reflect other in-school factors which are proven to influence pupil performance.

In recent years too much significance has been attached to numerical pupil outcomes from statutory tests alone in judging school performance. Data generated by such tests simply represents how a relatively small group of pupils performed in a set of narrow tests, focussed on a small segment of the curriculum, at a given moment in time. Taken in isolation, test results are not a good proxy for judging school effectiveness; there are better ways.

One of the most stable findings in educational research is the impact of students' background on achievement, especially their parents' level of education and earnings (Clarke, 2007). Home environment is also an important factor in stimulating the development of early literacy and numeracy skills (Bonci, 2008). Consistent with this research, PIRLS and TIMSS data on the home background of students shows a strong positive association between primary student achievement and home educational resources, parents' emphasis on early literacy and numeracy activities, and children's literacy and numeracy skills when entering school (Martin and Mullis, 2013).

Research therefore supports the fact that judgement of a school's success or failure on the basis of statutory tests is unjust and unreliable. No intervention should be triggered on the basis of test data alone. Rather, the results from statutory assessments should trigger further discussion leading to a qualitative expert judgement. We should also remember that superficially good test results can be achieved at a high price in terms of curriculum breadth, extra-curricular activity, pupil welfare and school sustainability - none of which are evident in the raw data. Over-reliance on data is simply naive and in some instances dangerous.

Finally, many schools have small cohorts of pupils who vary significantly in their composition from year to year. This can produce fluctuations in data with no relation to the school's underlying effectiveness. The statistical significance of this data in very small schools is highly dubious.

4 The statutory assessment system should be accessible to pupils of all abilities and recognise their progress

A basic expectation of any assessment system is that it should recognise the progress made by all children. The current interim framework and assessment materials fail to do this. Simplistic, overarching labels such as 'working below the expected standard' mean that the progress of too many children is ignored and too many children are effectively labelled as failing and the cumbersome bureaucratic language does not conceal this perception from pupils or their families. This is not only unhelpful to the school but it also sends entirely the wrong message to our pupils, potentially having an impact on their future motivation.

In the final report of the Independent Review of Key Stage 2 testing, assessment and accountability in June 2011, it was acknowledged that the reading test should be accessible to all pupils. The report recommended that the development of the new reading tests should take into account the balance of text and reading time, establish a clear order of difficulty for both texts and questions, and ensure that the texts themselves are accessible to all pupils. Feedback on this year's reading test highlights that these recommended design features to ensure accessibility were severely lacking and it is hoped that significant improvements will be seen this year.

Strictly timed test conditions get in the way of some pupils properly demonstrating what they can do and the progress they have made. Although it is recognised that fluency is a skill in reading that forms part of the statutory assessment, this should not be given so much significance as to disadvantage many pupils. Evidence shows that, in reading, scores of both SEND and non-SEND pupils are improved with additional time. In 2004, QCA commissioned a study to investigate the impact of additional time on the Key Stage 2 tests in English reading and science. In reading, the scores of both SEN and non-SEN pupils were improved significantly. We believe that the focus of tests should be on capturing what children can do, not on mastering test technique at such a young age.

5 Progress should be valued over attainment in statutory assessment

Whilst any form of data from statutory assessment alone should not be used to judge school effectiveness, if such data is to be used as part of the wider picture when holding individual schools to account then the fairest way to do this is by focusing on the progress pupils make. Attainment is important and all teachers want as many children as possible to reach the highest standards. However, when it comes to holding schools to account, it would be grossly unfair to base comparisons on attainment when children's starting points can be so different.

However, we need to be careful not to assume that a certain starting point is an indicator of future performance. Not only can this act as a ceiling on the child's progress, the individual circumstances of a child may have a negative effect on the progress that they make.

The group discussed at length the possible use of a Contextual Value Added measure (CVA). The group was clear that when measuring progress, those holding schools to account should take full account of the context that the school operates in (including factors such as mobility, levels of SEND, levels of deprivation, etc.) As mentioned in principle 2, data should only ever be part of the picture and should act as a starting point for further discussion. After much consideration, the group decided that a CVA measure would not be in line with this principle and would give undue credibility to a single measure. Context really does matter, but it should not itself be reduced to a single data point.

Even though progress measures are preferable to attainment measures to inform a judgement about school effectiveness, they are not the complete answer when making such a judgement. Any statistical comparisons both within and between schools using progress data produced using the results of statutory tests need to be treated with a great deal of caution. If progress measures are to be used, it is vital that those drawing inferences from such evidence understand the issues and take steps to avoid erroneous inferences. Thus far there has been little to no attempt to communicate any of the issues identified alongside published performance tables or in guidance documents issued to school leaders (Perry, 2016). If progress measures are to continue then the limitations must be carefully communicated to all relevant stakeholders, including Ofsted, RSCs and parents.

6 The number of statutory assessments in the primary phase should be minimised

The disproportionate emphasis on statutory testing for the assessment of primary aged pupils must be ended. One way of redressing the balance is to reduce the number of statutory assessments primary age children are expected to participate in.

Research evidence suggests that there is no obvious correlation between testing frequency and pupil outcomes. As each participating country takes part in the same assessment, the results of TIMSS and PIRLS 2011 are a useful resource for comparing pupil performance and frequency of testing from country to country. In Singapore, pupils' average scaled score was significantly higher than the average scaled score achieved by pupils in England in all three subjects (Martin and Mullis, 2013), and there is only one statutory national test at primary level in Singapore (Arora et al, 2012a). In Hong Kong, pupils experience very similar levels of testing as pupils in England, but perform significantly better in the international assessments. Denmark's results in maths, reading and science were not statistically significantly different to England's, but Danish primary school pupils experience a higher level of testing than English children. Pupils take tests in Danish (with an emphasis on reading) in years 2, 4 and 6 of primary school. In addition, a maths test is taken in year 3 and in the final year of primary school (Danish Ministry of Education, 2016).

Most studies conducted on factors which influence performance do not even mention the amount of testing in their analysis, as societal, economic and cultural factors have a far more significant impact on outcomes.

A FUTURE MODEL

The aim of this section is to present a simple overview of an alternative model for statutory assessment, which embodies the guiding principles identified earlier in this report.

Have two statutory assessment points for primary pupils

Statutory assessment in primary school should be restricted to two points, Reception and Year 6, in order to create the space in between for schools to focus on delivering a broad and balanced curriculum, appropriate to the needs of all children. Throughout the primary phase, schools should be free to determine their own processes and procedures for pupil assessment, informed by widely available evidence of best practice, that allows teachers to maximise pupil learning and progress.

High stakes testing narrows the focus of the curriculum to that which is tested. The group do not believe statutory testing should be used by the government to influence teaching, learning and pedagogy. The various screening checks deployed by the government, including phonics and the proposed multiplication tests, should instead become part of the national sampling framework.

Introduce a start of primary school statutory assessment

In order to establish a baseline from which to measure progress, teachers would carry out an observation-based assessment during a child's first year in primary school. This should take the form of a single, nationally agreed assessment to avoid a repetition of the problems experienced in 2015/16. We anticipate that a moderation process would be necessary to support this. Great care would need to be taken when designing such an assessment, with significant input from Early Years experts. It is important that the results of this assessment should not be used to set targets for individual pupils or as a predictor of their future progress. Instead, the data from this baseline should be used solely as part of a cohort level measure of progress at school, local and national level.

Whilst it was relatively clear that the end point would be the summer term of year six, agreeing on the best 'start point' or baseline proved one of the most challenging issues the group faced. There was general agreement that the initial assessment or 'start point' should be as early as possible in a child's time in school, in order to take full account of the progress they make throughout their primary schooling. There is much to consider regarding any baseline assessment and these issues are outlined later in this report.

Remove end of Key Stage One statutory assessments

In the proposed model there would be two statutory assessment points. One at the start of a child's time in primary school and one at the end. The key measure arising from statutory assessment should be the progress children make between these two points therefore end of Key Stage 1 assessments should be removed as a statutory requirement.

Streamline and improve Key Stage Two statutory assessments

At the end of year six, in the medium term, we envisage statutory assessments in reading, maths and writing would continue in some form. Reading and maths would continue to be assessed through a national test, externally set and marked. Writing would remain teacher assessed through an improved system that focuses on the overall quality of a child's writing rather than the component parts. Early evidence suggests that comparative judgement may provide a workable and valid alternative to current arrangements for teacher assessment of writing.

Make statutory tests accessible and enable pupils to show progress

Statutory assessments and tests must be designed in such a way that the majority of children are able to access them. At the very least, tests should be structured so that the questions, and where appropriate any texts, appear in order of difficulty. Serious consideration should be given to removing the hard time limits for statutory assessments, particularly in reading, and replacing these with a minimum and maximum time limit so that children can focus on demonstrating what they can do rather than test technique. Inevitably there is likely to be a very small proportion of children with more significant special educational needs who are not able to access the tests. The Rochford Review has offered some interesting and potentially useful recommendations in this specific area which should be considered fully.

Introduce national sampling and assessment banks

Within this model, the government would have the option of carrying out national sampling if there were a need to monitor standards in particular subjects or aspects of the curriculum. The data produced through sampling should be used to gain an understanding of national standards. It should not be used to hold individual schools to account but could provide national data against which schools can evaluate themselves. In the long term, there is potential for national sampling to replace the current model where every pupil takes every test at the end of Key Stage 2.

All schools would be expected to have robust assessment processes in place and to be able to explain how they use these to support pupils' learning, to identify and intervene where pupils are falling behind, and to report to parents. Schools should be mindful of the recommendations made in the Commission on Assessment Without Levels Final Report when designing such processes (DfE, 2015). To support teachers and schools, a national bank of assessment materials should be made available. Such resources would also help teachers in assessing the progress children are making against national expectations.

Report pupil performance as a score on the national scale

The terminology used to describe pupils' attainment in 2016 (working towards the expected standard, working at the expected standard or working at greater depth within the expected standard) was unhelpful, arbitrary and demotivating. Such an approach also fails to recognise and celebrate the progress that a significant group of pupils have made. The group were particularly concerned about the effect on those pupils who, despite making significant progress, could only be judged to be working below expectations at both the end of Key Stage 1 and the end of Key Stage 2. Stopping the use of such terms and simply reporting a child's scaled score would be a positive step forwards.

Accept data is only one part of the picture of school effectiveness

It is important to reiterate that this model should be viewed in light of the overarching recommendation that any data produced from such statutory assessments should be seen as only one element when judging school effectiveness. Schools should not be held to account on the basis of this data alone. It is also important to recognise that such statutory assessments will never be able to capture all aspects of a child's progress or all the different ways in which a school contributes to the progress a child makes.

No one single set of results should lead to negative consequences for the school. All data should be considered over a rolling three year period. There needs to be a recognition that cohorts of pupils vary; a dip in results in one year does not necessarily equate to a decline in school effectiveness. Basing interventions on such a short-term approach is unlikely to be helpful or indeed valid.

End floor and coasting standards as determinants of intervention

The use of floor standards and coasting standards to determine intervention in individual schools should be stopped. Instead there should be a greater level of dialogue between schools and those that seek to hold them to account, including RSCs. The starting point should be a discussion around the data to understand the context and story behind it. Any intervention at this point should be supportive, recognise the knowledge and understanding of the professionals working within the school and be based on working with the existing leadership team in the school.

In an ideal world, data from assessments should be used as part of the inspection process. The results of the inspection may, if appropriate, trigger supportive intervention, and the RSCs should base their work on the inspection results rather than independent evaluations. This streamlines the accountability system without reducing rigour, inserts the necessary expert judgement into the process, reduces conflict and duplication, and minimises the level of fear and uncertainty.

ISSUES FOR FURTHER CONSIDERATION

START OF SCHOOL ASSESSMENT

There was considerable support within the group for the view that progress is the fairest way to measure school performance. The logic of taking a baseline measure as early as is practicable, so that the impact of the reception year is properly recognised, is inescapable. However, what a baseline measure should include; when it should be conducted; and how an accurate assessment can best be made proved to be highly contentious issues in the group.

The negative experience of many schools of the recent reception baseline trial, still fresh in the mind, was recognised as having muddied the water in terms of taking an objective view of the pros and cons of such an approach. The group considered the evaluation of the pilot and wider research in order to identify the key evidence which the government would need to take account of in the design and implementation of an alternative model.

Baseline assessment trial

In March 2014, the Department for Education announced plans to introduce a baseline assessment for all children at the start of reception in September 2016, with schools able to opt in early to a pilot phase from September 2015. The assessment was conducted during the first few weeks of the autumn term of the Reception year and was designed to produce a 'baseline' figure on the basis of which progress during the primary years could be measured.

As part of the pilot, schools were able to select one of three baseline assessment schemes approved by the Department for Education; Early Excellence Baseline Assessment (EExBA), Centre for Evaluation and Monitoring baseline (CEM) and National Foundation for Educational Research (NFER) Reception baseline (Standards and Testing Agency and Department for Education, 2015).

On 7 April 2016, the government announced the results of a study on the three baseline assessments. This concluded that the assessments were "not sufficiently comparable (with each other) to create a fair starting point from which to measure pupils' progress". As a result, these baseline assessment outcomes would not be used for school accountability purposes.

What should a baseline measure include?

There are many different views, and much research published, on what a baseline measure should include, in order to provide the most accurate assessment of starting point or indeed predictor of future potential.

In its requirements for an appropriate baseline assessment, the Department for Education stated that the 'clear majority of the content domain must be clearly linked to the learning

ASSESSMENT REVIEW GROUP

and development requirements of the communication and language, literacy and mathematics areas of learning from the EYFS' (Brogaard Clausen et al., 2015; Davidson et al., 2015), although additional areas of learning may be included at the discretion of the individual scheme providers (Kirkup, 2015). This approach was widely criticised, with some commentators suggesting that the baseline assessment focused too heavily on children's early literacy and numeracy skills; therefore providing information on a narrow range of knowledge and skills which were not predictive of children's later progress (e.g. Brogaard Clausen et al., 2015, Whitebread, 2016). Members of the group shared this concern and were minded to support the development of an assessment that goes beyond a narrow range of knowledge and skills.

A significant body of research has systematically highlighted the importance of children's early social-emotional skills in their later development and academic outcomes. Children with better emotional wellbeing make more progress in primary school, as social-emotional skills are crucial to their academic achievement and development (Brogaard Clausen et al., 2015). Child behavioural skills account for a substantial portion of children's early academic achievement (Montroy et al., 2015). In particular, early self-regulation has been identified as a key predictor of both current and later academic achievement and has been linked to better academic achievement for children in primary school, even for those at-risk for underachievement (Montroy et al., 2015). However, none of these things are easy or straightforward to assess in a formal assessment.

When should a baseline be conducted?

Many argue that a standardised assessment in the first six weeks of school is developmentally inappropriate. Research suggests that the new and unfamiliar school context and routine are not conducive to children showing their 'true potential', and that young children in their first six weeks of school lack the confidence to demonstrate what they are capable of doing (see Bradbury & Roberts-Holme, 2016; Brogaard Clausen et al., 2015; Dubiel, 2016).

The majority of the group were supportive of a future baseline assessment taking place in the second half of the Autumn term, in order to give children time to settle in to their new environment whilst still being early enough in the school year to reflect a near-entry level. A two-week submission window is suggested in which judgements should be finalised and submitted, to ensure consistency across schools.

How best can an accurate assessment be made?

There was no desire within the Assessment Review Group for children to take formal 'tests' at such a young age. Overwhelmingly the group believed that observational teacher assessment offered the most appropriate way forward for assessing pupil starting points. Evidence suggests that observational teacher assessment can be robust and consistent when supported by high quality training, exemplification and a framework for moderation. It was noted within the group that the considerable money saved by removing statutory testing at Key Stage 1 could well contribute towards CPD for teachers and higher quality moderation at baseline.

ASSESSMENT REVIEW GROUP

As an alternative, it was suggested to the group that aspects of the Early Years Foundation Stage Profile could possibly provide the information required for a baseline assessment. The group considered a range of views for and against (summarised in box, below). Members of the group remained concerned that there was considerable risk in attempting to adapt the EYFSP for statutory baseline purposes and that any developments in this area should not risk interfering with effective Early Years practice.

Early Years Foundation Stage Profile - A valid alternative?

The EYFS approach is proposed to have several advantages over a baseline test, in that the assessment (Bradbury & Roberts-Holme, 2016):

- is developed from on-going observational assessment across contexts and over time;
- is contributed to by parents and others who know the child well;
- reflects a child's responses to challenges and embedded skills and knowledge with the child uses independently in a range of situations;
- is holistic; focusing on prime areas (personal, social and emotional development; communication and language; physical development) as well as the specific areas (literacy; mathematics; knowledge of the world; expressive arts and design) and the characteristics of effective learning (playing and exploring, active learning, creating and thinking critically);
- the EYFSP does not have to be conducted in English, so can still be used for those children who do not speak English confidently.

Research by Bradbury & Roberts-Holme (2016) found strong teacher support for the EYFS profile; 82 per cent of respondents agreed that 'The EYFS Profile helps me to monitor the development of the reception class and plan for learning.' This is also reflected in the choice of baseline that many schools implemented last year; over 70 per cent of primary schools selected Early Excellence baseline (known as EExBa) (TES, 2015a), which is based on observations, like the existing EYFSP (Bradbury & Roberts-Holme, 2016). This is further supported by the fact that the 2015 NAHT reception baseline survey found that the overwhelming majority of respondents who signed up to Early Excellence felt that a key benefit was how well it aligned with the EYFS (92 per cent).

However, respondents to the survey by Bradbury & Roberts-Holmes (2016) also provided some negative comments, largely related to the scale and content (particularly for maths) of the EYFS Profile (Bradbury & Roberts-Holmes, 2016), suggesting that teachers do not find it without flaws. Other limitations of using the EYFSP, as an alternative to a separate baseline measure are:

- The EYFSP assessment takes place at the end of the first year; a progress measure constructed using this information would fail to capture the impact of the school in that first year (Davidson et al., 2015).
- The problem of low scoring within a high stakes accountability culture is not unique to baseline assessment; similar tactical responses have been found in relation to the EYFS Profile (Bradbury, 2013, in Bradbury & Roberts-Holme, 2016).
- The EYFSP was not intended to be used as a high-stakes accountability measure and as such, cannot be guaranteed to be fit for this purpose (Davidson et al., 2015).

The use of assessment data

The purpose of a baseline assessment at the start of a child's time in school should only be to provide a starting point for the progress measure at the end of Key Stage 2. Concerns were expressed that such a baseline assessment should not be used as a predictive measure and certainly not to set targets for an individual child to achieve at the end of Key Stage 2. The results should provide a cohort measure at school, local and national level, with only national and local data being published, and be used as an indicator of what might be expected from that cohort of children at the end of primary school. To fulfil this purpose, it would be necessary to implement one consistent national baseline assessment.

The group was largely supportive of the concept that a start of school baseline would be collected and 'black-boxed' until end of primary school assessments have taken place (seven years later) in order to calculate a cohort progress measure.

NATIONAL ASSESSMENT BANKS AND SAMPLING

The burden of high stakes standardised national testing, where the results are used as narrow measures of accountability, are well known. The Organisation for Economic Cooperation and Development (OECD) conducted international research on evaluation and assessment, reported in *Reviews of Evaluation and Assessment in Education: Synergies for Better Learning* (2013) and notes that standardised assessments are essentially narrow measurements that are limited to measuring a selection of curriculum goals. More complex competencies are generally measured through some form of classroom assessment where it is possible to achieve richer and deeper assessment, with a greater opportunity for more diverse and innovative approaches. Overall closed-task 'paper and pencil' assessments may be chosen for the purposes of objectivity, fairness and impartiality, whereas performance-based assessments better capture higher level skills such as problem-solving, creativity and collaboration, but are more costly, time-consuming and harder to judge on a larger scale (OECD, 2013b).

The report noted the risk that evaluation and assessment systems can distort how and what students are taught: if teachers are judged largely on results from standardised student tests, they may 'teach to the test', focusing solely on skills that are tested and giving less attention to students' wider developmental and educational needs. To mitigate this risk, OECD argued that pupils' needs must be placed at the centre of evaluation and assessment, in order that pupils are engaged in their learning and empowered to assess their own progress.

Other countries approach the assessment of pupils' achievement in quite different ways, giving much greater autonomy over these matters to schools themselves, and placing a greater emphasis on formative assessment conducted by teachers. Critically, there is often a strong focus on assessment as an integral part of the on-going process of learning, which informs both the learning objectives for individual pupils and teachers, and for schools' self-evaluation. In other systems these considerations, rather than a limited

ASSESSMENT REVIEW GROUP

range of attainment data, are used as key components when measuring school effectiveness, forming the basis for external accountability measures by school inspectorates. The international trend is towards greater school-level responsibility and the developing view that evaluation and assessment are important tools in delivering improved student outcomes, rather than being ends in themselves (OECD, 2013a).

One of the strategies identified by OECD includes ‘...initiatives at the central level to build up a knowledge base, tools and guidelines to assist evaluation and assessment activities. These typically include detailed plans to implement student learning objectives, including guidelines for schools and teachers to develop student assessment criteria’ (OECD, 2013a). Other examples are tools for teachers to use in the assessment of their students (e.g. banks of test items), internet platforms proposing formative teaching and learning strategies, tools for the self-appraisal of teachers, instruments for school leaders to undertake teacher appraisal, and resources for school self-evaluation.

National sampling

‘Government [should] consult widely on methods of assuring school accountability which do not impact on the right of children to a balanced education ...the purpose of national monitoring ...is best served by sample testing to measure standards over time, and that cohort testing (i.e. whole school cohort) is neither appropriate nor desirable for this purpose.’

House of Commons, Children’s, Schools & Families select committee on Testing and Assessment, Third report (2008), Vol1, p66.

One of the key recommendations of the Assessment Review Group is that there should be a maximum of two sets of statutory assessments in the primary phase. It is recognised, however, that the government may have a need to monitor standards of the impact of policy decisions in a particular subject or aspect of the curriculum at a national level. Our view is that national sampling should be used for this purpose, rather than introducing a new test for the whole cohort. These results should not be used to hold individual schools to account but could provide national data against which schools can evaluate themselves, celebrate success, and plan for improvement. At a national level, this data would provide evidence of performance and should be used to inform policy decisions to improve or maintain standards in the area of testing.

These principles should apply to the current phonics screening check as well as the proposed multiplication check. Such tests are in place to influence teaching and learning in schools, highlighting aspects of the curriculum where it is believed that additional focus is needed to improve standards across the country. An initial representative sample would set a benchmark, and sampling over following years would demonstrate improvements in performance whilst teaching and learning strategies were embedded into classroom practice.

ASSESSMENT REVIEW GROUP

This type of sampling would require the government to produce a national assessment, to be externally marked, which schools would deliver to pupils. Those schools in the sample should receive their results back but these should not be published or used to hold individual schools to account. Schools could compare their results to the national data produced to inform school improvement. The national assessment should also be made available to schools which are not part of the sample, to give them the option of delivering the assessment, or aspects of it, to their pupils and enabling them to compare their results with the national data.

In the long-term, we could reach a point where national sampling is the norm rather than compulsory testing of the whole cohort.

National sampling: Belgium (Flemish speaking communities)

The Flemish government conducts Periodical Surveys, a set of large scale tests for a sample of schools and pupils. The surveys provide reliable and objective pupil performance data, giving insight into the quality of the Flemish education system, which in turn are used to inform education policy for both schools and government.

The surveys provide answers to a range of questions, including:

- To what extent have pupils achieved the final/development objectives at the end of a particular educational level? Which are being successfully achieved and which are presenting difficulties?
- Are there systematic differences between schools in the percentage of pupils achieving the final objectives/development objectives? Do these differences remain when the pupil population is taken into account?
- To what extent are performance differences associated with certain pupil, class or school characteristics?
- In the case of a repeat survey, have pupils performed better or worse than they did previously? (Eurydice, 2012)

Two surveys are conducted annually for each of the different phases of compulsory education; these generally favour testing a diversity of subjects, cross-curricular themes, the curriculum as a whole, or general skills. Surveys are repeated over time to provide longitudinal data, and may include practical tests to gain a picture of mastery of less easily measurable skills. Anonymous questionnaires for pupils, parents, teachers and management teams are used to supplement the test findings making it possible to refine and explain the results, and to identify the factors leading to better or poorer performance. Participating schools receive a feedback report which can be used for future school improvement, while those not chosen to participate can opt to take similar tests. The schools can choose to input the pupils' answers into a secure feedback system and receive a feedback report, comparing their results with the Flemish average and with schools that have a comparable pupil population, providing a value-added measure.

National assessment banks

The Assessment Review Group have recommended that the government should support the profession by setting up a national assessment materials bank. This would contain tools and materials for teachers to use when assessing pupils against the national curriculum objectives. Such a resource could support teaching and learning not only in primary schools, but also at Key Stage 3 where there have been significant concerns raised about pupil progress.

A national bank of assessment resources would increase consistency and reliability as well as being a driver for high quality teaching and learning. Resources organised in relation to units of work of the national curriculum could be used to assess pupils throughout the year and should include a wide range of activities, including tests. Exemplar materials and mark schemes should be provided. Professional training of teachers and school leaders would support consistency of standards across schools.

Although needing centralised management, it is key that such an assessment bank is not linked in any way to formal accountability. It would enable schools to assess and collect data on pupil performance and progress to inform self-evaluation and improvement.

National assessment banks: Scotland

In Scotland, schools are responsible for monitoring and evaluating their performance and progress, and must produce an annual report. Self-evaluation is regarded as the starting point for improvement, and is underpinned by *How good is our school?* published by the Scottish Inspectorate (Eurydice, 2016).

This process is supported by Scotland's National Assessment Resource (NAR), a joint development between Education Scotland and the Scottish Qualifications Authority. This provides a single place in which assessment materials can be stored to support Scotland's Curriculum for Excellence; the 3-18 curriculum that sets out the knowledge, skills and attributes that pupils and students must develop to meet the curriculum's four key capacities, across curriculum areas, stages and qualifications. The NAR includes assessment materials developed by SQA, Education Scotland and also by practitioners.

One of the key functions of the NAR is to provide a way for teachers and early years practitioners to develop a shared understanding of standards and expectations. The NAR is designed to support teachers in deepening their understanding and expertise in assessment and to develop their capacity to make sound judgements about progress and achievement. Ultimately, as confidence and understanding grow, it is expected that increasingly teachers will be in a position to add their own assessment resources to the NAR for others to use.

ALTERNATIVE MODELS FOR THE ASSESSMENT OF WRITING

The Assessment Review Group recognised that the system of national assessment of writing needs significant review. Writing should remain teacher assessed but focus on the holistic quality of a child's writing rather than the component parts. There is a need to move to a system where schools are working collaboratively in groups to moderate their judgements and improve the consistency and reliability of the process.

The group considered that comparative judgement may provide a workable and valid alternative to current arrangements for teacher assessment of writing being based on a whole rounded judgement of writing rather than a check list of individual components.

Early evidence seems to demonstrate that it is an efficient process which can produce reliable outcomes. One particular strength is that moderation could take place between many people rather than decisions resting on just one person. This could prove very useful in clusters of schools.

However, the group raised concerns regarding the potential for superficial judgements about children's writing and that it could become a mechanistic process. Depending on the way in which it is implemented, the introduction of comparative judgement as a national system of assessing children's writing could be seen as a reintroduction of a form of national writing test; this is something to be cautious of in light of the limitations and disadvantages of the previous incarnation of such a test.

Comparative judgement is currently the focus of various research and pilots and the group will await the findings and results of these to further inform the debate.

What is comparative judgement?

Comparative judgement is a way of assessing pupils' work that encourages teachers to make judgements about the overall quality and effect of pupils' writing instead of focusing on component parts.

Rather than assessing writing against a pre-determined list of criteria or a rubric, teachers are presented with two pieces of writing side by side on screen and simply have to decide which is better. This can be done for individual pieces of writing or for portfolios of work. A score is then provided for each piece or portfolio based on the judgements that have been made.

The approach has the additional advantage of allowing a higher number of people to make such judgements than would be the case through traditional moderation methods. Early small-scale studies have found high levels of reliability when compared with the standard rubric-based approach.

Single and cross key stage schools

A small but significant number of schools only cover one key stage or do not align with the standard key stages. These schools have long been poorly-served by the statutory accountability system. The Assessment Review Group has considered a range of alternative arrangements for measuring pupil progress in these schools including the idea of retaining current arrangements and statutory testing at end of Key Stage 1.

One alternative option is to apply exactly the same measure for pupils attending these schools to those that attend all through primaries. Reception teachers would carry out a start of school, observation-based assessment and Year 6 teachers would administer an end of key stage two assessment. The removal of statutory assessments at Key Stage 1 would, we hope, help to improve transition arrangements between infant and junior, first and middle schools, providing an incentive for schools to collaborate on sound and robust assessments at transition.

Statutory assessments at the beginning of primary and end of Year 6 can still provide some evidence about the provision in such schools. However, schools will also have their own assessment information which provides a wider picture of the progress of pupils in all year groups. All of these sources of data should be evaluated when holding these schools to account. A national assessment bank would allow such schools to make robust assessments of their own pupils attainment and progress to consider alongside statutory assessment data, including for pupils in the early years of Key Stage 3 in middle schools.

AFTERWORD

Nick Brook - Deputy General Secretary of NAHT

In convening the Assessment Review Group, NAHT called on the government to work with the profession to develop a fairer, more coherent system for assessment. I believe the proposals within this report have the potential to deliver on that ambition. They present an opportunity to reset our focus, so that assessment for learning is prioritised over preparation for statutory tests; to create space in the curriculum, by limiting statutory assessment to the start and end of primary school; and to recognise and value the work of all pupils beyond binary concepts of pass and fail.

There is much still to define and areas where a broad consensus is far from assured, particularly as we push beyond discussion of 'what' is needed into 'how' it should be done. We hope that it prompts early reflection and debate on these critical issues and provokes more people to make their voices heard within the government's forthcoming assessment consultation. The consultation is likely to focus on the nuts and bolts of a future assessment system. However, assessment reform cannot happen in isolation. Whilst designing the best possible assessment system we must not lose sight of the wider challenges and constraints that will directly affect the impact of reforms.

Firstly, we must look again at how data from statutory assessment is used to hold schools to account. Over-reliance on statutory assessment data raises the stakes of testing and ultimately distorts curriculum emphasis and outcomes. Unless we address some of the worst aspects of the current accountability system, including acceptance of the inherent limitations of data, even the most sensible assessment arrangements will become skewed. Floor and coasting standards cast a shadow of fear over many schools and school leaders. Poor test results can trigger an avalanche of interventions, based on a presumption of school failure, which are distracting at best and career ending at worst. It is easy to understand why schools in this shadow struggle to recruit teachers and leaders. There needs to be better join-up amongst those that hold schools to account and a more constructive approach to intervention. Most importantly, we need to replace the presumption of failure with an expectation of support.

Secondly, better governance of the assessment system is needed, leading to a stable, proportionate cycle of design, evaluation and implementation for every national assessment. Effective national test design is a complex skill which requires careful thought and substantial evaluation. The scale of national assessments in a system the size of England means that effective implementation of change is a major challenge in itself. Frequent reforms and constant tinkering around the edges can therefore have a negative impact on quality.

Thirdly, assessment for learning is not an intuitive skill possessed by all. There needs to be substantial investment in the training and development of staff in schools if this is to be done universally well. Not all schools or academy chains will have in-house expertise to draw upon and external support will come at a cost. We know that school budgets

are already at and beyond breaking point, following real-term cuts since 2010. More resources are required. Additionally, the development of national assessment banks will require investment to ensure the highest quality materials are available to schools. These cost pressures should however be offset by savings achieved by reducing the amount of statutory testing required within these proposals.

The new Secretary of State, Justine Greening, has shown a desire to listen and a willingness to set right mistakes of recent years. With political will and genuine engagement with the profession these challenges are far from insurmountable.



Nick Brook
Deputy General Secretary of NAHT

APPENDIX A

MEMBERSHIP OF THE ASSESSMENT REVIEW GROUP

The Assessment Review Group comprised of:

James Bowen, Director of NAHT Edge

Nick Brook, NAHT Deputy General Secretary

Robert Coe, Professor of Education at Durham University

Tony Draper, CEO of Lakes Academies Trust

Nansi Ellis, ATL Assistant General Secretary

Dave Ellison, Deputy Head Teacher of Foxfield Primary School, part of The Inspire Partnership

Louise Griffith, Assistant Head Teacher of St George's Church of England Primary School

Sarah Hannafin, NAHT Policy Advisor

Amanda Hulme, Head Teacher of Claypool Primary School

Kathryn James, NAHT Deputy General Secretary

Anne Lyons, Head Teacher of St John Fisher Catholic Primary School

Theo Smith, NAHT Policy Research Analyst

Professor Dame Alison Peacock, Executive Headteacher of The Wroxham School and CEO of the Chartered College of Teaching

Michael Tidd, Deputy Head Teacher of Edgewood Primary School

With thanks to the following contributors to the review:

Shahed Ahmed OBE, Head Teacher of Elmhurst Primary School

Alex Smythe, Head Teacher of Newcroft Primary School

Jan Dubiel, National Director of Early Excellence

Judy Shaw, Head Teacher of Tuel Lane Infant School and Nursery

BIBLIOGRAPHY

- Arora, A., Castle, C., Centurino, C., Martin, M., Minnich, C., Mullis, I. and Stanco, G., (2012)a. TIMSS 2011 Encyclopedia: Education Policy and Curriculum in Mathematics and Science. Volume 2: L-Z and Benchmarking Participants. Massachusetts: TIMSS & PIRLS International Study Center. Available at: http://timssandpirls.bc.edu/timss2011/downloads/TIMSS2011_Enc-v2.pdf [Accessed 12 December 2016]
- Arora, A., Castle, C., Centurino, C., Martin, M., Minnich, C., Mullis, I. and Stanco, G., (2012)b. TIMSS 2011 Encyclopedia: Education Policy and Curriculum in Mathematics and Science. Volume 1: A-K. Massachusetts: TIMSS & PIRLS International Study Center. Available at: http://timssandpirls.bc.edu/timss2011/downloads/TIMSS2011_Enc-v1.pdf [Accessed 12 December 2016]
- Bethell, G., Kellaghan, T. and Ross, J., (2011). National and International assessments of student achievement. Department for International Development. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/67619/nat-int-assess-stdnt-ach.pdf [Accessed 12 December 2016]
- Bew, P., (2011). Independent Review of Key Stage 2 testing, assessment and accountability. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/176180/Review-KS2-Testing_final-report.pdf [Accessed 12 December 2016]
- Bonci, A., (2008). A research review: the importance of families and the home environment. National Literacy Trust. Available at: http://www.literacytrust.org.uk/assets/0000/7901/Research_review-importance_of_families_and_home.pdf [Accessed 12 December 2016]
- Bradbury, A. and Roberts-Holmes, G., (2016). "They are children... not robots, not machines." The Introduction of Reception Baseline Assessment. NUT, ATL and UCL: Institute of Education. Available at: <http://webcache.googleusercontent.com/search?q=cache:RoQJE3Os354J:www.teachers.org.uk/files/baseline-assessment-final-10404.pdf+&cd=1&hl=en&ct=clnk&gl=uk> [Accessed 12th December 2016]
- Brogaard Clausen, S., Guimaraes, S., Howe, S. and Cottle, M., (2015). Assessment of Young Children on Entry to School: Informative, Formative or Performative? International Journal for Cross-Disciplinary Subjects in Education (IJCDSE), 6(1), pp. 2121-2125.
- Clark, C., (2007). Why families matter to literacy: a brief research summary. National Literacy Trust. Available at: http://www.literacytrust.org.uk/assets/0000/2038/Why_families_matter.pdf [Accessed 12 December 2016]
- Danish Ministry of Education, (2016). Evaluation, Tests & Student Plans. Available at: <http://eng.uvm.dk/Education/Primary-and-lower-secondary-education/The-Folkeskole/Evaluation-Tests-Student-and-Plans> [Accessed 11 December 2016]
- Davidson, H., Kempton, J. and Thong, C., (2015). Progress matters in Primary too: Holding schools to account consistently. CentreForum & Pearson. Available at: <http://socialwelfare.bl.uk/subject-areas/services-activity/education-skills/centreforum/progress15.aspx> [Accessed 12th December 2016]
- Department for Education, (2015). Final report of the Commission on Assessment without Levels. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/483058/Commission_on_Assessment_Without_Levels_-_report.pdf
- Dubriel, J., (2016). Effective Assessment in the Early Years Foundation Stage (2nd Ed.). SAGE Publications Ltd.
- European Commission/EACEA/Eurydice, (2012). Belgium (Flemish Community). Available at: https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/Belgium-Flemish-Community:Quality_Assurance_in_Early_Childhood_and_School_Education#Tests_for_schools



European Commission/EACEA/Eurydice, (2016). United Kingdom (Scotland). Available at: https://webgate.ec.europa.eu/fpfis/mwikis/eurydice/index.php/United-Kingdom-Scotland:Quality_Assurance

House of Commons, 2008. Children, Schools and Families Committee. Testing and Assessment, Third report, Vol. 1.

Kirkup, C., (2015). Baseline assessment. NFER. Available at: <https://www.nfer.ac.uk/pdf/baseline-assessment.pdf> [Accessed 14th December 2016]

Martin, M. and Mullis, I., (2013). TIMSS and PIRLS 2011: Relationships Among Reading, Mathematics, and Science Achievement at the Fourth Grade – Implications for Early Learning. Massachusetts: TIMSS & PIRLS International Study Center. Available at: http://timssandpirls.bc.edu/timsspirls2011/downloads/TP11_Relationship_Report.pdf [Accessed 12 December 2016]

Montroy, J., Bowles, R., Skibbe, L. and Foster, T., (2015). Social skills and problem behaviors as mediators of the relationship between behavioral self-regulation and academic achievement. *Early Childhood Research Quarterly*, 29, pp. 298–309.

OECD, (2013)a. Reviews of Evaluation and Assessment in Education: Synergies for Better Learning SUMMARY, pp. 14-17. Available at: https://www.oecd.org/edu/school/Synergies%20for%20Better%20Learning_Summary.pdf

OECD, (2013)b. Reviews of Evaluation and Assessment in Education: Synergies for Better Learning. Available at: <http://www.oecd-ilibrary.org/docserver/download/9113021e.pdf?expires=1483616767&id=id&acname=guest&checksum=38F1FFEF2B3CEB694A9149226CFB9E5A>

Ofsted, (2016). The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2015/16. (HC 821, 2015/16) London: The Stationary Office. Available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/574186/Ofsted_annual_report_education_and_skills_201516_web-ready.pdf [Accessed 13 December 2016]

Perry, T., (2016). English Value-Added Measures: Examining the Limitations of School Performance Measurement. *British Educational Research Journal*, 42(6), pp. 1056-1080. DOI: 10.1002/berj.3247

Stiggins, R., (1999). Assessment, student confidence, and school success. *Phi Delta Kappan*, 81(3), pp. 191-198.

Volante, L., (2004). Teaching to the test: What every educator and policy-maker should know. *Canadian Journal of Educational Administration and Policy*, 35(6). Available at: <https://www.umanitoba.ca/publications/cjeap/articles/volante.html> [Accessed 12 December 2016]

Whitebread, D., (2016). Which abilities of 4 year olds predict later academic achievement? Developmental evidence and implications for early assessment. Baseline Assessment: What research is telling us. Newman University, Birmingham, Wednesday 24 Feb, 2016. Available at: <http://www.newman.ac.uk/files/w3/24feb/pdf/Which%20abilities%20of%204%20year%20olds%20predict%20later%20academic%20achievement%20...%200.2%20-%20David%20Whitebread.pdf?q=540> [Accessed 14th December 2016]

Yeh, S., 2005. Limiting the Unintended Consequences of High-Stakes Testing. *Education Policy Analysis Archives*, 13,

Report of the
ASSESSMENT REVIEW GROUP

Redressing the balance

January 2017

NAHT

1 Heath Square | Boltro Road
Haywards Heath | West Sussex | RH16 1BL

 naht.org.uk

 [@NAHTnews](https://twitter.com/NAHTnews)

 press.office@naht.org.uk