



FINAL REPORT

Testing the Water

How assessment can underpin, not undermine,
great teaching

By Will Millard, Iesha Small, and Loic Menzies (LKMco)



Pearson is the world's learning company, with expertise in educational courseware and assessment, and a range of teaching and learning services powered by technology.

Our mission is to help people make progress through access to better learning. We believe that learning opens up opportunities, creating fulfilling careers and better lives.

Find out more about Pearson at www.pearson.com/uk.



LKMco works across the education, youth and policy sectors. We help organisations develop and evaluate projects for young people and carry out academic and policy research and campaigning about the issues that experience tells us matter.

Find out more at www.lkmco.org, or get in touch at info@lkmco.org.



Will Millard is a Senior Associate at LKMco, where he has led research into a wide range of issues including assessment, enterprise education, and oracy. Will has also conducted evaluations for organisations including Arts Council England and Researchers in Schools, and presented at a wide range of events and conferences across the UK. He began his career as an English teacher at a large secondary school in North West London, and then worked at a national support service for school leaders and governors. Will has visited schools around the world, has two Masters degrees in public policy, and lives in Hackney.



Ilesha Small combines her work as a Senior Associate at LKMco with a maths teaching role at Kings Langley School, Hertfordshire, where she was previously assistant head teacher. She has over 12 years' experience in UK schools in London and the East of England and is also an experienced governor in the Early Years phase. Ilesha has also been an effective independent consultant and coach where she has specialised in developing the expertise and capacity of middle and senior leaders. During her educational career, Ilesha has been seconded to the Department of Education. She was part of the team responsible for the reforms to Initial Teacher Training that have led to the current variety of school-based routes into teaching.



Loic Menzies is Director of LKMco. He specialises in education policy and research, youth development and social enterprise. He was previously a tutor for Canterbury Christ Church's Faculty of Education, an Associate Senior Manager and Head of History and Social Sciences at St. George's R.C. School in North West London and a youth worker. He holds a degree in Politics, Philosophy and Economics from Magdalen College, Oxford. He is a trustee of The Kite Trust which supports LGBT+ young people and a volunteer with the homeless charity Jimmy's Cambridge.

Contents

1. Executive summary.....	4	5. How can the accountability system change to enable teachers to shift emphasis back on to assessment that supports learning?.....	43
1.2 Summary of the problems and challenges, and ways forward.....	8	5.1 Challenge: The accountability system often has a damaging influence on teachers' day-to-day assessment practices.....	44
2. What is this research for?	10	5.1.2 Ways forward.....	45
2.1 The aims of this consultation.....	10	5.2 Challenge: The evidence is unclear on how assessments can best support evaluations of school- and system-level performance.....	47
2.2 Why are we doing this?.....	10	5.2.2 Ways forward.....	51
2.3 The process.....	11	5.3 Challenge: Teachers and parents have misperceptions about the purposes of statutory assessments.....	55
2.4 The assessment policy landscape.....	14	5.3.2 Ways forward.....	56
3. What would make teachers feel more confident and skilled when conducting assessment in the classroom?.....	16	6. How can the workload associated with assessment be reduced?	58
3.1 Challenge: Teachers' access to assessment training is inadequate.....	17	6.1 Challenge: Schools often have inefficient assessment practices.....	59
3.1.2 Ways forward: Improving access to assessment training.....	21	6.1.2 Ways forward.....	61
3.2 Challenge: The quality of assessment training is inadequate.....	23	7. How can statutory assessments and tests help all young people demonstrate their academic abilities, while providing trustworthy results?.....	66
3.2.2 Ways forward: Improving the quality of assessment training.....	21	7.1 Challenge: Statutory assessments and qualifications do not adequately capture all of pupils' achievements.....	67
4. How can assessment get the right information to the right people, at the right time?	26	7.1.2 Ways forward.....	69
4.1 Challenge: Teachers, parents, governors, the government, employers, and young people, all need and want different things from assessment.....	27	8. How can unnecessary stress about assessment be reduced for young people and their teachers?	71
4.1.2 Ways forward.....	28	8.1 Challenge: Testing can create considerable stress.....	72
4.2 Challenge: Assessments are too often focused on providing information for accountability and reporting purposes.....	31	8.1.2 Ways forward.....	73
4.2.1 Ways forward.....	35	9. Testing the Water – Pearson's Response.....	74
4.3 Challenge: Communication about assessment between teachers, and between teachers, governors and parents, is too limited.....	38	Appendices.....	77
4.3.1 Ways forward.....	40	References.....	85

1. Executive summary

The consultation objectives and research design

Teachers and parents have voiced widespread concerns about how statutory tests and exams can distract schools from their core job: supporting excellent teaching and learning.ⁱ Tests and exams are blamed for: the narrowing of the curriculum and teaching to the test; disproportionate and destructive levels of stress among teachers and pupils, and; only valuing 'certain' sorts of achievement to the detriment of broader conceptions of success.

LKMco and Pearson therefore launched 'Testing the Water', a national consultation on the future of assessment, in late 2016. The consultation defined 'assessment' broadly, taking into account impromptu verbal feedback during lessons through to statutory tests and exams, and everything in between. The project progressed in two phases:

- **Phase 1** involved focus groups and an online consultation, which enabled us to speak to and hear from hundreds of teachers, school leaders, parents, governors and young people. This led to the publication of our Interim Report in May 2017ⁱⁱ, which presented an overview of main challenges facing assessment.

- **Phase 2** has tackled the challenges from phase one head on, drawing on: a national poll of over 1,000 teachers in England; domestic and international case studies; a literature review, and; a series of think pieces from leading global experts in assessment.

What would make teachers feel more confident and skilled when conducting assessment?

Understanding and using assessment should be a "fundamental competency for all educators".ⁱⁱⁱ However, we find that:

- **Only one third of classroom teachers feel 'very confident' conducting assessment** as part of their day-to-day teaching.
- **One in five classroom teachers would not know where to look for information on assessment** if they needed it.
- **Under half of teachers received training in assessment as part of their initial teacher training**, and teachers' access to assessment training over the course of their careers is far too limited.

- **The need for training is greatest at the chalkface**; classroom teachers are less likely than their more experienced colleagues to have access to ongoing professional development in assessment.

A lack of time and money severely curtails schools' abilities to provide assessment training, and most of the available training is regarded by teachers as being low in quality. As a result, **assessment training for teachers only has a very limited impact on day-to-day practice.**

A Central Assessment Bank should therefore be created to give teachers free access to high quality assessments that serve a wide range of purposes. The Bank could build on the resources already available to teachers through websites such as 'Diagnostic Questions'.

New teachers must be better supported in developing their understanding of assessment while training, and **trainee teachers should be expected to pass a test in assessment at the end of their training year, before qualifying.** This should be part of a broader assessment of the initial teacher training curriculum. Assessment organisations and universities should provide greater **access to in-person and online training in assessment to support teachers and school governors.**

How can assessment get the right information to the right people, at the right time?

Far too much of classroom teachers' work is geared towards summative assessment.

Lesson time is wasted attempting to grade pupils' performance, rather than conducting assessments and providing feedback that identify and support next steps in learning. A lack of technical understanding of assessment and pressure to produce data for reporting and accountability swamps teachers' ability to use diagnostic formative assessment.

Communication between schools and parents about assessments is poor, and many parents feel the summative judgements their children receive from tests and exams are too vague (including in relation to age-related expectations at primary and the new numbered GCSE grades at secondary).

There is also a lack of understanding and communication about assessment reliability among teachers, governors and parents. Without an understanding of how reliable their assessments are, teachers "can't use those assessments to make smart decisions", argued Professor Dylan Wiliam.

Schools should therefore strictly limit the number of summative assessments they conduct

so that teachers can focus on diagnostic formative assessments. **Schools should also make greater use of standardised tests**, to benchmark how their pupils compare with others nationally while also shifting some of the burden of summative assessment away from classroom teachers.

Teachers need to understand assessment reliability better so that decisions they take based on assessments are appropriate and proportionate. Better training, access to quality resources, and more accessible information from assessment providers would facilitate this.



How can the accountability system change to enable teachers to shift emphasis back on to assessment that supports learning?

The accountability system exerts a powerful and often negative influence on day-to-day classroom assessment because teachers feel compelled to produce data for the purposes of reporting on and tracking pupils' progress, rather than uncovering what pupils have and have not remembered and understood and planning next steps accordingly. **Teachers have often felt unsupported during periods of accountability, curriculum and assessment reform.** Sometimes, though, **teachers and parents' understanding of the purposes of statutory tests and exams is inaccurate.**

In addition to supporting access to quality assessment training and resources, **the Department for Education must ensure schools receive appropriate time and support to implement**

curriculum and assessment reforms. We also recommend that **the Department for Education should develop a system of matrix sampling for assessing more of the National Curriculum** (while minimising the number or length of tests pupils need to sit) to monitor standards and identify where teachers need additional support in delivering subject content. In assessing more of the curriculum, such a system has the potential to reduce the incentive for teachers to teach to the test.

While statutory assessments form the bedrock of the school accountability system, **the evidence is unclear about how statutory tests and exams can best support school- and system-level improvement.** **The Department for Education should therefore build experimentation and evaluation into assessment and accountability reform,** to better understand how its reforms impact upon standards. **Schools' headline data should be published as a three-year rolling average,** to present a more rounded picture of performance and reduce the impact of year-on-year volatility.

How can the workload associated with assessment be reduced?

The workload associated with assessment is enormous, and unrealistic and unsustainable expectations are often placed on classroom teachers. **Schools often have inefficient assessment practices,** and in particular an over-reliance on:

- Heavy marking, and;
- Mock tests and exams.

High quality training and resources for teachers and school leaders would increase their knowledge and understanding of alternative assessment strategies. **New technology could also be used to cut teachers' workload** and improve the accuracy, reliability and validity of classroom assessment.



How can statutory assessments and tests help all young people demonstrate their academic abilities, while providing trustworthy results?

Many teachers, parents, governors, and young people feel that **statutory assessments do not adequately capture pupils' achievements**. This is in part because **the accountability system incentivises schools to pick certain qualifications over others**.

The Department for Education should therefore stop reporting schools' performance using the 'EBacc' performance measure, and instead focus on Attainment and Progress 8. Attainment and Progress 8 incorporate the EBacc subjects, but strike a balance between these and other valuable options, including arts subjects, and technical qualifications.

How can unnecessary stress about assessment be reduced?

Statutory assessments cause considerable and unnecessary stress for pupils, teachers and parents. This is because of their high stakes nature, but also because they are directly linked to judgements about teachers' and schools' performance. Stress is often 'passed down', from school leaders to teachers, and from teachers to their pupils.

This report suggests **schools should completely decouple pupils' test results from teachers' formal performance evaluations**. Results can be a useful starting point for developmental conversations about where teachers need additional support, but are not a reliable or fair way to measure one teacher's performance. **Pupils should also take a greater number of low stakes assessments** such as weekly multiple-choice quizzes, which provide significant educational benefit and could increase pupils' confidence.



1.2 Summary of the problems and challenges, and ways forward

Question	What are the problems and challenges?	What are the ways forward?
What would make teachers feel more confident and skilled when conducting assessment in the classroom?	Access to assessment training is inadequate and curtailed by schools' lack of spare time and money. This means: <ul style="list-style-type: none"> Class teachers lack confidence in conducting assessment as part of their day-to-day work. Class teachers are less likely than their more senior colleagues to know where to access support in assessment. Governors lack knowledge and confidence in assessment. 	<p>The Department for Education alongside other organisations should develop a Central Assessment Bank.</p> <p>Trainee teachers should pass an assessment test before qualifying.</p> <p>Assessment organisations and universities should provide in-person and online training to share their expertise with practitioners.</p>
	The quality of assessment training is inadequate, and generally only of limited help to teachers.	<p>The Department for Education alongside other organisations should use their websites to signpost towards quality assessment resources, products, and training.</p> <p>Training must cover both the theory and practice of assessment.</p> <p>Training must cater to the needs of professionals working in different roles, including:</p> <ul style="list-style-type: none"> Senior leaders; Middle leaders; Class teachers; Teaching assistants, and; Governors and parents.
How can assessment get the right information to the right people, at the right time?	Teachers, parents, governors, the government, employers, and young people, all need and want different things from assessment.	Teachers must be clear on the purpose of their assessments, identifying what information is needed, and by whom. They can then select different assessments based on this.
	Assessments too often focus on providing information for reporting and accountability purposes, meaning: <ul style="list-style-type: none"> Teachers spend more time conducting summative assessments than they would like. Schools produce lots of poor quality summative data, which can distract from – and reduce the quality of – diagnostic formative assessments. 	<p>Schools must cut the time class teachers spend conducting summative assessments so that they can focus on conducting diagnostic, formative assessments.</p> <p>Schools should use standardised tests to benchmark and report pupils' achievements.</p>
	Communication about assessment between teachers, and between teachers, governors and parents, is too limited and, in particular: <ul style="list-style-type: none"> Assessments often do not provide parents with meaningful information. No assessment is perfectly reliable and not enough people realise this. 	<p>Everyone – but especially teachers – should know how reliable assessments are.</p> <p>Assessment providers should provide easily accessible and digestible information about the reliability of their assessments.</p>

Question	What are the problems and challenges?	What are the ways forward?
How can the accountability system change to enable teachers to shift emphasis back on to assessment that supports learning?	The accountability system often has a damaging influence on teachers' day-to-day assessment practices, because: <ul style="list-style-type: none"> • Statutory tests and exams can dictate what and how pupils learn. • Teachers face undue strain during accountability, curriculum, and assessment reforms. 	The Department for Education should develop a system of matrix sampling for assessing more of the National Curriculum (while minimising the number and length of tests pupils need to sit). Teachers must be given a sensible timeframe in which to implement curriculum and assessment reforms.
	The evidence is not clear on how assessment can best support school- and system-level performance.	The Department for Education should build experimentation and evaluation into assessment and accountability reform, to better understand how its reforms impact upon standards in schools. Ofsted should enhance assessment training for its inspectors. The Department for Education should present schools' performance data as three-year rolling averages.
	Teachers and parents have misperceptions about the purposes of statutory assessments.	The Department for Education and other organisations (including schools) must ensure teachers can access high quality assessment training and resources. The Department for Education must work with stakeholders including teachers' unions to 'myth-bust' statutory assessments.
How can the workload associated with assessment be reduced?	Schools often have inefficient assessment practices, and in particular an over-reliance on: <ul style="list-style-type: none"> • Heavy marking, and; • Mock tests and exams. 	Improved access to quality training and resources could shift schools towards more efficient assessment practices, specifically: <ul style="list-style-type: none"> • Reduced use of marking; • Keeping the number of mock tests and exams to a minimum, and; • Using technology effectively.
How can statutory assessments and tests help all young people demonstrate their academic abilities, while providing trustworthy results?	Statutory assessments and qualifications do not adequately capture all pupils' achievements, because: <ul style="list-style-type: none"> • They are based upon age-related expectations. • The accountability system incentivises schools to select certain qualifications over others. 	The Department for Education should stop reporting schools' performance using the 'EBacc' performance measure, and instead focus on Attainment 8 and Progress 8.
How can unnecessary stress about assessment be reduced for young people and their teachers?	Statutory assessments result in significant and unhelpfully high levels of stress. Linking test results to teachers' performance management can be counter-productive.	Schools should use more low stakes assessments. Schools must decouple pupils' test results from teachers' performance evaluations.

2. What is this research for?

2.1 The aims of this consultation

In December 2016 Pearson and LKMco launched 'Testing the Water', a consultation seeking to better understand concerns about assessment among teachers, school leaders, parents, governors and young people, and to identify ways of addressing these concerns.

2.2 Why are we doing this?

A stimulus paper written for Pearson by Professor Peter Hill, and a Pearson-commissioned survey in June 2016, found assessment is a major concern for classroom teachers.^{iv} The survey highlighted key concerns among teachers and parents around assessment including:

- The stress assessments – and particularly 'high stakes' tests and exams – can cause young people, their teachers, and parents;
- The impact of assessments (and especially statutory tests) can have on teaching and learning in the classroom;
- The use of assessments to judge school and teacher performance, and;
- The extent to which assessments provide a fair and accurate judgement of young people's achievements.

The challenge for any education system is that assessment "is the result of a large number of trade-offs."^v This research seeks to help – as Dylan Wiliam, Emeritus Professor of Educational Assessment at the UCL Institute of Education, has suggested – make the tradeoffs "explicit, so that we are better able to judge whether the balances we strike are ones with which we are comfortable."

Pearson's survey revealed widespread support for a cross-sector debate to address some of the challenges relating to testing and assessment, and so Pearson and LKMco have sought to facilitate this debate, and help find solutions.

What do we mean by 'assessment'?

Assessment is a fundamental element of teaching and learning. Professor Wiliam writes that if pupils always learned what they were taught, "we could instead just keep records of what we had taught."^{vi} Since it is not possible to predict how different pupils will respond to the same material, assessment becomes "the bridge between teaching and learning – it is only through assessment that we can find out whether what has happened in the classroom has produced the learning we intended."

'Assessment' is a very broad term, and denotes a variety of activities that serve a number of different functions. For example:

- Diagnostic formative assessment in the classroom will help a teacher and pupil identify where gaps

exist in the young person's understanding, and consequently what the appropriate next lesson activities might be.

- Summative assessments seek to evaluate how a pupil has performed in a test in comparison with his or her peers.
- Assessments are sometimes used to provide information to parents and other groups outside schools about pupils' learning.
- Assessments are sometimes used for accountability, and used to judge teachers' and schools' performance.

This research explores 'assessment' in the widest sense, incorporating all of the activities listed, above, in order to do justice to the broad array of functions assessments serve, and the hugely varied responses people have to them.



2.3 The process

This project progressed in two main phases.

Phase 1: Identifying key concerns and challenges regarding assessment

The first phase involved speaking to a broad range of stakeholders about their concerns regarding assessment, and led to the publication of our “Testing the Water: Interim Report” in May 2017.^{vii} Findings from the first phase of the project are referenced throughout this report, and are updated to incorporate online consultation responses received since May 2017.

Phase 1 of the project focused on areas highlighted in a stimulus paper by Professor Peter Hill, an expert in international assessment systems, and in Pearson’s 2016 survey. These led to the initial set of consultation questions, which are outlined, below.

Workshops

We ran 17 workshops across the country. Because of their critical role in conducting assessment day-to-day, 12 of these workshops were with classroom teachers and school leaders. We also ran workshops with parents, school governors and trustees, and young people themselves. In total, we spoke to over 150 people during this process.

We worked with a wide range of organisations including the NAHT, ASCL, Challenge Partners, and the National Governance Association to help arrange these workshops, although the views expressed and conclusions drawn in this report do not specifically reflect those of any single organisation mentioned. A full list of the workshops is given in Appendix 1.

Workshop questions

1. Are teachers’ negative attitudes towards assessment inevitable, or could this change? If so, how?
2. Why do teachers and parents feel more negatively about some forms of assessment than others? What can we do about this?
3. What explains parents’ belief that their children are not adequately prepared for assessments and tests? How could this be improved?
4. What might be done in the longer term to resolve tensions between the validity, reliability and credibility of assessment?
5. Could we redress the balance between assessment for learning, assessment for accountability and assessment for qualifications?
6. What should teachers’ role in assessment be?
7. How can we make assessment (particularly formative assessment) more efficient and effective? Can technology help?
8. Whose responsibility is innovation in assessment?
9. Do you have any other thoughts not covered by these questions?

Online consultation

To complement the depth offered by the workshops, we concurrently ran an online consultation, asking respondents three questions:

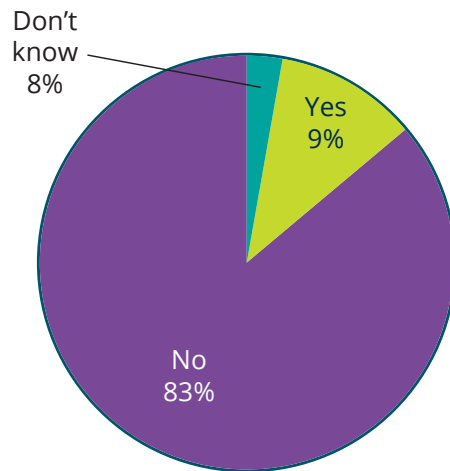
Online consultation questions

1. What is your biggest concern about assessment?
2. What is assessment useful for?
3. How would you improve current approaches to assessment?

This survey was open for the duration of the project, and in total we collected 752 responses. We heard from:

- 422 respondents currently working in a school (with 17 working in Early Years settings, 58 in primary, 310 in secondary, and 106 in post-16);
- 208 school governors or academy trustees, and;
- 186 parents of children currently attending school.
- More than four in five respondents said they had never responded to a consultation on assessment before.

Have you responded to a consultation about assessment before? (n=657)



Respondents were asked for their postcodes and, as the mapped responses indicate, they work and live across the country.

A more detailed breakdown of our respondents, including by school type and job role, is given in Appendix 2.

Responses to the online consultation were coded inductively, helping to identify broader trends in how people feel about assessment.



Phase 2: Identifying solutions and ways forward

Following phase 1, six solution-orientated questions were chosen for phase two:

Questions underpinning phase 2:

1. What would make teachers feel more confident and skilled when conducting assessment in the classroom?
2. How can assessment get the right information to the right people, at the right time, so that practitioners, young people, their parents, governors, employers and government obtain the information they want and need?
3. How can the accountability system change to enable teachers to shift emphasis back on to assessment that supports learning?
4. How can the workload associated with assessment be reduced for teachers?
5. How can statutory assessments and tests help all young people demonstrate their academic abilities, while providing trustworthy results?
6. How can unnecessary stress about assessment be reduced for young people and their teachers?

Expert interviews

Ten experts in assessment were interviewed, and each was asked the questions outlined, above. These experts were:

- **Alex Quigley**, Director of Learning and Research, Huntington School.
- **Becky Allen**, Director, Education Datalab.
- **Christine Counsell**, Director of Education, Inspiration Trust.
- **Daisy Christodoulou**, Director of Education, No More Marking.
- **Dylan Wiliam**, Emeritus Professor of Educational Assessment, the UCL IOE.
- **Harry Torrance**, Professor and Director of the Education and Social Research Institute, Manchester Metropolitan University.
- **Michael Tidd**, Headteacher, Medmerry Primary School, and education blogger.
- **Rob Coe**, Professor of Education and Director of the Centre for Evaluation and Monitoring (CEM), Durham University.
- **Simon Knight**, Director of Whole School SEND, part of the London Leadership Strategy.
- **Tim Oates CBE**, Group Director of Assessment Research and Development, Cambridge Assessment.

These experts' input is synthesised and outlined throughout this report, although unless otherwise stated the recommendations do not reflect the views of any specific expert.

National survey of teachers

In addition to our online consultation, YouGov was commissioned by LKMco and Pearson to survey teachers regarding their views on assessment.

In total, 1,002 teachers were surveyed. Fieldwork was undertaken between the 5th and 16th October 2017, and carried out online.

A breakdown of where respondents work, and of their roles, is given in the appendices.

Case studies

A series of case studies was selected based on the literature and discussion with experts. Case studies were chosen because they help illustrate innovative, constructive practice in relation to the six questions underpinning phase two of the research. They were conducted between August and October 2017.

The case studies highlight work taking place first and foremost in schools as well as other organisations across England.

Three international case studies were also conducted, examining how other jurisdictions tackle some of the challenges highlighted in phase one. These international case studies were selected, again, with reference to the literature and in consultation with experts.

The inclusion of international case studies is not intended to imply there is a 'magic bullet' solution to any of the challenges discussed. The context in which assessment and wider educational policies are enacted is hugely influential in governing their success. This report is not suggesting any idea be simply 'copied and pasted' into the English system. Their inclusion is intended to stimulate discussion,

but also ensure important examples of innovative practice beyond this country are acknowledged.

A full list of the case studies undertaken for the report is given in the appendices.

300-word 'magic wand' think-pieces

In addition to speaking to a range of experts, a selection of educationalists were asked to contribute short thinkpieces, addressing the statement: 'If I could wave a magic wand and change one thing about assessment in English schools, it would be...'. These are incorporated throughout this report.

Thinkpieces were contributed by:

- **Adam Boddison**, Chief Executive, nasen.
- **Dame Alison Peacock**, Chief Executive, Chartered College of Teaching.
- **Allana Gay**, Deputy Headteacher, Lea Valley Primary School.
- **Daisy Christodoulou**.
- **Professor Dylan Wiliam**.
- **Emma Knights**, Chief Executive, National Governance Association.
- **Geoff Barton**, General Secretary, ASCL.
- **Professor Harry Torrance**.
- **Mary Bousted**, General Secretary, National Education Union.
- **Michelle Doyle Wildman**, Acting CEO, PTA UK.
- **Neil Carmichael**, former Chair of the Education Select Committee.
- **Paul Whiteman**, General Secretary, NAHT.
- **Professor Rob Coe**.
- **Tim Oates CBE**.

2.4 The assessment policy landscape

2010 – 2011:

- The 'Importance of Teaching' White Paper, published in November 2010, set out the government's ambition to reform the curriculum and assessment at primary, secondary and post-16.^{viii}
- The 'English Baccalaureate' ('EBacc') was announced in 2010, and was included for the first time in 2010 performance measures. It includes English, maths, science, a language and a humanities GCSE.
- Lord Bew's 2011 review of Key Stage (KS) 2 testing and accountability found schools should be held accountable for the education of their pupils, but suggested placing greater emphasis on teacher assessment judgements in the accountability system.^{ix}

2012 – 2013:

- A new universal phonics screening check was rolled out in 2012.
- Pupils sat a grammar, punctuation and spelling test at the end of KS2 for the first time in 2013.

2014:

- Ofqual released its review of exam marking in A-levels, GCSEs and other academic qualifications in February 2014, suggesting better use could be made of on-screen marking.^x
- Schools taught the new National Curriculum from September 2014.
- Levels - which had been the system used by most primary and secondary schools for monitoring pupils' attainment and progress - were abolished in September 2014.

2015:

- Pupils sat KS1 and KS2 assessments under the old National Curriculum for the final time in the summer, 2015.
- The 'Commission on Assessment Without Levels' released its report in September 2015, highlighting the conflicting pressures assessment could place on teachers, and the need for alignment between assessment, curriculum and accountability policy.^{xi}
- New GCSE courses in English language and literature, and maths, were first taught from September 2015 (to be sat in summer 2017, results in August 2017). The new GCSEs were assessed mainly through exams, and graded 9 to 1. The remaining GCSE subjects would be introduced over two years from September 2016.
- New AS and A-level courses were taught from September 2015 (with further courses to be introduced in 2016 and 2017). The new courses would not be split into modules, and assessed mainly by exam.



2016:

- The three independent teacher workload review groups published their respective reports in March 2016, outlining concerns about marking, planning and resources, and data management, and possible ways of addressing the concerns.^{xii}
- The government scrapped plans for the Reception baseline assessment (first proposed in 2014).
- Pupils took KS1 and KS2 assessments based on the new National Curriculum for the first time in the summer.
- The Education Select Committee launched its inquiry into primary assessment.
- The Rochford Review, was published. It focused on the assessment of pupils with special educational needs or disabilities (SEND) and recommended moving away from using P scales to assess children working below the standard of National Curriculum tests.^{xiii}
- The government published findings from its review of the Standards and Testing Agency in November 2016, finding the body had shortcomings but could continue to develop and deliver primary assessments.^{xiv}
- Ofsted's Annual Report 2015/16, published in December 2016, found evidence to suggest schools' curricula were being narrowed because of a focus on core subjects in statutory testing.^{xv}

2017:

- The Education Select Committee published findings from its inquiry into primary assessment, arguing that high stakes testing has been "harming teaching and learning in primary schools".^{xvi}
- In August, pupils were awarded grades 9 to 1 for the first time following new English and maths GCSEs.^{xvii}
- Ofqual announced it would run annual National Reference Tests – one in English, and one in maths – to monitor changes in standards over time at GCSE.^{xviii}
- In September the government published its response to its primary assessment consultation, explaining it would – among other things – introduce an Early Years baseline assessment, remove the statutory requirement to report KS2 teacher assessment data, introduce a year 4 multiplication check, and introduce a more flexible approach to writing assessment.^{xix}
- The government also published its response to the Rochford Review recommendations about the assessment of pupils working below the level of the National Curriculum, saying it would remove the requirement to assess pupils engaged in subject-specific learning using P scales, and pilot an approach to peer-to-peer moderation.^{xx}

- Ofsted published its new 2017 to 2022 strategy, explaining among other steps, that it would seek to better understand the impact of its grading system.^{xxi}
- Ofsted published the findings from its curriculum review, flagging its concerns around the amount of debate and reflection in schools about the curriculum, and the narrowing of primary and KS3 curricula.^{xxii}



3. What would make teachers feel more confident and skilled when conducting assessment in the classroom?

What are the problems and challenges?	What are the ways forward?
<p>Access to assessment training is inadequate and curtailed by schools' lack of spare time and money. This means:</p> <ul style="list-style-type: none"> • Class teachers lack confidence in conducting assessment as part of their day-to-day work. • Class teachers are less likely than their more senior colleagues to know where to access support in assessment. • Governors lack knowledge and confidence in assessment. 	<p>The Department for Education alongside other organisations should develop a Central Assessment Bank.</p> <p>Trainee teachers should pass an assessment test before qualifying.</p> <p>Assessment organisations and universities should provide in-person and online training to share their expertise with practitioners.</p>
<p>The quality of assessment training is inadequate, and generally only of limited help to teachers.</p>	<p>The Department for Education alongside other organisations should use their websites to signpost towards quality assessment resources, products, and training.</p> <p>Training must cover both the theory and practice of assessment.</p> <p>Training must cater to the needs of professionals working in different roles, including:</p> <ul style="list-style-type: none"> • Senior leaders; • Middle leaders; • Class teachers; • Teaching assistants, and; • Governors and parents.

3.1 Challenge: Teachers' access to assessment training is inadequate

Teachers' access to training is curtailed by schools' lack of spare time and money

Limited funds

School budgets are tight, perhaps making senior leaders hesitant to spend money on assessment training. The Teacher Development Trust has found that over 21,000 teachers are employed in schools reporting zero or near-zero expenditure on CPD, and the median spend on CPD across the sector is 0.7% of schools' overall budgets.^{xxiii}

Poor quality training can discourage schools and teachers from participating

The training teachers do receive in assessment (as in other areas of their practice) can be low in quality. This creates a vicious cycle, whereby teachers do not seek CPD because what they have experienced in the past has been poor.

This vicious cycle leads to the creation of another, whereby teachers do not access training that could give them strategies that would lighten their workload. This high workload in turn reduces the likelihood that they access training.^{xxiv}

Training is often not sustained for long enough

England's teachers work longer hours, on average, than teachers in other jurisdictions, and spend considerably less time in on-going training.^{xxv}

This is in spite of evidence that suggests effective professional development should span at least two terms, and preferably a year or longer.^{xxvi} In relation to assessment, this presents a key challenge because, as Rob Coe, Professor of Education and Director of the Centre for Evaluation and Monitoring (CEM) at Durham University Coe argued:

"If you want to change teacher's behaviour, and you want to engage with something that's as complex as assessment, you're not going to do this in half day or twilight session."

Professor Rob Coe

Class teachers lack confidence in conducting assessment as part of their day-to-day work

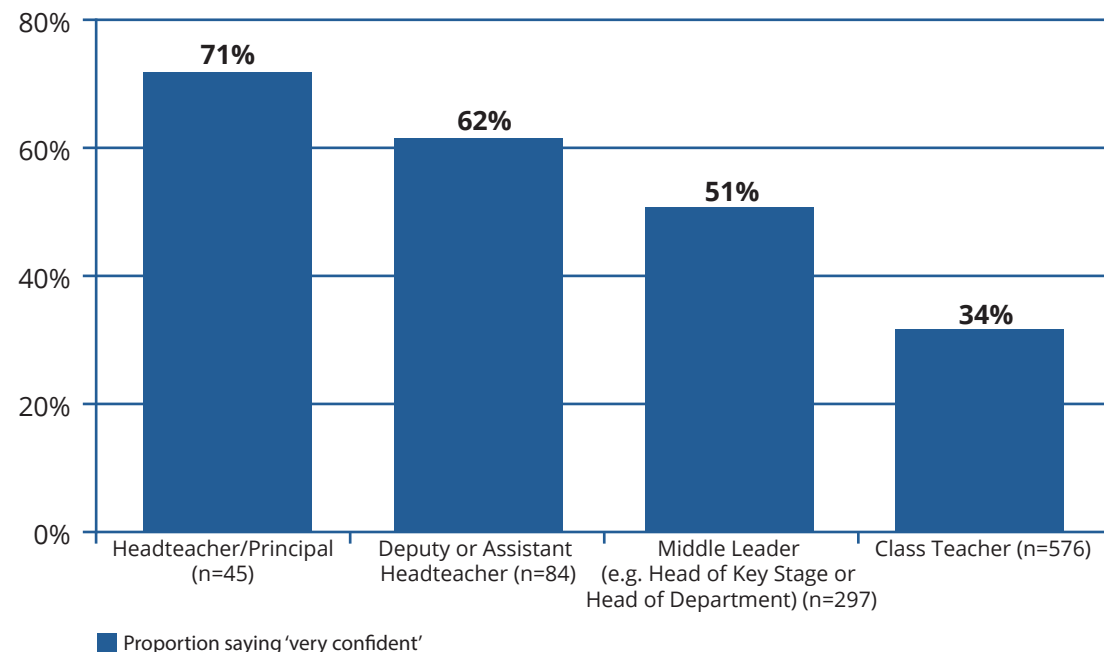
Confidence varies with seniority

There is widespread agreement in academic literature that assessment literacy should be seen as a "fundamental competency for all educators":^{xxvii}

"[E]ducators' inadequate knowledge in assessment can cripple the quality of education. Assessment literacy is seen as a sine qua non for today's competent educator."

Popham, 2009: p. 4

How confident do you feel day-to-day conducting assessment as part of your teaching?



However our poll of teachers reveals that only one a third of classroom teachers feel 'very confident' conducting assessment as part of their day-to-day teaching, and our online consultation revealed concerns (particularly amongst parents and governors) about teachers' ability to conduct assessment, and participants in our focus groups stressed the need to improve access to training.

Practitioners' confidence increases with seniority: headteachers and principals were the most likely to report feeling 'very confident'.¹

These findings are consistent with previous research. For example, the NAHT's Commission on Assessment in 2014 found that teacher training in assessment across the board, including on-going professional development, "was not of a sufficiently high or rigorous standard."^{xxviii} It argued:

"[T]eachers are not automatically equipped to assess, even though there is an apparent assumption that this is the case. They need practical training in assessment methodology and practice and an ongoing programme of CPD."

NAHT, 2014: p. 17.

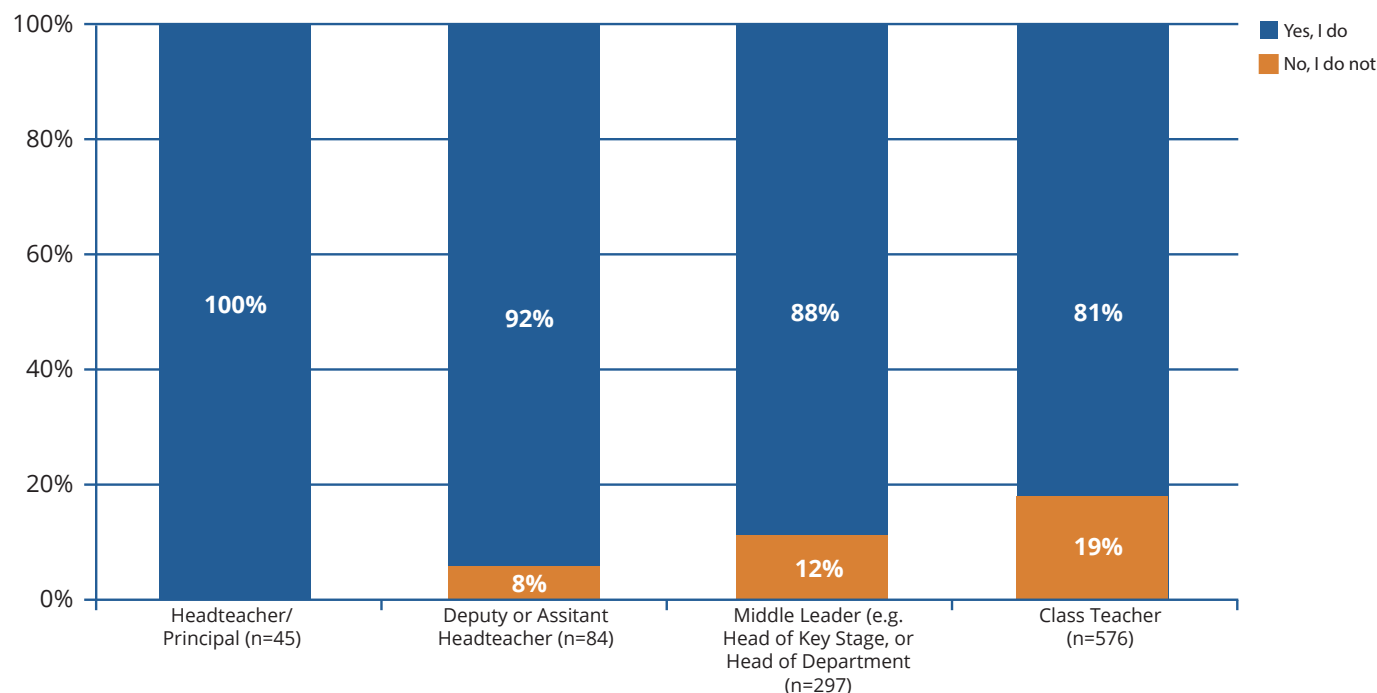
Furthermore, the NFER's 2016 Teacher Omnibus Survey found over three quarters of senior leaders and classroom teachers agreed they or their staff would benefit from "some type of additional training on assessment without levels."^{xxix}

The Carter review of initial teacher training found in 2015 that, of all the areas in which trainees receive support, the most significant improvements were needed in relation to assessment.^{xxx} Furthermore, the review found "significant gaps in both the capacity of school and ITT providers in the theoretical and technical aspects of assessment."

Confidence varies by subject

Whilst most science teachers (53%, n=92) report feeling very confident assessing, fewer than half of English (44%, n=134) and maths (40%, n=78) teachers share this confidence. Science teachers were more likely to have attended within-school or external professional development in assessment.

If you wanted or needed information about assessment, do you know where you would be able to get it from?



¹ 45 headteachers and principals responded to the poll. With this size sample, the finding should be regarded as indicative only.

Class teachers are less likely than their more senior colleagues to know where to access support in assessment

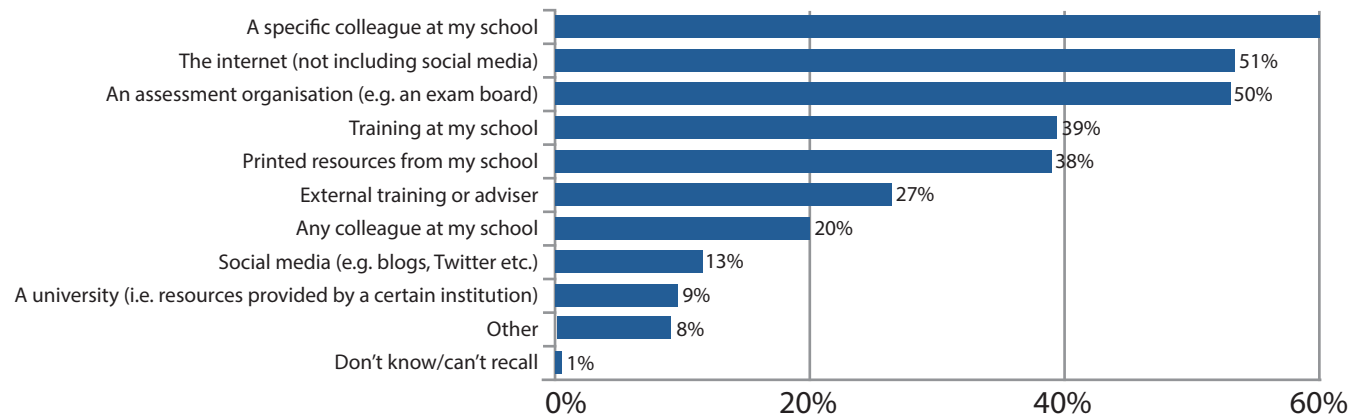
85% of all respondents said they would know where to find information about assessment if they wanted it. However, practitioners' access to information about assessment appears to increase with seniority, with 100% of headteachers and principals saying they would know where to find information, compared to only 81% of class teachers.

Given that only a third of class teachers feel 'very confident' in conducting assessment day-to-day it is worrying that one in five would not know where to look for support given the critical role of assessment in class teachers' work.

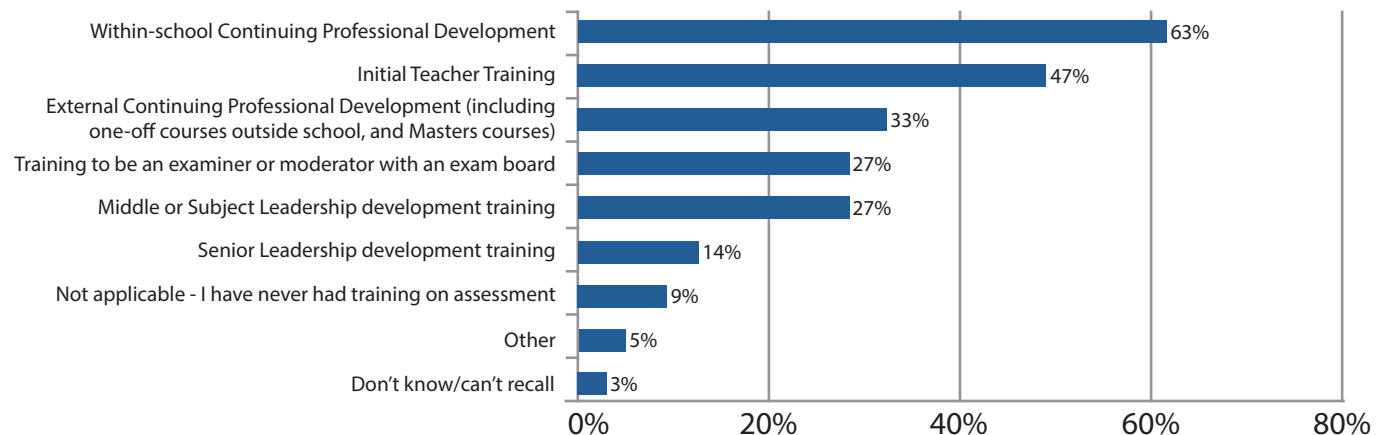
Teachers rely on colleagues for support with assessment

The most commonly cited source of information or advice on assessment is asking a specific colleague. Patterns vary to some extent by phase and, to a lesser extent, by subject, with primary teachers (n=285) significantly more likely than their secondary (n=424) or FE (n=106) colleagues to speak to someone else at their school, use a printed resource, or attend within-school training. Secondary practitioners, however, are significantly more likely than their Early Years (n=47), primary, post-16 (n=50) or FE colleagues to use social media, with 18% reporting that this is where they would look.²

Where would you go for this information? (n=851)



Have you ever received training on assessment as part of...? (Select all that apply) (n=1,002)



² With fewer than 50 Early Years respondents, this finding should be taken as indicative only.

Access to training varies by phase

Primary school practitioners were the most likely to have attended within-school training, with 70% saying they had participated in such training. Early Years (57%) and FE (53%) practitioners were the least likely to have attended within-school training. Primary practitioners were also significantly more likely than their secondary colleagues to have attended senior leadership development in assessment.

Access to training varies by subject

A significantly higher proportion of English (70%) and science (74%) teachers have attended within-school training in assessment compared to maths teachers

(46%), and a significantly higher proportion of science teachers (45%) said they had attended external assessment training than maths teachers (29%).

Governors lack knowledge and confidence in assessment

Many contributors – not least governors themselves – noted governors could lack knowledge and understanding of assessment, including:

- Underlying theories of assessment, such as the difference between formative and summative assessment;
- The rationale for, and administration of, statutory tests and exams, and;

- How to analyse assessment data, including internal tracking data, and the results of external tests and exams.

A key issue is that governing bodies often rely on their headteacher for the information they have access to, and then how to interpret this information. This impedes governors' ability to hold their headteachers to account and to ask effective questions.

Training and support in assessment – Finland

Teacher training in Finland is looked upon with envy by teachers in other countries. Training is highly selective, with only around 10% of applicants (and often fewer) accepted onto the courses, which last five years. However, some Finnish teachers are sceptical about the extent to which their training supported their use of assessment in the classroom. Several said they did not recall receiving any specific initial training in assessment at all, and Pekka Peura, a teacher and principal, said, "I think they tried to teach me something about how to assess or evaluate but I didn't understand anything at the time!"

Instead, teachers often learn about assessment

on the job. "Teachers are very active in developing their own professional skills", explained Minna Welin, a vice principal. Schools often have 'pedagogical teams', which support one another in particular subject areas or with specific year groups by discussing challenges, conducting developmental lesson observations, or team teaching.

Reflecting on the level of autonomy teachers have in their classrooms, Pekka said teachers "want a lot of training, because nobody is telling [us] what to do. We have to figure it out on our own." Furthermore, there has been limited training in the curriculum reform introduced in 2016, and Mikaela Sumeli, an elementary school teacher, felt teachers would benefit from greater guidance in this respect.



3.1.2 Ways forward: Improving access to assessment training

The Department for Education alongside other organisations should develop a Central Assessment Bank

Access to training is critical, Professor Wiliam explained, because “there’s nothing so motivating as being good at something, so the key to improving teachers’ confidence with assessment is to make them better at assessment.”

Practitioners’ confidence and ability to conduct different sorts of assessment could be improved by establishing a Central Assessment Bank, containing a large range of quality assured formative and summative assessment resources and training materials. In effect this would be an assessment ‘one stop shop’. Teachers would pick, choose and adapt relevant resources based on their curricula.

A Central Assessment Bank could provide quality assured resources online. Quality assurance could be based on:

- 1 An expert panel or contracted organisation vetting content using agreed criteria, before it is made available publicly.
- 2 Users’ ratings and reviews. However, whilst cheaper and more manageable, this would be vulnerable to poor use given the lack of understanding about some elements of assessments among teachers.

Work already exists in this area. For example the Diagnostic Questions website,^{xxxii} allows teachers to create their own assessments or access a bank

of questions with which to diagnostically assess pupils. Other assessment item banks also exist, but these are often behind pay walls, or have been developed in other jurisdictions (and therefore relate to different curricula).^{xxxii} A Central Assessment Bank could therefore be developed in partnership with existing providers.

The Department for Education has an important role to play in these proposed initiatives, but should work with professional organisations including the Chartered College and unions, academic institutions including universities, and assessment organisations.

Trainee teachers should pass an assessment test before qualifying

Given the critical role assessment plays in supporting teachers’ practice, trainee teachers should be expected to pass a test in assessment at the end of their training year, before qualifying. This should be part of a broader assessment of the initial teacher training curriculum. The test could assess trainees’ understanding of overarching assessment principles, such as those outlined in section 3.2.2, and in particular concepts of validity (including reliability, and sources of unreliability). Training in other areas of assessment should then form an integral part of teachers’ further professional development.

Furthermore, the test could evaluate practitioners’ ability to draw appropriate inferences from different sorts of assessment evidence. This would *not* involve grading or levelling work; rather, this could entail using multiple-choice questions or ‘true or false’ statements about what conclusions can reasonably be drawn from different assessments, possibly including real examples of pupils’ work. This part of the test could be phase- or subject-specific.

Assessment triangles – Ontario, Canada

Following a 2013 policy memorandum by the Ontario Ministry of Education about improving diagnostic assessments, School Boards (loosely the equivalent of local authorities in England) have worked with teachers’ unions in certain districts in Ontario to create tiered assessment ‘triangles’ in reading and numeracy.

The triangles give teachers a menu of assessments to choose from, in order to help them unpick and better understand pupils’ needs. Assessments are grouped into three tiers:

- 1 The bottom tier contains diagnostic and summative assessments to help teachers make judgements about pupils’ progress against age-related expectations in reading or numeracy.
- 2 The middle tier provides more specific assessments that address possible developmental gaps in reading or numeracy.
- 3 The top tier contains assessments to help establish where pupils may need more extensive and targeted support, and include some assessments of SEND.

Jenn Clark, a school administrator (the school’s most senior manager) argued that the triangles are helpful because they support teachers’ professional judgement rather than mandating particular courses of action. Andrea Gillespie, a school superintendent agreed, saying “there are so many assessment choices that are out there, we felt a need to support our teachers in using their professional judgement, but that it be grounded in research”.

Training providers would need to support trainees acquire the knowledge and skills needed to pass the test.

Assessment organisations and universities should provide in-person and online training to share their expertise with practitioners

Assessment providers and other organisations with extensive knowledge of assessment, including universities, should offer in-person and online training to teachers and governors.

Many assessment organisations offer training for practitioners who want to become examiners and moderators. However, they should offer access to broader training that helps practitioners develop their assessment knowledge and skills.

Training should adhere to the principles of quality professional development (discussed at the start of this section), and be differentiated for practitioners with differing levels of assessment expertise. Training should also be available for governors.



If I could wave a magic wand...

Creating a professional learning culture around assessment: Professor Rob Coe, Director of the Centre for Evaluation and Monitoring (CEM), Durham University

Exactly how magic is this wand? Presumably it is cheating to say the one thing I would change is to make all assessment perfect, in every way, forever?

If I have to be more specific, I think there are three main things I would like to see improved. The first is about the quality of assessments. Quality here includes things like the precision of scores (reliability), alignment between the things learners have to demonstrate to achieve high scores and the kinds of learning we value and promote (construct validity), freedom from biases, and a lack of both construct-irrelevant variance and construct under-representation. The second is about the capacity to use the information and feedback from assessment processes to inform the learning process in ways that are optimal. The third concerns the ways assessment is used as part of the learning and consolidation process. These include retrieval practice, the testing effect (in which weighing the pig really does make it heavier), and a consideration of forgetting as a natural and expected part of learning, but one that can be overcome by design and good use of assessment.

So that looks like three things, and my magic wand can only deliver one change. But I don't think this is cheating, because all three depend on the same change: building the skill, understanding, expertise and experience of teachers in their use of assessment. Hence, my magic wand will create an infrastructure and culture that promotes and requires substantial, sustained and effective professional learning about assessment as a routine expectation for every teacher. From that, the rest follows.

3.2 Challenge: The quality of assessment training is inadequate

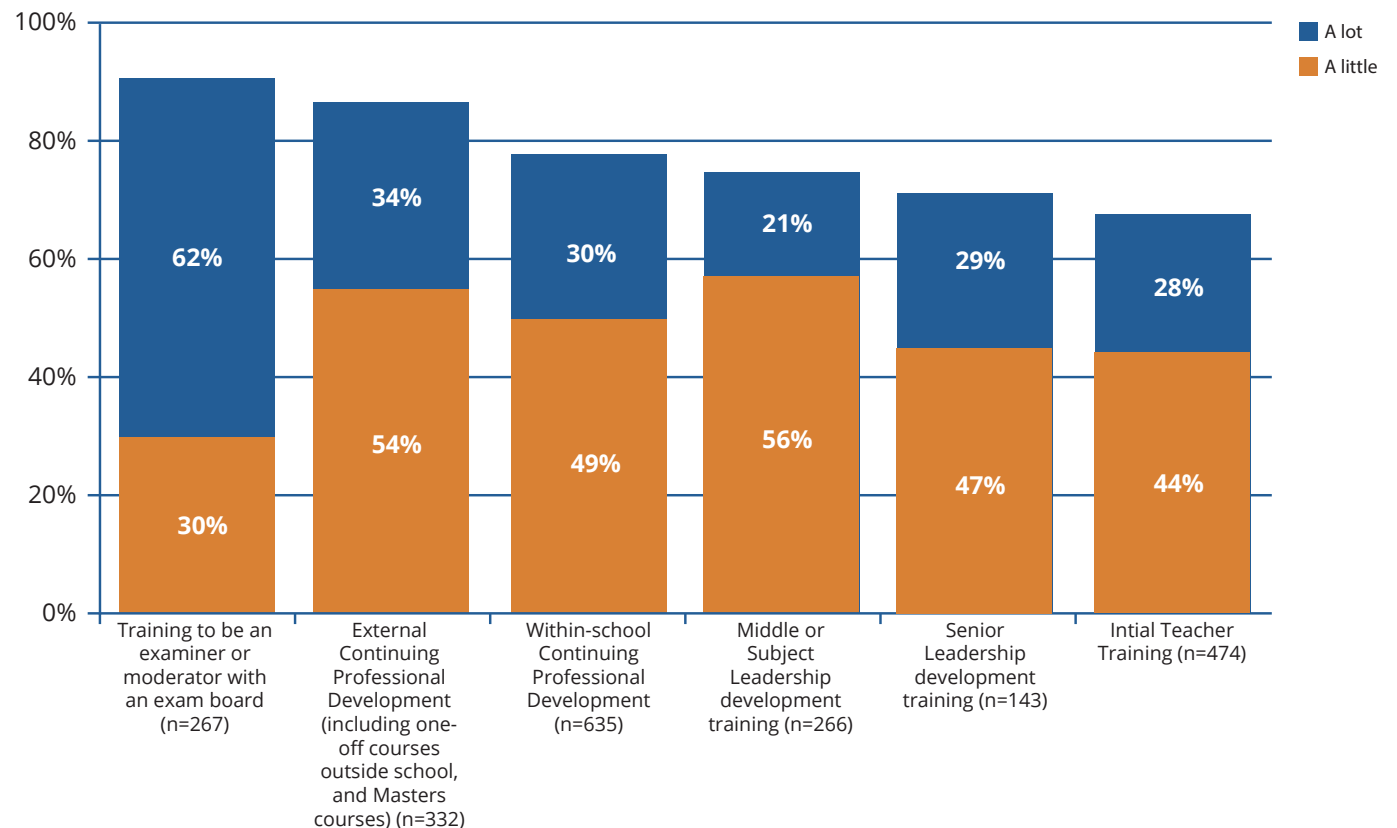
Assessment training is generally only of limited help to teachers

When teachers do participate in assessment training, they tend not to find it useful in improving their practice.

Responses to our poll indicate practitioners feel training as an examiner or moderator is a comparatively useful form of assessment training, perhaps because the skills involved in these roles relate directly to the content they teach. However, Menzies argues that such training has important shortcomings, not least that many teachers feel the quality of training they receive to become an examiner is very variable.^{xxxiii} Furthermore, such training disproportionately benefits secondary and post-16 teachers, who are preparing pupils for exams.

Teachers are less positive about the impact of initial teacher training on their assessment competence, corroborating the weaknesses of this training as identified by the Carter Review. Meanwhile middle and senior leadership development opportunities are also not as effective as they could or should be.

To what extent would you say training you received on assessment whilst undertaking each of the following has improved your practice of assessment?



3.2.2 Ways forward: Improving the quality of assessment training

The Department for Education alongside other organisations should use their websites to signpost towards quality assessment resources, products, and training

Definitions of 'quality' may vary between organisation, so the rationale for signposting towards particular assessment materials or training should be made explicit. By signposting towards materials, the Department for Education would not be mandating any resources but simply improving the information available to practitioners when making decisions.

Training must cover both the theory and practice of assessment

Training that covers both the principles and practice of assessment would improve classroom instruction, helping teachers better diagnose and address pupils' needs. As Daisy Christodoulou, Director of Education at No More Marking, points out:

"In order to interpret the information provided by assessments, teachers need to understand fundamental assessment concepts."

Daisy Christodoulou

An assessment training curriculum should therefore include the following elements:

1. Assessment in theory

Training should support practitioners' understanding of assessment theory. The following areas are especially important, and should be sequenced logically over time with the core foundations (in particular different sorts of inferences and issues of validity) included during initial teacher training:

- Formative and summative inferences, and the assessments that can support these;
- Validity (including concepts of reliability, and sources of unreliability);
- The interrelationship between assessment, curricula, and pedagogy. This should help teachers and school leaders understand how knowledge ties together and the ramifications this should have on curriculum structure. This would lead to greater understanding of how assessment can be used to maximise pupils' grasp of key knowledge and skills.
- An overview of how memory works, including the distinction between working and long term memory, and how different assessments can develop memory;
- How forms of SEND can impact upon children's development (and how this might be identified through assessment, and what appropriate subsequent courses of action might be);
- Approaches to grading, including norm, criterion and cohort referencing;
- Statistical measures, including significance and confidence intervals, and;
- The function of statutory assessments, how results are calculated, and how results are used.

2. Assessment in practice

It is also critical to develop practical assessment skills. Again, these should be sequenced logically over time, although the priority for trainee and new teachers will likely include the design and use of formative assessments and interpreting the results of these. Training should help teachers and school leaders:

- Ask pupils great questions;
- Design and undertake summative assessments;
- Interpret and act upon assessment data;
- Calculate the reliability of assessments administered during class;
- Build curricula, using assessment to support the transferral of knowledge and skills;
- Use assessment to identify pupils' deeper-rooted developmental needs, where necessary, and;
- Use standardised tests to make comparisons with pupils nationally.



Training must cater to the needs of professionals working in different roles

Training must meet the needs of different individuals in different roles. In particular:

- **Classroom teachers:** these individuals would particularly benefit from learning about drawing formative inferences from assessments about pupils' learning;
- **Senior and middle leaders:** these individuals tend to be responsible for developing department- and school-level assessment policies. Training for these groups should therefore ensure they can decouple formative and summative assessment;
- **Teaching assistants:** they have an important role to play in making classroom assessment more effective, although their job is not to replace the class teacher. Teaching assistants must be able to judge when to defer to the class teacher, and feel comfortable with the notion that they need not be responsible for solving every learning problem they see. There is excellent practice with regards to this in the Early Years, where teaching assistants do not perform a diagnostic function but support data-collection by acting as an extra pair of "eyes and ears" for teachers, as Rob Webster, Director of the Maximising the Impact of Teaching Assistants Project at the UCL IOE explained.
- **Governors and parents:** these groups play a key role in supporting young people's education but often lack the knowledge to unpick the assessment taking place at school.^{xxxiv}

Increasing teachers' assessment literacy – Assessment Academy

Assessment Academy is a training platform for school leaders and teachers, developed by Evidence Based Education (EBE). It was established in response to a perceived gap in the market in relation to assessment training for teachers. Namely, while Masters courses could be expensive and time-consuming, other courses were often quick but ineffective.

The programme adheres to evidence on what makes CPD effective, and Jack Deverson, EBE's Managing Director, explained:

- At least two participants per school must take part, including one senior leader with responsibility for assessment, and a middle leader heading up a phase or subject department. This is to facilitate peer support and collaboration, and encourage buy-in for implementation of new practices.

- The programme lasts for about 50 hours, spread out across three terms (meaning 1 to 2 hours of training a week). A little under half this time is spent learning about assessment theory. The next stage of the learning is guided and structured practice, whereby participants put the ideas, tools and resources into use in their classrooms. Finally, participants support colleagues in their departments or phases to develop effective practices in assessment.

Participants have access to a range of tools to support assessment in their settings, including an assessment reliability calculator. Importantly, participants do not just get the tools, but are supported in using them to improve the quality of their assessments and, therefore, the quality of information they provide. Jamie Scott, EBE's Head of Partnerships, said, "After all, if you don't know the reliability of an assessment, how confident can you be in the information it produces?"

4. How can assessment get the right information to the right people, at the right time?

What are the problems and challenges?	What are the ways forward?
Teachers, parents, governors, the government, employers, and young people, all need and want different things from assessment.	Teachers must be clear on the purpose of their assessments, identifying what information is needed, and by whom. They can then select different assessments based on this.
Assessments too often focus on providing information for reporting and accountability purposes, meaning: <ul style="list-style-type: none"> Teachers spend more time conducting summative assessments than they would like. Schools produce lots of poor quality summative data, which can distract from – and reduce the quality of – diagnostic formative assessments. 	Schools must cut the time class teachers spend conducting summative assessments so that they can focus on conducting diagnostic, formative assessments. Schools should use standardised tests to benchmark and report pupils' achievements.
Communication about assessment between teachers, and between teachers, governors and parents, is too limited and, in particular: <ul style="list-style-type: none"> Assessments often do not provide parents with meaningful information. No assessment is perfectly reliable and not enough people realise this. 	Everyone – but especially teachers – should know how reliable assessments are. Assessment providers should provide easily accessible and digestible information about the reliability of their assessments.



If I could wave a magic wand...

Helping parents understand what assessments are for: Michelle Doyle Wildman, Acting CEO, PTA UK

Assessments would be undertaken without fuss and drama, and with clear purpose, so that parents recognise their importance and better understand what taking these assessments means specifically for their children.

It's a reality that some assessments are used to ensure schools are delivering the appropriate standard of education to the children in their care. They reveal areas for improvement and in so doing increase the quality of education in schools. But all too often the term 'assessment' is bandied around without the information or context to help parents or their children understand what they are for, how the results will affect them, and why they are important to children as individuals.

Most parents have probably heard about the SATs test at the end of Key Stages 1 and 2, which are intended to highlight the performance of the school. Yet by contrast, CATS tests taken in many schools in Years 4 and 5 tend to have less fanfare but could give parents a good understanding of how well their child is performing.

A better understanding of what assessments are being sat, when, and the purpose and the value of these assessments, could help parents understand that many assessments are there to open doors and support opportunity, rather than pigeon-hole their children.

4.1 Challenge: Teachers, parents, governors, the government, employers, and young people, all need and want different things from assessment

Providing relevant and timely information on pupils' achievement for a range of different stakeholders is difficult because different groups want different things from assessment. Consequently as Professor Coe put it, there is unfortunately no "single simple thing" that will work for everyone:

"Parents want one thing. School leaders want one thing. Class teachers want one thing. The students themselves want something else."

Professor Rob Coe

Tests and exams at the end of Key Stages 2 and 4 give an example of how different groups have differing needs when it comes to assessment. These tests also show how misperceptions of assessments can lead to frustration.

SATs tests are intended to provide the government, Ofsted, school leaders, and governors with information about schools' performance, so that perceived underperformance can be addressed.^{xxxv} SATs are also required to calculate the Progress 8 performance measure at secondary school. However, many teachers and parents' responses to the online consultation and in focus groups suggested SATs tests should:

- Provide feedback that help pupils develop their learning, and;

- Assess a broader range of areas so as to celebrate pupils' wider achievements.^{xxxvi}

While the tests offer a 'macro level snapshot' of how schools and the system are performing in relation to the measures tested, Michael Tidd, headteacher at Medmerry Primary School, said that greater understanding is needed that this is all they are intended to (and consequently will) do. Different assessments – and, in all likelihood, more assessment – would be needed to fulfil any other functions.

Furthermore, expanding the function of SATs tests at KS2 could result in them falling into the same trap as GCSEs, which serve multiple conflicting functions. Pupils, colleges, universities, and employers rely on GCSE exams to grant qualifications. However, the government and Ofsted also rely on GCSE results to hold schools to account, and measure standards over time.



If I could wave a magic wand...

Being clearer about the purposes of different assessments: Daisy Christodoulou,
Director of Education, No More Marking

I would like to see more of a focus on the purposes of assessment. I'd like to see teachers, schools, Ofsted and the government give more attention to the purpose a particular assessment has been designed for, before attempting to draw any inferences from it.

At the school-level, many assessments are expected to provide both formative and summative information. That is, one assessment may be used to generate a grade that shows how a pupil is doing relative to their peer group, but it may also then be expected to provide diagnostic information about what a pupil has to do next to improve. But the kind of assessment that is ideal for these two purposes is very different. Summative assessments have to test large domains of content and be taken in standard conditions. For example, a task on a GCSE English paper might involve an essay on A Christmas Carol. But when reading such an essay, it can be quite hard to work out precisely what a pupil's weakness is, and precisely what they need to do next to improve. In order to provide a more helpful next step for the pupil, a smaller assessment might be better – a quiz on the events in A Christmas Carol, perhaps, or a writing activity where pupils edit the ambiguous pronouns in a paragraph.

And at the system level, things are just as bad – we routinely expect GCSEs to provide us with completely different types of information. For example, employers expect GCSEs to tell them if a school-leaver has the skills to do a certain job. Universities want GCSEs that tell them if pupils have the knowledge and skills to start a particular degree course. The government wants the results from the same examinations to tell us whether a school is succeeding. And many of us also want to use the results of GCSEs to tell us if standards are rising over time. These competing purposes pull in many different directions, and the perfect test for each individual purpose would not look like the current GCSE.



If I could wave a magic wand...

Generating more (and better) discussion about what makes a quality assessment:

Tim Oates CBE, Group Director of Assessment Research and Development, Cambridge Assessment

Some might immediately launch into 'wishes' about technology-enhanced formative and summative assessment – but these are coming anyway. Far more important, in my view, is the question of the quality of assessment and what we do with the outcomes of assessment. I remain pretty nonplussed as to whether 'consequential validity' does or does not encompass both of these – looking to the future, they both are vital concerns.

Education is being swamped with assessment-oriented technology – applications written in young educationalists' kitchens, by the big technology companies, by publishers and by assessment agencies. I am pessimistic about an 'invisible hand' of competition driving rapidly towards high quality assessment. 'High quality' assessment requires consideration of a diverse set of characteristics. Each assessment needs to measure well. It needs to encapsulate and promote the right standards; standards which drive both curriculum quality and meet the requirements for progression. Then there is utility (cost and manageability), comparability, security and so on. And there is too much evidence of 'invisible hands' in markets working poorly where users do not possess good knowledge regarding quality – information asymmetries – or where incentives and drivers are stacked up the wrong way.

Fingers point at accountability – 'that's the thing that is wrecking assessment...'. But it's naïve to think that accountability is going away. Government has a well-grounded duty to gather information on the quality of education. It needs to do this with probity and accuracy – but all successful systems exercise this duty, albeit using different models for so doing. And it's wrong to perpetuate myths such as 'Finland has no accountability'. Plenty of attention is given in Finland to assessment outcomes – and there is loads of testing in primary to make sure that children do not fall behind.

So here is my one desire for the future: that there is far more discussion about what makes good assessment – only then can all parties converge on quality.

4.1.2 Ways forward

Teachers must be clear on the purpose of their assessments, identifying what information is needed, and by whom. They should then select different assessments based on this

In order to meet different stakeholder's needs, the purpose of each assessment needs to be made clear. Importantly, schools should not see assessments or the data they produce as an end in themselves, but rather as a means of supporting better conclusions and more targeted action. Professor Wiliam suggests we should:

"Start from the decisions people need to make and then think about how we can then inform those decisions. ...Why give reports to parents at the end of the year when the school year's just finished? Give the parents information and be clear about what it is you expect people to do with this information."

Professor Dylan Wiliam

Ensuring assessment supports mastery, attainment and progress – Michaela Community School

Michaela Community School in Wembley, London, adopts a three-pronged approach to assessment. Deputy Headteacher, Katie Ashford, explained that while the school believes its approach will benefit pupils, the school has yet to receive GCSE results and so drawing firm conclusions from it would be premature.

1. The first element to assessment at the school is 'mastery', which focuses on content-specific understanding and knowledge. In essence, Katie explained, "does the teacher know whether or not the pupils have learnt what they are supposed to have learnt?" The design of these assessments varies depending on the content being taught, but will often involve multiple-choice questions and other quizzes. The timings of the assessments also varies. Some test what pupils learnt during that lesson, others, the previous lesson, and others recapping content taught earlier in the year. Recapping knowledge is especially important because "if [pupils] can't remember it, then it hasn't been learnt."

2. The school also undertakes assessments to identify pupils' attainment levels. These are separate to their assessments of progress, and involve testing whether pupils have understood content at a unit level. For example, while pupils may have learnt specific facts about a play, their teacher still needs to know whether they can

write a competent essay about it. Consequently, there are two assessment points each year (one in January, and one in June) to help teachers make these judgements. Katie explained that administering the tests is challenging, because it is difficult to pitch the tests at the right level and to ensure teachers' marking is consistent.

Teachers currently mark using rubrics, although the school would like to move away from these as rubrics can lead to less reliable judgements (the school is working with No More Marking – described in section 6.1.2, below – to increase the reliability of its writing assessments). Pupils receive marks following these tests, but because the tests get harder for each year group the school does not necessarily expect pupils' marks to increase (although this would be desirable). Training and moderation exercises take place before, during and after teachers mark the assessments to ensure judgements are in line with one another.

3. Michaela also assesses pupils' progress using standardised reading, English, and maths tests from GL Assessment. These assessments are conducted at the beginning and end of year 7, and the end of years 8 and 9. Using standardised assessments is important because it shows the school how pupils' progress compares to other pupils of the same age nationally. Katie said the information provided is more useful at the cohort level, showing how in general a year group is progressing. These assessments can also help triangulate the results of the bi-annual assessments, mentioned, above.



Putting pupils at the centre of the assessment process – Chailey Heritage School

Chailey Heritage School in Sussex is a non-maintained special school for pupils aged 3 to 19 with complex physical disabilities, high health needs, sensory impairments, and associated communication and learning difficulties. The purpose of assessment at the school is first and foremost to support the pupils as they develop educationally, socially and physically. This takes priority over providing accountability data, or comparing pupils' achievements with their peers in other settings.

The school previously felt that pressure to 'prove' what pupils had accomplished meant it had tried to use commercial assessment software packages. However the school came to feel that it was using assessment in ways that did not always prioritise pupils' development and that these packages did not provide teachers with information that related directly to the next steps pupils needed to make. Worse, such packages could exert undue influence over the curriculum if teachers felt pressured to 'tick off' certain objectives (even if these objectives were not developmentally relevant or appropriate for the pupil).

Instead, headteacher Simon Yates explained that every pupil at the school now has a series of profiles relating to different areas of his or her

development, including communication, social and emotional wellbeing, physical, and access technology (the technology the pupil uses in their day-to-day life). These profiles form the cornerstone of the school's curriculum, and each profile contains a series of bespoke targets for each pupil. A range of people feed into these (including teachers, speech and language therapists, occupational therapists, physiotherapists, and parents). The number of targets each child has varies depending on need and are revised as and when necessary.

The most important part of what teachers at the school do is observation, explained Simon and his colleague, class teacher Julie Tilbury. The school has experimented with ways of tracking pupils' progress, although Simon recalled "any kind of prescribed structure we've had before couldn't possibly work for everybody." The school produces two reports about each pupil annually, containing prose about what the child has achieved. Simon said this information does not translate into numbers, and that he is unclear what the purpose of producing numbers would be. Instead, Simon and his team focus on ensuring teaching and support are as good as possible: "then the progress made by that child, whatever that is, will be the very best that the child could have made."

Julie said pupils' parents are "astonished at how much detail there is, but also how accurately it captures their child."



Using verbal questioning to stretch pupils' knowledge and understanding – Parkwood Primary School

For a number of years, Parkwood Primary School in Keighley, West Yorkshire, has been focusing on developing the quality of teachers' and pupils' talk during lessons. Verbal questioning is one of the most efficient and effective ways for teachers to probe and extend pupils' learning, assistant headteacher Sacha Nelson explained, and so the school helps teachers do more of it. Natalie Gregory, a year 3 teacher, explained, "if you ask questions on the spot you get a great sense of where the kids are."

Teachers are expected to think about the questions they ask pupils while planning their lessons by using Chris Quigley's B.A.D. framework, and the Tower Hamlets Language Functions.

The B.A.D. framework suggests different types of question can support pupils develop 'basic', 'advancing' and 'deep' understanding of subject content:

- 'Basic' questions require pupils to, for example, 'list', 'describe', 'locate', 'define', or 'recall'.
- 'Advancing' questions ask pupils to, for example, 'modify', 'explain', 'predict', or 'compare'.
- 'Deep' questions ask pupils to, for instance, 'argue', 'evaluate', 'imagine', or 'design'.

Tower Hamlets 'Language Functions' include argument, comparison, deduction, hypothesis and sequencing. Teachers use the B.A.D. framework and language functions to help them plan lesson questions and activities.

Both Sacha and Natalie said teachers need excellent subject knowledge to use questioning effectively. Natalie said that "it's really hard to stretch pupils further if you don't know the whole, entire subject area."

Pupils at the school said they find answering questions helpful, both because it helps them develop their knowledge and understanding, and because it helps the teacher understand how they are doing.

4.2 Challenge: Assessments are too often focused on providing information for accountability and reporting purposes

Many practitioners feel assessments used for reporting and accountability purposes take up too much time. These assessments may include:

- Internal tests, intended to provide marks or grades that can highlight any concerns about pupil performance that are reported to middle or senior leaders, and governors. Results from such assessments are sometimes reported to parents and governors.
- External summative tests and exams, intended to provide schools and the government with information about pupils' achievements, and possible areas of underperformance.

Schools feel pressure to produce such data because of a perceived need to have 'up to date' assessment information about their pupils. However, assessments that serve a reporting or accountability function are less useful in providing detailed feedback to inform next steps in learning. Furthermore when asked what the primary function of assessment should be, respondents to our online consultation overwhelmingly said it should check pupils' learning in order to identify next steps for teaching.

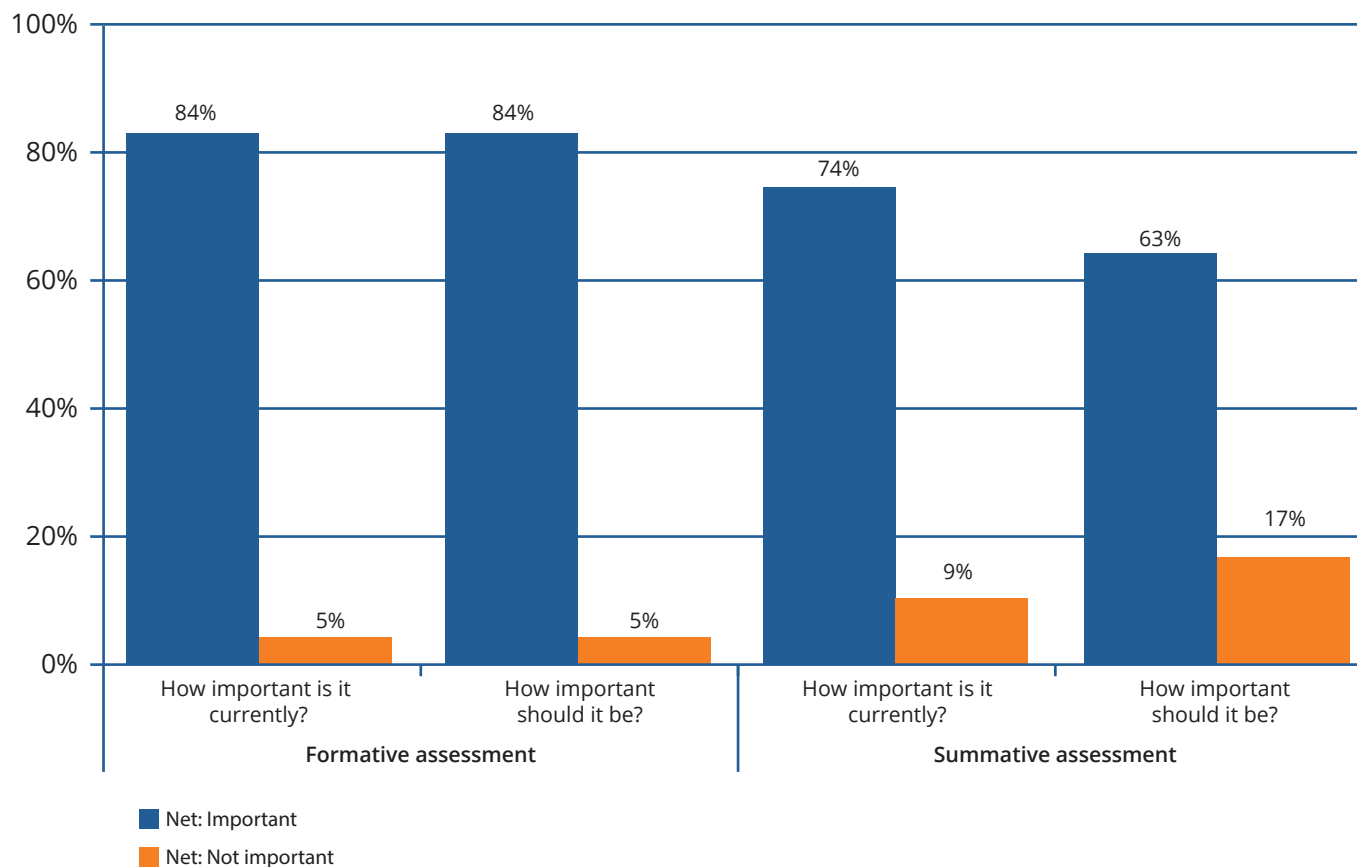
Teachers spend more time conducting summative assessments than they would like

As part of our national poll of teachers, respondents were asked how important formative and summative assessments currently are to the job of teaching, and how important these should be. The terms were defined as follows:^{xxxvii}

- **Formative assessments** include the day-to-day practices that help teachers and pupils understand what has and has not been learnt, and put in place actions to address this on an ongoing basis.
- **Summative assessments** provide a shared and consistent understanding of pupils' achievements.

While nearly three quarters of teachers reported that summative assessments currently form an important part of the job of teaching, less than two thirds felt this should be the case. 17% said summative assessments should not be important to the job of teaching.³

How important are - and should be - different forms of assessment to the job of teaching? (n=1,002)



³ Respondents were asked to rank their responses on a scale of 7 to 1, with 7 being 'very important' and 1 being 'not at all important'. Respondents were also given the option to say 'Don't know'.

Practitioners tended to give approximately equal weighting to the importance of formative assessment regardless of the phase or subject area they worked in. However, respondents in secondary schools have a significantly greater focus on summative assessment in their roles (with 79% saying it is important to their jobs) compared to their Early Years and primary colleagues (with 64% and 66% respectively saying summative assessment is important).

Differences in perspectives were also apparent between practitioners in different roles, with summative assessments seen as more important to teaching by middle leaders including heads of department and phases leaders than class teachers (with 79% middle leaders saying summative assessment is important to teaching, in comparison with 70% teachers).

Classroom teachers were significantly more likely than middle leaders to say summative assessment should **not** be important to their job (19% versus 13%). This is perhaps unsurprising, as middle leaders' roles often entail tracking pupils' achievements and progress in their phase or subject area. Worryingly, this may mean that middle leaders and classroom teachers are pulling in different directions, with classroom teachers spending more time than they would like undertaking summative assessments.

Schools produce lots of poor quality summative data, which can distract from – and reduce the quality of – diagnostic formative assessments

Schools' focus on summative assessment can distract teachers from other forms of assessment that would

help them understand pupils' learning better. This can happen for several reasons.

1. Pupils practise assessments in their final form rather than concentrating on mastering the component knowledge and skills

Many schools blur the line between formative and summative assessments. Christine Counsell, Director of Education at the Inspiration Trust, explained this is often due to the erroneous belief that formative assessments should relate "in a direct way to the ultimate summative assessment." Counsell argued this "strange cosmic marriage" often involves getting young people to "practise the thing in its final form", and that this represents a misunderstanding about how to best help children achieve the outcomes tested in summative assessments:^{xxxviii}

"The last thing we want to be doing with ordinary formative assessments ... is building an assessment around the final form, trying to replicate the GCSE. Because what the children have to do to succeed in that is everything from learning to conjugate their verbs to practise and become successful through drilling in various aspects of communication."

Christine Counsell

In turn, this dramatically reduces the efficacy of classroom assessment in showing pupils and teachers where improvements are needed.

Michael Fordham argues that giving pupils formative feedback on full form assessments such as essays can be beneficial, when pupils have the wider knowledge necessary to make sense of this feedback.^{xxxix} However, this sort of feedback could be confusing if pupils do not know enough about the topic area to make sense of it. In contrast, diagnostic formative

assessment seeks to identify and test the causes of underperformance; "these kinds of assessment will not necessarily look like the final performance: indeed, they might well look very different."

2. Schools prioritise generating tracking information, which can detract from understanding next steps in learning

Schools regularly collect data for 'tracking' purposes, that is, to monitor pupils' progress throughout the school year. The vast majority of schools in England do this at least termly.^{xl} Tracking is generally intended to provide information about pupils' broader achievements in a subject area at a given point in time.

Research indicates that teachers and school leaders feel anxious to 'prove' what their pupils have covered.^{xli} This can mean teachers spend less time undertaking formative, diagnostic assessments and more time collecting evidence to demonstrate pupils are making 'progress'. However, tracking can reduce the quality of information available to teachers and pupils about next steps in learning if:

- Teachers spend more time than they would like conducting summative assessments that serve little or no formative function;
- Schools try to use summative assessments formatively, even though these assessments are often too broad to provide the granularity of feedback necessary to provide helpful information about how and where pupils can improve;
- Schools track pupils' performance in formative assessments, even though data produced from formative assessments should "*duck and dive*" as pupils grapple with new content and

teachers evaluate how and where to focus next. Counsell explained that using such data to inform summative judgements about progress is inappropriate because the content tested by formative assessments should be very specific. Additionally, teachers might worry if, week-on-week, pupils are not getting better scores, even though they may be being taught different content;

- It supports “an obsession with grade boundaries,” meaning schools attempt to measure the proportion of pupils getting ‘good’ grades internally, based upon assessments that have results that are not reliable enough to draw such conclusions. Counsell suggested this in turn can lead to a narrowing of the curriculum and a focus on preparing for tests in a way that actually undermines pupils’ chances of making good progress;
- The reliability of these assessments is unknown, or low, or;
- It creates excessive and unmanageable workloads for teachers, which Counsell argued “is completely unnecessary.”

3. The curriculum can end up overly based on test and exam syllabi

Respondents expressed concern that end of Key Stage test or exam syllabi can disproportionately influence schools’ entire curricula.^{xiii} Ironically, broader curricula permit more rounded learning for pupils and are ultimately more effective in improving achievement.^{xiii} In particular, Counsell suggested, a broader curriculum can disproportionately benefit pupils from poorer backgrounds:

“The best way to make sure that children, particularly disadvantaged children, are going to succeed by the end of Year 6 in reading and writing is to give them a broad curriculum, it’s not to practise those tests.”

Christine Counsell

Parents said that while their children’s performance in English and maths is clearly important, they also want to know how their children are getting on across a far broader range of areas in school. Many expressed concern that, particularly at KS2, the curriculum is disproportionately shaped by the content of statutory tests.

Another reason this is problematic is that test and exam syllabi can become models of progression where, for example, the knowledge and skills associated with a grade ‘3’ are seen as a pre-requisite for moving on to grade 4. However, the GCSE grade descriptors 9 to 1 are designed to summatively capture learning and achievements at the end of a course of study, rather than provide a sequence for learning wherein schools cover the content for grade ‘4’ before moving onto grade ‘5’. This is a particular problem where schools hang their KS3 curriculum off KS4 requirements, since KS3 should focus on different content. Adopting GCSE based grades further down the school therefore raises two key questions:

- If they are adapted in order to be ‘age-appropriate’, how can schools be confident the grades provide a reliable measure against pupils of the same age nationally?
- If they are not transposed, is it fair or helpful to compare a year 8’s achievement on one set of topics with a year 11’s achievement on different topics?



If I could wave a magic wand...

Delineating between formative and summative assessment: Neil Carmichael, former Chair of the Education Select Committee

There are, essentially, two forms of school assessment; firstly, ‘summative assessment’, where test and examination results are aggregated for the purposes of measuring school performance. The risks associated with such league tables are unintended consequences where schools might, for instance, encourage pupils to pursue academic pathways not ideally suited to them.

This form of assessment as operated in England enables ‘informed’ parental choice between schools and is a tool to drive school improvement.

‘Formative assessment’ is the other form, and the purpose is about identifying and responding to the needs of the individual pupil. Often formative assessment across different education systems is informal, optional and interactive, but it is an integral part of the teaching and learning process.

In England the lines between summative and formative assessment systems are blurred. This causes confusion, and is also associated with the persistent concerns over ‘teaching to test’ and the stress levels of pupils being tested at several stages.

Assessment of individual pupil performance and of schools are two essential functions in any education system. In England, there is an urgent need to properly clarify the boundaries between summative and formative assessments. School leaders are naturally concerned about the summative assessment system but they should also understand the power of effective formative assessment in delivering impressive aggregate outcomes.

4.2.1 Ways forward

Schools must cut the time class teachers spend conducting summative assessments so that they can focus on diagnostic, formative assessments

Formative assessment may be easier where teachers do not have responsibility for summative assessment:

“An Olympic coach for the high jump has no influence over whether somebody clears the bar or not at the Olympic Games so all their work, all their coaching is focused on, ‘how can I make this athlete the best high jumper possible’, so all of their assessment is formative.”

Professor Dylan Wiliam

By ensuring that the purpose of classroom assessment is clear and, importantly, that it is separated from tracking of pupils’ longer term progress, teachers will be freer to focus on evaluating learning and identifying relevant next steps.

School leaders, and middle leaders in particular, should therefore ensure class teachers have the space and confidence to undertake regular diagnostic formative assessments rather than focusing on summative assessments.

Schools and teachers could do this by:

- Undertaking only a limited number of summative assessments in each subject, each year. The number of summative assessments should not exceed the number of topics pupils are studying. Subject and phase leaders should specify the timings of these assessments in advance so teachers know when to expect additional marking;

- Using summative assessments to make holistic judgements about pupils’ performance so as to provide marks or grades (rather than to provide formative feedback);
- Using mock tests or exams to compile a list of common mistakes or misconceptions and

address these in a follow-up lesson, rather than using them to provide formative feedback at the individual pupil level, and;

- Making greater use of standardised testing packages.

Standardising assessment judgements across schools – Ark Schools

Ark Schools is a charity and multi-academy trust containing 35 schools in England. Jasper Green is Head of Secondary Curriculum and Assessment. Ark designs curriculum and assessments for some subjects centrally. While the schools do not have to use these resources, Jasper explained the majority do, because they offer a rigorous framework that can guide learning and enable schools to compare their pupils’ progress with others across the network. This approach provides the reassurance of a ‘minimum standard’. It also means that good practice can be identified, something Jasper said “provides a route to sharing expertise.” As Ark develops the provision for other subjects, they are looking to share and build upon the curriculum and assessment expertise that already exists within its network.

In schools using Ark’s common curricula and assessments, pupils sit summative assessments at the end of each term. These take place in school, although significant care is taken to ensure the assessments are administered consistently, because the “conditions of the assessment are key. There’s not enough acknowledgement that how you sit an assessment and how you mark it makes a massive difference to the marks

you’re comparing.” Teachers administer and mark the assessments, perform moderation and standardisation, before uploading the results to a central database. The assessment conditions are vital, therefore, because “if the data from the beginning is not good, the whole thing is a waste of time.” Once the results are uploaded, the central team create age-related numbered GCSE grade boundaries. To support the consistency of judgements both centrally and in schools, schools administer standardised GL Assessment tests at specific points during KS3.

To support the consistency of their marking and feedback, school leaders and teachers attend moderation sessions and network days three times a year. These provide an opportunity for teachers to use centrally compiled portfolios of pupils’ work exemplifying different grade standards to rank their own pupils’ work. Double marking also takes place for some subjects. This helps prepare teachers and subject leaders for the next round of assessments and provides an opportunity to explore modifications or interventions that might support learning with their classes through re-teach back at school. For Jasper, these moderation and network sessions are critical, because “accurate data does nothing, unless it leads to some action. It’s the action that’s important.”

Schools should use standardised tests to benchmark and report pupils' achievements

A number of standardised tests exist,^{xliv} which help schools compare pupils' performance with other children of the same age across the country. According to the NFER's teacher omnibus survey, a greater proportion of respondents from secondary schools use standardised tests (62%) than primary school respondents (22%).^{xlv}

Schools should make greater use of standardised testing packages to assess, benchmark and report on pupil achievement in specific knowledge and skills areas. Commercial packages are available offering standardised scores in reading and maths. Using such packages can:

- Provide schools with an indication of how their pupils compare with their peers nationally;
- Reduce the burden for teachers and middle leaders in some subjects to regularly create their own summative assessments, and;
- Reduce teaching to the test, as many standardised packages randomise the questions they ask.

Increased use of standardised assessments by schools may also address consultation respondents' concerns about the lack of common language for describing pupil achievements, particularly at primary level.

Using standardised tests to enhance within-school assessment – Plumcroft Primary School

Teachers at Plumcroft Primary School in Greenwich, London, wanted assessments to help them:

- Make more reliable judgements about pupils' performance, so as to;
- Ensure the activities taking place in classrooms and across the school are as efficient as possible.

The school uses Rising Stars 'Star Reading' and 'Star Maths' standardised tests to achieve this. These assessments serve as an "honesty check", headteacher Richard Slade explained, complementing the other sources of information gathered by teachers, and ensuring the senior team do not need to rely exclusively on Ofsted inspectors or external performance data for validation that pupils are moving in the right direction.

Pupils sit the assessments each half term. The tests are administered during a two-week window prior to the half term, using iPads. The children take the assessments in silence, and in the company of either the class teacher or a teaching

assistant. Richard explained, though, that because the questions are computer adaptive (meaning questions get harder as they are answered correctly, or easier if answered incorrectly), the risk of cheating is minimised because pupils are unlikely to have the same questions on screen at any given point. The senior leadership team conduct random moderating visits to ensure the conditions under which pupils in different classes are sitting the assessments are suitable.

Results are presented online, and used to identify pupils in need of additional support during lessons (as well as pupils who may need more specialist support). The information also helps Richard and other line managers monitor different classes' performance. Teachers' performance management does not hinge on the results of the tests, although the test data provides a useful discussion point.

Teachers said they like using the tests, because the results feel reliable and fair. It is not possible to predict what questions will come up, so teachers feel incentivised to teach broadly. Dave Witham is the school's Upper KS2 Phase Leader and a class teacher, and said "from our point of view, it's a completely independent summative assessment."

Using pupils' feedback to guide curriculum content and teaching – Japan

The Japanese word for assessment is 'Hyōka', although the word has two simultaneous meanings:

- The first is 'evaluation', making absolute judgements about ability and competence.
- The second takes the meaning of assessment more broadly, as something formative that helps teachers and students learn something about themselves.

Mr Ohno is a teacher of Biology at Kunitachi High School. He feels most teachers in Japan take 'assessment' to mean evaluation: "I think this is a very big problem, [because] they don't think about how to stimulate and [help students] grow." Furthermore the emphasis of classroom assessment becomes about making summative judgements and providing grades, rather than informing next steps in teaching and learning.

Consequently, Mr Ohno has developed a system whereby his students complete reflection sheets at the end of every lesson. The reflection sheets ask pupils:

- To record the date, and content covered;
- For three words that students associate with the content;
- What they did not understand during the lesson;
- To think of a new question about the topic they covered that lesson;
- For something they found particularly interesting and important, and;
- About the quality of the lesson, giving it an A (good), B, or C (bad).

These reflections can take up to 10 minutes to complete. Afterwards, Mr Ohno reviews all the students' responses

and, on the same sheet, gives them a grade, evaluating the quality of their reflections. However, he feels "the most important thing is [to] motivate the students."

Initially it took him around an hour to mark each class's responses, although it now only takes around 15 minutes to cover 40 reflections. He then addresses common misconceptions in the next lesson, and answers some of the students' questions. (Students in Mr Ohno's class, and many of his colleagues' classes, sit regular mini-quizzes to test their understanding of academic content).

Motivated by a similar desire to use assessments to better inform the curriculum and classroom teaching, Ohyu Gakuen Girls High School in Setagaya-ku, Tokyo, conducted a large, one-off evaluation in the late 2000s. This used data from a wide range of sources, including:

- Psychometric tests, evaluating pupils' levels of self-confidence and stress;
- Evaluation forms, completed by both pupils and teachers, reviewing the ways in which pupils contributed during lessons (such as the quantity of questions asked, and the completion rates for homework), and;
- Academic performance, according to standardised testing.

The school's principal, Mr Yoshino, explained that the school collated this data, examining both how trends in pupils' responses changed over time, and also what 'average' responses in each year group looked like. The school found that the younger pupils needed far more support to develop self-confidence and resilience, and modified curriculum content and teaching strategies with these age groups accordingly. The teachers of these age-groups were encouraged to develop more 'active' styles of teaching, getting the girls to work in groups and to take greater responsibility for their own learning.



4.3 Challenge: Communication about assessment between teachers, and between teachers, governors and parents, is too limited

There is an urgent need to improve communication around assessment, and in particular between practitioners, and between practitioners and parents.

Assessments often do not provide parents with meaningful information

Nearly half of parents do not feel statutory tests are a fair measurement of their children's achievements.^{xlvi} This shows that the focus of these tests and methods of assessment are too narrow, or that the tests are misunderstood.

The removal of National Curriculum levels has resulted in a lack of a common language to describe pupils' achievements and progress according to many teachers, parents and governors. However, many also acknowledged that even whilst levels existed, they were not always used in a way that provided useful information about pupils' achievements.

Parents can also feel that the new attainment bands 'working towards', 'meeting' or 'exceeding' age-related expectations are too broad to be helpful at primary. Meanwhile at secondary level, the transition

to new numbered GCSE grades has been confusing for many. This is partly because parents are unsure how the new grades compare with the old. Some parents want additional information about where children sit relative to their peers and what they and the school can do to help their child improve. This concern was raised during focus groups and in the online consultation in relation to the Early Years, and tests at Key Stages 2 and 4.

Special school practitioners said it was particularly important to consider the quality of communication in relation to young people with forms of SEND. It can be challenging for schools to accurately communicate profiles of need (including academic performance) about pupils with SEND, in ways that are not open to misinterpretation, said Simon Knight, Director of Whole School SEND, part of the London Leadership Strategy. Communication about these pupils' need is often not specific enough.

No assessment is perfectly reliable and not enough people realise this

Statutory assessments such as those sat by young people at the end of Key Stages 2 and 4 need to provide a shared understanding of the general level a pupil is working at, across large subject areas. To assess everything a pupil knows could take days, and since time and money are both in short supply, tests 'sample' pupils' knowledge, testing certain areas of knowledge and skill in order to make inferences about their broader abilities.

However, tests can be unreliable because of:^{xlvii}

- **Sampling unreliability**, caused by the type or difficulty of the questions;
- **Inconsistencies in students' performance**, which can vary day-to-day, and depending on the test conditions, and;
- **Marker unreliability**, caused by the difficulty in ensuring assessors give marks in the same way.

The need for reliability introduces trade-offs

Assessing someone for longer, or assessing a narrower range of topics, can increase the reliability of a test. KS2 tests assess pupils more narrowly in certain aspects of English and maths, whereas assessments in the Early Years, KS4 and KS5 tend to be broader in scope (covering more of the curriculum). This has consequences for how teachers spend their time, since knowing that particular topics or subjects are more likely to be tested can result in 'teaching to the test'.⁴ This challenge is well documented in the literature,^{xlviii} and Ofsted flagged this concern in its primary and secondary curriculum research.^{xlix}

Teaching to the test has an undesirable impact on both pupils and teachers. Young people told us how learning could become formulaic and dull when the focus is narrowed to passing tests. Teachers complain it undermines their professionalism, is stressful, and diminishes their love for their subjects and for teaching itself.

⁴There is nothing wrong with this, suggested William and Oates during their interviews, if teachers teach to the right tests, and everyone agrees on the purpose of these tests. Such tests will sample randomly across the entire taught curriculum, so it is not possible to predict what will come up.

Yet trade-offs are inevitable since testing every young people on a wider range of content would mean more time and money spent assessing. Becky Allen, Director of Education Datalab, explained:

"If you want to have a system where you collect some information on every single child and you want to cover a very, very wide range of parts of the curriculum then you are going to be testing children for a long length of time. And arguably that is not ideal."

Becky Allen

Statutory tests are not perfectly reliable

An assessment's reliability is often reported using a reliability coefficient between 0 and 1, with '0' denoting a score that is nothing but measurement error (and therefore totally unreliable), and a '1' indicating no measurement error whatsoever (and complete reliability). Daniel Koretz (Professor of Education at Harvard University) explains that "in large-scale assessment programmes, the most reliable tests have reliability coefficients of .90 or a bit higher."ⁱ

However, Koretz suggested that even in tests with a reliability of .9 ("close to a best-case scenario"), if anywhere between 30% and 70% of students pass, between 12% and 14% would receive different grades were they tested a second time because of fluctuations in the pupils' performance or in how their responses are marked. Professor Wiliam has written and spoken extensively about the reliability of assessments.ⁱⁱ Assuming assessments for the previous KS2 levels system were reliable at .9 (which is optimistic) Wiliam suggests that 23% of pupils would be reclassified were they to retake the assessments.



4.3.1 Ways forward

Everyone – but especially teachers – should know how reliable assessments are

Teachers, young people, governors, parents, ministers, universities and employers need to embrace the idea that assessments are not perfectly reliable. Far from spending longer testing young people (with one estimate suggesting twice as much assessment would be needed per test to increase a reliability score of .75 to .85),ⁱⁱⁱ a greater understanding of reliability could lead to:

- More accurate reporting of pupils' achievements by teachers and schools, leading to;
- More informed and appropriate decision-making based on the results of these assessments, and;
- A more honest and informed debate between educationalists, and discussion about alternative ways of conducting assessment.

Teachers and schools must understand how reliable their assessments are, so their decisions based on assessments can take this into account. Wiliam said "if you haven't got any handle on the size of the error, you can't use those assessments to make smart decisions." Teachers therefore need to be able to calculate the reliability of their assessments, or access resources that can do this for them. For example, the Assessment Academy, a training programme from Evidence Based Education, supports its participants in accessing and using an assessment reliability calculator.



If I could wave a magic wand...

Calculating and communicating assessment reliability: Dylan Wiliam, Emeritus Professor of Educational Assessment, UCL IOE

Of course I want to change many things about assessment in English schools, but if I had to pick just one, it would be this—no assessment result of any kind should be reported to students, parents, or anyone else, without some indication of how accurate it is.

No assessment is perfectly reliable. With a slightly different selection of questions, with the student completing the assessment on a different day, with a different marker, or even the same marker at different time of the day, the result would be slightly different. And if you have no idea how different it might be, you have no business placing any weight on the result.

So, rather than telling a student that she has scored 65 on a test, we might say, for example, that the student scored 65, plus or minus 10. If we regard a score of 70 as a passing score for a particular course, then when the student asks "Did I pass?", we could say no more than "Probably not, although you might have done." If a parent then asks, "Why don't you know?" the teacher would say, "Because no assessment is perfectly reliable." At this point the parent might well say, "Well, why can't you make the assessment more reliable?" to which the teacher would say, "Well of course we could make the assessment more reliable, but that would mean making it longer, and we think we have better things to do with that time, like teaching your child."

In other words, we need all stakeholders in education to not just reluctantly accept but actually embrace the unreliability of assessments. Stakeholders need to realise that a certain amount of uncertainty is not unfortunate; it's optimal. My hunch is that, once people are confronted by the margins of error typical in educational assessments, they might start taking other sources of evidence into account, and make more sensible decisions about how much weight to place on any one assessment.

Individual teachers should know about the reliability of their assessments, but not be expected to communicate this to parents (unless the assessment has especially high consequences for the pupil) as doing so would be onerous. Instead, schools should communicate how assessments inform decisions about pupils to parents. This could be done as part of:

- One-off meetings for new and existing parents;
- Parents' evenings;
- Ongoing dialogue between class teachers and parents, and;
- Communications through school websites.

The media and the government also have an important role to play in ensuring they communicate not only the results of statutory tests and assessments, but information about the purpose and reliability of assessments.

Assessment providers should provide easily accessible and digestible information about the reliability of their assessments

While increased use of standardised assessments by schools would allow more reliable comparisons between pupils' performance and others nationally, doing so could lead to misplaced complacency as the reliability of these tests is by no means perfect.

"[Standardised tests] are always going to be noisy measures of how good children are at things at a particular point in time. ... These tests are really noisy and you should expect quite a mess in your data at the level of the individual child."

Becky Allen

Many assessment organisations provide detailed information about their assessments' reliability. However, this is not always readily available to teachers, parents and pupils. Assessment organisations should therefore ensure users of their assessments know how to access and interpret this information.

Setting targets for pupils that go beyond academic objectives – Frank Wise School

Frank Wise School in Banbury, Oxfordshire, serves pupils aged between 2 and 19 with severe learning difficulties and profound and multiple learning difficulties. Class teams and parents collaborate to set targets, and teachers then track pupils' progress against a wide range of objectives that encompass their academic, social and emotional, communication, and physical development.

These targets are seen as overarching objectives, and inform each child's curriculum and assessment arrangements for the remainder of that year, explained deputy headteacher Heidi Dennison. The targets may also specify the circumstances in which the children may complete the target (for example, whether at a specific location within class, elsewhere within school, or even offsite).

The school tracks pupils' progress against their targets throughout the year, and once pupils achieve them they will continue to work on them in generalised contexts (rather

than these necessarily being the specific focus of lesson activities, although they could well be). Each time an assessment addressing a pupil's specific target is carried out, teachers will note the level of support given, indicating whether physical, gestural or verbal support was needed or whether the pupil could complete a task unaided. The school generally feels a pupil is confidently able to perform a task when he or she does so unaided three times.

At the end of the year, data is collected on the extent to which children met their targets. Currently this data is binary – children either achieved their target, or they did not – although the school is exploring ways of showing in data terms partial success. Each individual target is then written up into an annual review report. Class teachers meet with subject leaders as part of this, in order to set targets for the next year.

The school expects pupils to meet approximately 80% of their targets a year. More than this and the targets may not be challenging enough; less than this, and the targets are probably too challenging.

Communicating pupils' achievements in the round – Japan

Students do very well academically in Japan. In terms of overall performance the country came second only to Singapore in its 2015 OECD PISA results^{liii}. Many teachers feel, though, that while students' academic scores are impressive, the school system as a whole is imbalanced, and too focused on the high-stakes university entrance exams that students take when they are 18.^{liv}

The content of the entrance tests depends on which university a student applies to, although many – particularly for the most academically competitive universities – focus on students' proficiency in language, science, and maths. This has 'washback' into the secondary school curriculum, and while there are national guidelines about what content should be taught and how, it is the entrance exams that direct teachers' behaviour. Teachers and school principals said that much of the work they do at the upper secondary level is orientated entirely towards the entrance exams. This is less the case for teachers in elementary and lower secondary schools, although they feel immense responsibility to give students a solid academic foundation to support success in the exams, but also get into the best possible school further up the chain (many lower and all upper secondary schools are selective).

Concerned that students worked too hard and lacked independence, the Japanese government introduced reforms in the 1990s and early 2000s,

cutting back the curriculum and introducing school-based 'Integrated Study Time' giving students space to pursue their own interests. These reforms had a mixed reception amongst teachers and students who in many cases felt the reforms had minimal impact on the overarching pressure to do well in entrance exams. The reforms have since been scaled back.^{lv} Universities are acutely aware of the challenge, although staff at Hosei University in Tokyo said that academic tests – for the time being – remain the most effective means of measuring students' potential (both academically, but also socially).

To counteract this imbalance between academic knowledge and 'everything else', the government has published guidance for teachers in assessing students' 'soft skills'. At Ryogoku Secondary School in Sumida, Tokyo, for example, teachers evaluate the pupils' soft skills as they progress through the lower school using the national guidelines. The skills assessed vary in different areas of the curriculum, although teachers generally take into account pupils' willingness to study, curiosity, and presentation skills. Each term, lower secondary school teachers award pupils an 'A' (excellent), 'B' (average), or 'C' (below average) grade in these skills. These grades matter, because they are combined with pupils' academic results at the end of lower secondary school to produce a '5' to '1' grade, indicating how successful the pupil has been in all areas of school life. These final grades affect which upper secondary school they can get into.

The school's vice principal, Mr Kobayashi, and his

colleague, class teacher Ms Oki, explained that while teachers nod towards these soft skills in upper secondary, the emphasis shifts back onto pupils' academic progression (only in rare circumstances will universities take into account pupils' scores in these areas). They acknowledged that moderating teachers' judgements about soft skills could be challenging. Subject teachers come together to review pupils' achievements before the grades are awarded, and a grade will not be accepted unless the teachers agree on it. However, because the government guidance only specifies what a 'B' looks like (with schools creating their own criteria for grades 'A' and 'C'), it is difficult to say how consistent judgements across schools are. Furthermore, while Ryogoku is both a lower and upper secondary school (meaning pupils generally pass straight through), there is an incentive for teachers in other lower secondary schools to inflate pupils' soft skills grades to give them a better chance of getting into stronger upper secondary settings.

Approaches to 'soft skills' vary considerably between schools. Although teachers at Senzoku Gakuen Elementary School in Kawasaki believe developing children 'in the round' is critically important, they place less emphasis on assessing such skills. This is because staff feel pupils generally cannot develop soft skills in the absence of academic knowledge. Consequently, the school's approach is not to quantify pupils' soft skills, but to teach ethics, classes that encourage pupils to think about moral and societal questions from different angles to build their empathy and morality.

5. How can the accountability system change to enable teachers to shift emphasis back on to assessment that supports learning?

What are the problems and challenges?	What are the ways forward?
<p>The accountability system often has a damaging influence on teachers' day-to-day assessment practices, because:</p> <ul style="list-style-type: none"> • Statutory tests and exams can dictate what and how pupils learn. • Teachers face undue strain during accountability, curriculum, and assessment reforms. 	<p>The Department for Education should develop a system of matrix sampling for assessing more of the National Curriculum (while minimising the number and length of tests pupils need to sit).</p> <p>Teachers must be given a sensible timeframe in which to implement curriculum and assessment reforms.</p>
<p>The evidence is not clear on how assessment can best support school- and system-level performance.</p>	<p>The Department for Education should build experimentation and evaluation into assessment and accountability reform, to better understand how its reforms impact upon standards in schools.</p> <p>Ofsted should enhance assessment training for its inspectors.</p> <p>The Department for Education should present schools' performance data as three-year rolling averages.</p>
<p>Teachers and parents have misperceptions about the purposes of statutory assessments.</p>	<p>The Department for Education and other organisations (including schools) must ensure teachers can access high quality assessment training and resources.</p> <p>The Department for Education must work with stakeholders including teachers' unions to 'myth-bust' statutory assessments.</p>



5.1 Challenge: The accountability system often has a damaging influence on teachers' day-to-day assessment practices

Statutory tests and exams can dictate what and how pupils learn

The accountability system often negatively influences teachers' day-to-day classroom assessment, and this was a major concern expressed by teachers, governors, parents, and young people alike. Many felt the content of statutory tests and assessments narrow the curriculum and result in teaching to the test (for the reasons explored in section 4.3, above), therefore shaping what pupils learn. Furthermore this can shape how they learn, with teachers often setting full form mock tests and exams rather than developing pupils' grasp of more specific knowledge and skills.

Consequently, assessment and accountability can work at cross-purposes, subjecting teachers to different and often competing pressures. Many feel they must prepare their pupils for tests and exams in ways that do not feel optimal for learning (such as teaching to the test). Therefore assessment can, as one online consultation respondent put it, be "misaligned with the objective of learning."

The ramifications of this are explored in section 4.2, and include:

- Pupils practising 'final form' assessments that resemble statutory tests and exams rather than mastering the component knowledge and skills;

- Schools prioritising the generation of tracking information, which can divert teachers' attention from identifying how to support pupils' next steps in learning, and;
- Text and exam syllabi becoming the curriculum.

Teachers face undue strain during accountability, curriculum, and assessment reforms

Teachers and schools have not had enough time or support during recent periods of accountability, curriculum, and assessment reform, and many felt they were expected to just 'get on with it.' The most important examples given in responses to our online consultation and during focus groups were the removal of levels in 2014 with no suggestions about what schools could use in their place, and reforms to statutory tests and exams at Key Stages 2 and 4, which often felt 'rushed'.

The lack of time and support over the course of these reforms had a damaging impact on classroom assessment because:

- Teachers felt unsure about what and how they were 'meant' to conduct assessment. This was especially a concern among primary teachers.
- Teachers felt less confident in their subject knowledge in light of the increased expectations set by the new National Curriculum, affecting their ability to assess pupils during lessons.
- Teachers' energies were diverted into planning schemes of work and lesson sequences to accommodate the new curriculum, taking their attention away from the 'here and now' of day-to-day classroom teaching and assessments.

- School leaders and governors were unsure how their settings would be affected by new accountability measures, and sometimes attempted to second-guess how to improve their performance under such measures (for example by trying to predict Attainment and Progress 8 despite the inherent challenges associated with this, with many then using classroom assessment to track pupils' progress towards these targets).



5.1.2 Ways forward

The Department for Education should develop a system of matrix sampling for assessing more of the National Curriculum (while minimising the number and length of tests pupils need to sit)

The Department for Education should use matrix sampling to assess more of the National Curriculum while minimising the number and length of tests pupils need to sit. Matrix sampling is an approach to measurement in which not every pupil is required to sit the same test or exam, or required to sit a test at all. Such an approach offers real potential to monitor standards over time, either at a school, regional, or national level, while reducing the pressure on schools to narrow their curricula. It could involve:

- Every pupil in a cohort taking a test, or just a selection of pupils;
- Different pupils taking different tests in the same subject, to assess a subject area more broadly;
- Different pupils taking tests in different subjects, assessing the curriculum more broadly, or;
- Different styles of assessment, with some completing written assignments while others undertake group tasks.

Such a system could help highlight areas of the curriculum in which teachers need additional support. It should be used instead of KS2 SATs tests, and alongside GCSEs (in the process ensuring GCSEs can perform the function of providing pupils with qualifications rather than judging schools' performance).

Building on the approach already underway with the National Reference Tests,⁵ some pupils in KS4 would



If I could wave a magic wand...

Measuring what we value, not valuing what we measure: Emma Knights, Chief Executive, National Governance Association

It would be to ensure that we measure what we value rather than simply valuing what we measure. Governing boards need to strike a balance between the focus on academic outcomes and the wider aims of education.

Assessment has a different purpose for each stakeholder in a school. Governing boards need to have access to high quality and comparative performance data in order to hold schools to account for the standard of education they provide. Identifying where particular groups of pupils, such as those eligible for the pupil premium, are underachieving should prompt discussion about any barriers and how these are being overcome. It is crucial that data is looked at in context – those governing have to be careful not to jump to conclusions, particularly when dealing with very small cohorts.

The accountability system's focus on pupil progress and attainment across a limited range of subjects can mean that less 'academic' subjects and pupils' personal development are side-lined. Part of governing boards' roles in setting the ethos, vision and strategy of their schools is to make sure this does not happen. A good vision describes what pupils will have left the school having achieved in terms of attainment, progress and being prepared for the next stage of their education and life beyond school.

But it is as important that governing boards consider the culture around assessment in their schools: pupils and staff should not be experiencing excessive levels of stress. While the accountability system set by central government can contribute to anxiety about outcomes, there is much that can be done in school to promote healthy attitudes towards assessment. This can sometimes require governors and trustees stepping up to have difficult conversations with school leaders in the interests of pupils and staff.

⁵ Following a first national test in February and March 2017, Ofqual announced in September 2017 that it would run annual National Reference Tests – one in English, and one in maths – to monitor changes in standards over time at GCSE. No individual student- or school-level results will be published, and from 2019 exam boards may refer to its results when awarding GCSEs. However, this system will not fundamentally change how assessments inform evaluations of school performance, and cannot therefore be expected to negate any of the perverse incentives resulting from the current system.

be required to sit a sampling test in addition to their GCSEs. However, by making the purpose of these tests clear (that is, they are intended to evaluate the school, and have no bearing on pupils' final grades), and ensuring they take place outside of the main exam season, any additional stress for pupils could be kept strictly to a minimum.

A matrix sampling approach should be introduced alongside evaluations of how accountability can best support school- and system-level improvement, outlined in section 5.2.2, below. The Department for Education in consultation with teaching organisations and assessment experts should decide:

- Whether this system will be used to evaluate school-, regional-, or national-level performance, and in what areas of the curriculum, and;
- Whether the system will be rolled out nationally, or whether it will be used differently in different regions.

Other organisations have also called for the introduction of such an approach, which could be structured in a number of ways:^{lvii}

- **A rolling programme of tests could measure standards in English one year, maths the next, and then science, before restarting.** A statistically representative sample of students (at least several thousand) would be assessed. The focus of the assessments would not be announced in advance, eliminating the incentive to teach to the test. Harry Torrance (Professor, and Director of the Education and Social Research Institute at Manchester Metropolitan University), who favours such an approach, said it would give system-level information about pupils' performance in the subject areas under scrutiny

but miss out on school-level granularity, although he believes this "would be a price worth paying" given the distortive effect statutory tests can have on schools' practice at present.^{lviii}

- **Standards could be evaluated at the system-level, and schools could then compare their own performance against this data.** The NAHT have previously suggested this, adding results should not be used to hold individual schools to account.^{lix}
- **Individual schools could still be held to account on their performance, depending on how the sampling is conducted.** IPPR and Cambridge Assessment argue this would mean schools still have a strong incentive to achieve good results. If tests are intended to measure school- or overall system-level performance, not every pupil need sit the same tests.^{lx}
- **Pupils could be randomly allocated test questions and different styles of assessment.** Professor Wiliam proposed that different pupils could be asked to complete written assessments, a collaborative speaking and listening assignment, or prepare individual or group presentations. This would help not only broaden the amount of content tested, but also incentivise teachers to develop pupils' broader skills sets. This idea is presented in more detail in section 5.2.2.

An additional benefit of sampling would be that it could allow a wider set of objectives and outcomes to be evaluated, beyond the purely academic. For example:

"One could have a test combined with inspection results about behaviour and mental wellbeing. If all schools were striving to improve the mental wellbeing of children, a subsequent survey which adopted the right sampling frame would detect that change."

Tim Oates CBE

Teachers must be given a sensible timeframe in which to implement curriculum and assessment reforms

Reforms to accountability, the curriculum, and to assessment will result in some additional pressures on teachers, governors, parents, and young people. However, the Department for Education could offer far more support to these groups and especially teachers during periods of reform by:

- Consulting widely before reforms are made, to help identify reforms' potentially negative effects and how these might be mitigated.
- Publishing non-statutory guidance with information that helps schools navigate the reforms and implement changes. It is clear for example that following recent reforms, more guidance on how to assess without levels was needed.
- Ensuring practitioners receive access (in the ways outlined in section 3) to relevant and timely training and resources.
- Giving schools enough time to implement the reforms. These 'implementation windows' should be specified by the government in advance, following consultation with teaching organisations including unions and the Chartered College.

Supporting schools during periods of curriculum and assessment reform – Japan and Ontario, Canada

Curriculum and assessment in Japan and Ontario, Canada, are both undergoing reform. Practitioners in both countries have received considerable time and support while the reforms have been implemented.

In Japan, the government has given schools several years to implement curriculum changes (2016 reforms will not be fully implemented in schools until 2023). Furthermore, it provides freely-available training for teachers, in order to update them on the reforms and talk through how changes will affect schools. Principal Yoshida at Senzoku Gakuen Elementary School, Kawasaki explained, though, that teachers are often too busy to attend these sessions. Teachers also rely on textbooks to help them navigate the reforms. Some teachers feel this undermines their professionalism, although many accept it is a ‘necessary evil’.

Curriculum and assessment reform in Ontario is a “constant, iterative” process, with changes fed back into the curriculum in an ongoing cycle. Jenn Clark, a school administrator, explained that schools and teachers get lots of support. Specifically:

- School boards fund instructional leaders to work with principals, helping them update their school improvement strategies. During a recent round of reforms, Jenn said the school board provided one instructional leader in every two schools.
- Schools receive funding so that classroom practitioners can have relief time to update lesson plans and learning goals.
- School boards publish a series of ‘think papers’ to elucidate the rationale and implications of the reforms. Jenn said these are not diktats, but contain ideas to support practitioners.

Evidence on how accountability works is unclear

An improved understanding of how assessments can best inform school- and system-level accountability is needed, and should be linked to a better understanding of how interventions can address school underperformance.

Many people take it for granted that the process of school inspection (which relies on performance data) helps schools improve. The OECD’s 2015 PISA study found that pupils tended to score more highly

in science in systems where principals have greater autonomy over resources and the curriculum, “but especially so in countries where achievement data are tracked over time or posted publicly.”^{lxix}

However, while accountability is a feature of many educational systems around the world,^{lxxii} it can take many different forms.^{lxxiii} Furthermore, in their review of the evidence, Coe and Sahlgren find limited evidence to suggest that high stakes accountability results in improved academic or longer-term life outcomes for young people. They therefore conclude that “we know little about how high-stakes assessments should be designed to optimise outcomes [or about] how to produce the optimal accountability structure.”^{lxxiv} In their analysis of England’s accountability system, Education Datalab found limited links between Ofsted judgements, and schools’ subsequent trajectories.^{lxxv}

Statutory assessments’ role in informing accountability is problematic

Teachers and parents were often concerned that the assessments used to inform school accountability do not provide a fair measure of school performance. There were five main issues:

1. **Assessments are not perfectly reliable**, and should therefore not provide a sole measure of schools’ performance. Issues related to reliability were explored in greater detail in section 4.3, and the challenges of basing judgements about schools’ performance on potentially volatile data are well documented.^{lxxvi}
2. **Interpretation of school data to inform judgements can be unreliable**. Use of performance data is prone to human error and

5.2 Challenge: The evidence is unclear on how assessments can best support evaluations of school- and system-level performance

There was widespread consensus that the ways in which assessments inform evaluations of school performance could be improved, and there were two main reasons for this.

bias. School leaders can misinterpret particular results,^{lxvii} and inspectors' views of a dataset can bias their view of a school.^{lxviii} Furthermore, different inspectors can draw different conclusions from the same data.^{lxix}

3. **The validity of performance metrics is questionable.** Pupils' performance on specific assessments, such as the KS2 reading test, can rise while their performance on equivalent measures (such as commercially-available standardised reading tests) does not.^{lxx} This draws into question the ability of current performance measures to capture what they claim to.

4. **Teacher assessment is not valued highly enough.** Many focus groups participants, particularly teachers, felt statutory assessments and thus the accountability regime place too little emphasis on teachers' professional judgement. Previous research has suggested that statutory assessments can be perceived as a means of controlling teachers' behaviour in schools.^{lxxi} This was a widely held concern in our consultation. Professor Torrance said that teachers often feel assessment "is done to them, rather than anything they can own and attempt to improve."

5. **Teacher assessment should not contribute to statutory assessment results.** Other contributors stressed that while teacher assessment is critically important, it should not be given greater prominence in statutory measures. The two main reasons given for this were:

a An **objection in principle**, because using teacher assessment to inform school-level accountability measures represents a conflict of interest.^{lxxii}

b An **objection in practice**, because teachers' judgements are subject to bias,^{lxxiii} and a lack of reliability (often resulting from different interpretations of marking criteria).^{lxxiv} It is not only in essay subjects such as English, though, where teacher assessment produces unreliable results, and "even specialist mathematics and science teachers in the secondary schools cannot be relied on to generate acceptably valid assessments of

their pupils' achievement."^{lxxv} Increasing reliability of teacher assessments is possible, but expensive and time-consuming.^{lxxvi} Furthermore, in systems where teachers have greater responsibility for their students' grades (such as in the United States, or in Queensland, Australia), they spend insufficient time on assessment for learning because they are more worried about the adequacy of their summative assessments.^{lxxvii}



Emphasising teachers' professional judgements – Finland

Finnish teachers are commonly perceived to have greater autonomy than colleagues in England, in part because there is no school inspection or prescriptive central curriculum. Certainly, Finnish teachers said they have a high degree of professional autonomy that helps them prioritise classroom assessment as they see fit. The Finnish National Agency for Education's own website explains "the focus in education is on learning rather than testing."^{lxxviii}

The purpose of assessment at Saunalahti School in Espoo is to "guide and encourage students' learning process," said vice principal Minna Welin, as opposed to holding teachers or schools to account. In the absence of pressure on teachers to produce data for accountability purposes, they are able to focus on providing formative feedback to pupils without worrying about a 'data trail'. Therefore the vast majority of assessment decisions rest on teachers. Much of the feedback pupils receive is verbal, explained Pekka Peura, a teacher, school principal, and education blogger. This means the feedback can be highly individualised and given immediately.

Teachers also use resources including textbooks to assess pupils, checking whether or not key content has been learnt. Sonny Johnson spent time training in two Finnish elementary schools, and recalled "you see the kids walking through the snow with these huge backpacks with about 10 textbooks in their bag." The textbooks are provided by the school, and decisions about which textbooks to buy are taken by school principals. Teachers do not feel using commercially prepared resources undermines their professional authority, in part because the textbooks are often designed in close collaboration with teachers. Textbooks often contain test questions or other specific ideas for assessments, and accompanying software packages that can include online assessments. Many Finnish teachers see textbooks as part of their professional armoury, providing well thought through ideas that complement lesson activities, and help teachers identify meaningful next steps for their classes. Pupils are set homework regularly, and it is often not possible for teachers to mark everything in school. Mikaela Sumeli teaches in an elementary school in Helsinki, and marks "almost every Sunday, for two to three hours."

One drawback of this teacher-led assessment is that it makes it difficult to compare the consistency of teachers' feedback. Furthermore, 2016 reforms to the national curriculum sought to give pupils greater ownership over their assessment in the classroom. Several teachers said this could further exacerbate the challenge of ensuring assessment and feedback is of a consistently high standard.

Although there is no Ofsted-style school inspection in Finland, teachers are held accountable in other ways. One teacher said "nobody tells us what to do", although schools are subject to evaluation^{lxxix}. The difference, Tim Oates, explains, is in how accountability data is used; in Finland, results are published at a national level only (with school-level results shared with individual schools). What is more, during the 1970s and 1980s when Finland was transitioning to fully comprehensive system, inspections and testing were coordinated and overseen centrally.^{lxxx}

Taking data as a starting point for a conversation about school performance – Ontario, Canada

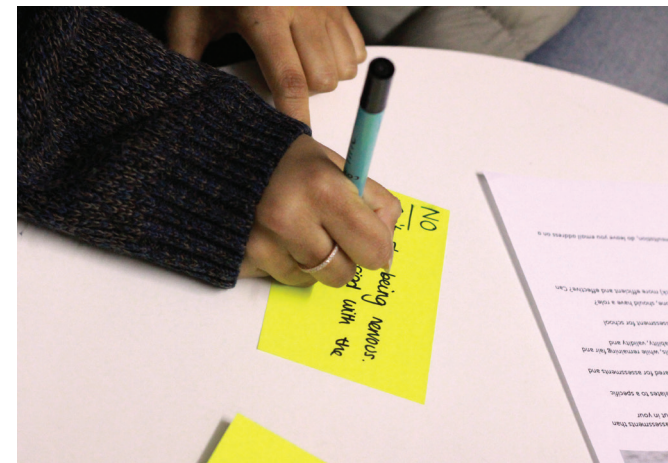
Pupils in Ontario, Canada, sit provincial assessments administered by the Educational Quality Accountability Organisation (EQAO) at the end of grades 3 (age 8) and 6 (age 11) in literacy and maths, grade 9 (age 14) in maths, and grade 10 (age 15) in literacy. From these, schools receive a report from the EQAO, containing 'IIRs', or 'Itemised Information Reports', providing analysis question-by-question of how students performed. The data can be accessed by the public online, and the EQAO does not rank schools (although several other research organisations do).

The EQAO tests feed into "robust conversations" about school performance, explained Andrea Gillespie, the superintendent of learning with the Trillium Lakelands District School Board, but these conversations are growth-orientated rather than punitive because super intendants and their teams work with schools over a period of time (rather than conducting one-off visits) to support improvements. She and others feel the EQAO provides a 'needs assessment', showing how schools are doing and what additional support they might need.

EQAO results do not result in knee-jerk reactions from staff either within the school, or within the

Board. Several reasons are given for this:

- Pupils are taught by multiple teachers, even before their first EQAO assessment in grade 3. Consequently laying blame at the feet of a single practitioner would be inappropriate.
- Actions that are seen as punitive, such as naming and shaming individual schools, are seen as undermining attempts to improve the system as a whole. Karen Dobbie at the Ministry of Education said "when there are problems, it's not just about that EQAO score." In other words, a low EQAO score is likely to indicate schools need professional or financial support (enabling a school to hire specialist maths training, for example). Removing individual school leaders or teachers could miss the problem, which is unlikely to be driven by one member of staff.
- The system relies upon collegiate working between school leaders and superintendents. Karen explained that school leaders approach superintendents when they need additional support. This openness would be undermined were accountability to focus on punitive responses to school underperformance. Jenn Clark, a school administrator (the school's most senior manager), said that while there is limited 'top-down' accountability in the Ontario system, teachers and school leaders feel very accountable to their pupils and colleagues.



5.2.2 Ways forward

The Department for Education should build experimentation and evaluation into assessment and accountability reform, to better understand how its reforms impact upon standards in schools

There is little consensus about how assessment can best inform accountability and school improvement. The accountability system – and how assessment informs this – should therefore be rigorously evaluated. This was emphasised in particular by Professor Coe.^{lxxxix} Doing so would help ensure more robust evidence is generated about the impact of policy decisions on system-level performance. It would also increase our understanding of how statutory assessments affect performance throughout different phases of education.

One approach would be different regions designing their own accountability and school support frameworks, within pre-defined parameters set by the Department for Education (which might, for instance, specify initially that every region had to assess pupils' performance in English and maths at baseline, KS2 and KS4, and publish the results). Regional Schools Commissioners would have an important role to play in designing these approaches. Each region's approach would be subject to rigorous evaluation in order to assess its impact across a range of outcomes. The 'best' models could then be adopted across other regions, re-evaluated, and so on.

Over time, evaluation of differing systems would begin to highlight the most effective ways of using accountability (and subsequent intervention) to support school- and system-level improvement.



If I could wave a magic wand...

Putting learning first: Dame Alison Peacock, Chief Executive, Chartered College of Teaching

Last week I attended a headteacher conference where a panel of young people from year seven were invited to reflect upon their experiences of primary school. The three youngsters had each attended a different school prior to joining their secondary school. There were nearly forty adults in the room but as the young panel began to settle into the situation they soon lost their nervousness and contributed their insights with razor sharp clarity.

In turn, each of them described the increased pressure, the narrowed curriculum, the lack of opportunity for science, drama, sport, and art, and how school trips were always followed by a written recount task. One of the children described how during SATs week her class was set additional tests by the teacher each afternoon in preparation for the main test the next day. By the end of the week she felt confused and overwhelmed, even though the actual SATs tests had 'been OK'.

Alijan was grateful for the quality of teaching he had received in year 6 because 'at the beginning of the year I was only emerging'. He regretted the fact that in years 4 and 5 supply teachers had taught his class; he felt no-one was there to help him. He described his year 6 teacher as very kind and helpful. She had taught him a great deal and by the end of the year he was proud to say he was now at the 'expected' level. It was painful to hear of the lack of opportunity for Alijan in comparison with his peers on the panel, and even worse to hear him describe himself as "only emerging".

Billy told us that secondary school was much better than he had thought it would be. Throughout his final year in primary he had been warned about how hard he needed to work in order that he would be able to cope in his new school.

The panel sent shock waves around the room. A narrowed curriculum, over testing, fear-mongering to ensure hard work, and relentless pressure. Assessment is not the enemy but hyper-accountability is. We need schools that have the courage to put learning first within the context of a rich and rewarding school experience. Such schools find that if the input is right the output looks after itself.

The Department for Education might then revise its parameters so that, for instance, not all pupils sit the same statutory assessments.

Withholding information about how assessments will contribute towards accountability measures

Assessments of pupil and school performance based on accountability measures tend to improve over time, yet research indicates that performance against equivalent measures does not.^{lxxxii} For example, pupils' KS2 reading levels have improved over time, while their standardised reading scores have not.^{lxxxiii}

One way to address this would be not to share how performance will be calculated with schools in advance. For example, in advance of the publication of performance data, schools might not know:

- How different subjects will be weighted;
- How many subjects will be included, or;
- How the performance of different sub groups of pupils (if any) will be weighted.

Such an approach would – quite rightly – be open to criticism that it is in some ways unfair to practitioners. However Professor Coe suggested it could form one of the approaches trialled in his above proposal.

Ofsted should enhance assessment training for its inspectors

As was argued in section 3, improved assessment training for teachers is urgently needed. Similarly, school inspectors also need additional training in assessment theory and practice to ensure they can constructively critique schools' systems. On the other hand, Alex Quigley, the Director of Learning and Research at Huntington School, pointed out there is risk here that this could veer into "telling people what to do."

The Department for Education should present schools' performance data as three-year rolling averages

In line with calls made by other organisations,^{lxxxiv} the Department for Education should present schools' headline performance data as a three-year rolling average. This is particularly important for primary schools, but should also apply to secondary schools. This would help reduce the volatility of the data, while ensuring action can still be taken in the case of perceived underperformance. Discussing the idea, Tim Oates said:

"We know that education can improve slowly and deteriorate fast. So we do know that one year's really bad results could mean a significant deterioration in the quality of the school. We know that is true and we are not saying that a bad year should not be submitted to scrutiny. But a bad year is not solely explicable by virtue of the school having done anything bad or different to what they were doing before."

Tim Oates CBE



Using teachers' judgement to award final grades

Many teachers and school leaders feel teachers' professional judgement should be given greater status in statutory assessments. However, as is argued above, this would be problematic because of poor reliability and potentially biased judgements. However, Wiliam suggested that teacher assessment could actually improve the process of allocating pupils with grades.

For Wiliam, the central challenge is "combining the knowledge that teachers have about their pupils with hard-edged information that allows us to ensure that consistent standards are being used".

^{lxv} One solution would be to sample pupils so as to define an 'envelope' of grades, which teachers could then use to allocate grades to pupils.

Wiliam proposed a system whereby:

- Teachers keep records of their pupils' achievements throughout the course of study.
- A formal external assessment then takes place at a pre-specified time (preferably towards the end of the course so pupils have a maximum-possible amount of learning time). Different pupils could sit a variety of different assessments, some taking

- forms of written assessment, while others prepare individual or group presentations, for example.
- Having submitted the names of all eligible pupils to the Department, pupils in the testing year are randomly allocated to receive particular assessments (perhaps involving different sorts of individual or group written tasks).
- All assessments are marked externally, and some (an increasing number in the future) by machine.
- Teachers receive the results as a profile of marks or grades, for example telling them that, in their maths class, four pupils achieved grade 9, five grade 8, seven grade 7, two grade 6, and so on.
- The teacher then allocates to pupils these grades, using the evidence collected throughout the year.

One advantage of such an approach is that a teacher's subsequent allocation of grades would be reliable, based on hundreds of hours of assessments throughout the course, while also aligning with the standards set across schools nationally (because he or she would receive an envelope of externally-set grades to award). Furthermore, the only way to 'teach to the test' would be to improve the capacity of all pupils to perform all the tasks.

Wiliam identifies three weaknesses of the model:

1. There is a lack of transparency, because a pupil's mark in the final assessment would not necessarily link to his or her final grade. This would need to be addressed by careful and clear communication between the teacher, pupil, and parents.
2. A teacher's personal bias might affect how they choose to allocate final grades. This could be monitored by looking at the correlation of the levels awarded and scores achieved during the assessments. This could be corrected for by checking that, on average, the pupils achieving the highest marks in the assessments are being given the higher grades.
3. Pupils need to be motivated to do well, even though the impact on their final grade of poor performance on the test is small.

The assessments would not give an especially reliable indication of a pupil's achievement on any specific task. The average achievement of the class on all tasks, though, would give a reliable indication of the distribution of achievement in the class.

The system would also allow for different styles of assessment to be introduced over time, alongside any necessary training and support to teachers.



If I could wave a magic wand...

Redesigning accountability to support assessment for learning: Mary Bousted,
General Secretary, National Education Union

I would reform the accountability system so that assessment could be used for its proper purpose of supporting pupils' learning.

Wouldn't it be great if our education system let us use assessment to support pupils with their learning, and also gave us valuable information about how the system is performing? I ask because the current assessment and accountability system fails in both those tasks, and worse, places an unacceptable strain on many teachers and pupils.

Instead of testing every pupil at multiple points in primary school for example, we could introduce a new, better system that separates assessment for learning from evaluation of national standards and trends. This would start with rigorous training and support so that teachers were confident in their formative assessment of pupils. Then we could introduce a national assessment bank, filled with standardised activities and tools that teachers could use to enhance their practice across all subject areas. A national sample testing system introduced across different phases of education and subjects could provide rich and meaningful data about how schools perform, identifying strengths and weaknesses, measuring how the system performs for different groups of children. Schools would then evaluate their own performance against the national picture, enabling critical reflection and genuine school improvement. Reports to parents would provide meaningful information, rather than the current system where results like these tell parents nothing useful.

The close relationship between the assessment and accountability systems means that government continues to attempt to control the behaviour of schools through high stakes testing of the next ministerial whim - be it phonics, grammar or times tables. Instead, assessment should enable the profession to support pupils' learning, to report openly and transparently on the strengths and weaknesses of schools and to hold ministers to account through sample testing for the impact of their policies on education.



5.3 Challenge: Teachers and parents have misperceptions about the purposes of statutory assessments

Perceptions are critically important, because they affect the extent to which different elements of the education system, including curriculum, assessment, pedagogy, and accountability, pull in the same direction (a feature described in the literature as system 'coherence').^{lxxxvi}

It is neither surprising nor unreasonable on the part of school leaders and teachers to worry about their pupils' results. As Becky Allen put it: "Heads roll on the basis of these tests." Consequently, school leaders "can only relax if everybody else (other school leaders) is relaxing," leading to anxious behaviours in schools. However, this can also lead to misperceptions of the purpose and ramifications of external assessments.

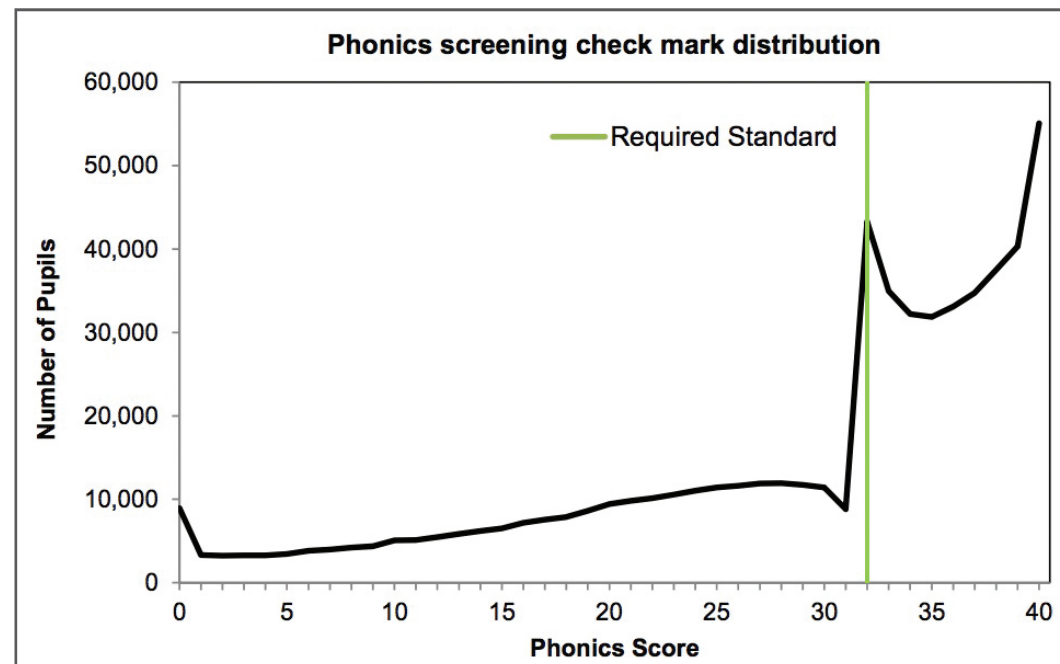
This was exemplified in the introduction of the phonics screening check, and the 'spike' in the mark distribution at the pass mark. The chart, right, is from a Department for Education publication,^{lxxxvii} and shows the number of pupils obtaining specific marks in the 2012 phonics check.

In 2014 the Department decided not to announce the pass mark for the check in advance, and the mark distribution levelled out to some extent.^{lxxxviii} Yet, the intention of the phonics screening check had never been to hold schools to account. Instead it was designed to show the government whether

pupils had reached a particular level of phonological awareness. The phonics screening check is not a metric used when calculating an individual school's performance,^{lxxxix} and school-level results are not published in performance tables.^{xc} However, because the check was perceived to be of consequence to schools, schools and teachers felt incentivised to give pupils 'the benefit of the doubt' around the pass mark.

Unreliable teacher judgements were also a key rationale for Ofqual's decision to stop speaking and listening assessments from counting towards GCSE English grades.^{xcj}

Ultimately, Christodoulou argues that almost irrespective of what Ofsted have specified centrally, hearsay and what "people think Ofsted want" with regards to assessments can lead to unhelpful practices in schools, particularly around the generation of internal data.

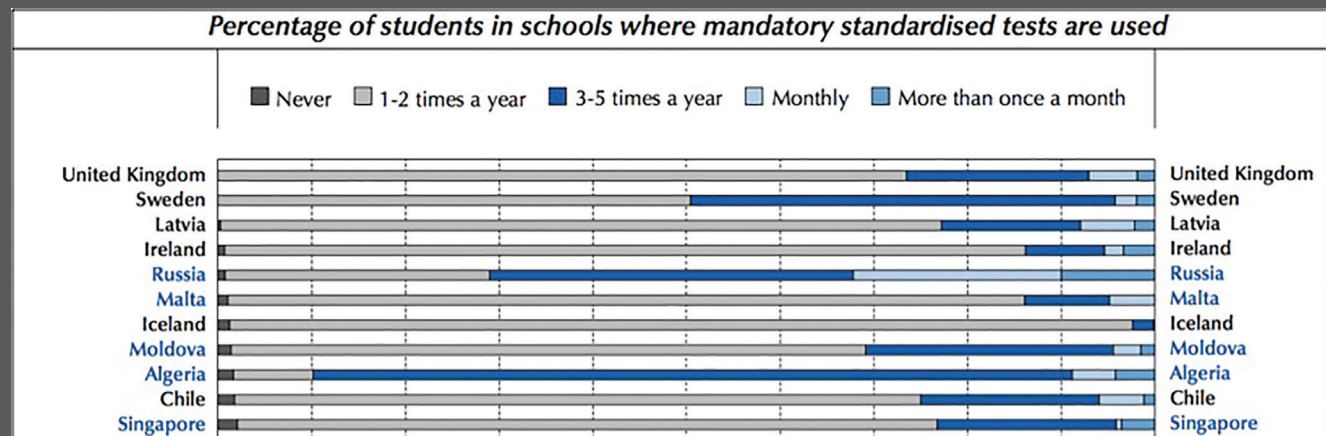


How many statutory tests and exams do pupils in England take, in comparison to other countries?

Teachers in England often feel that their pupils are subject to more testing than young people in other jurisdictions. This came through particularly in the online consultation responses, with many respondents suggesting pupils should sit fewer statutory tests to bring the country into line with others.

However, 2015 PISA OECD results shows that while virtually no young person passes through the UK

education system without sitting a mandatory standardised test during their education, a number of OECD jurisdictions administer such assessments with much greater regularity.^{xcii} Every OECD jurisdiction has a national assessment in place at either lower or higher secondary level, with the exception of Switzerland. What matters, therefore, is how tests are perceived, and how adequately teachers and pupils feel able to prepare for them. The picture, below, is an excerpt of a table in the 2015 PISA OECD report.^{xciii}



5.3.2 Ways forward

The Department for Education and other organisations (including schools) must ensure teachers can access high quality assessment training and resources

As was outlined in section 3, better assessment training for teachers is urgently needed. This would have the additional benefit of improving teachers' understanding of assessment reform, and help correct misperceptions around assessment and accountability.

The Department for Education must work with stakeholders including teachers' unions to 'myth-bust' statutory assessments

Ofsted has worked to publicly counteract misperceptions around school inspection through its myth-busting campaign^{xciv} and through clear statements from the Chief Inspector and her central team.^{xcv} Ofsted should continue to issue such statements.

In addition, the Department for Education should publish myth-busting material around the purposes of different assessments, including the Early Years profile, intermediate 'checks' (including the phonics and newly-announced multiplications check), SATs tests, and GCSEs. Such documents should clarify:

- How different assessments will contribute to evaluations of school performance throughout different phases;
- Why assessments are constructed as they are (for example, why there is a reliance on tests over teacher assessment results), and;
- How reliable these different assessments are, and consequently how confident practitioners and parents can be in the results.

The Department for Education must work with stakeholders including teaching unions to ensure these messages reach school leaders and teachers, and are seen as credible.

Furthermore, the Department for Education and Ofsted must ensure that their representatives provide consistent messaging about the purpose of different assessments in their day-to-day correspondence with schools.



If I could wave a magic wand...

Using assessments to support intelligent accountability: Paul Whiteman, General Secretary, NAHT

In order to improve the educational experience for all pupils and reduce burdens on schools, the current high stakes nature of statutory testing and exams, the punitive accountability system, and sanction-driven approach to intervention must be addressed.

Statutory assessments and national examinations will never be able to capture all aspects of a pupil's progress or all the different ways in which a school contributes to this. Data generated by these tests and exams 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.

As education professionals know, exams and tests aren't an appropriate method of assessment for all students or subjects. They are restrictive and don't allow students to demonstrate all that they can do; some tests are effective for assessing certain types of knowledge and less effective for others. The focus on testing and exams condemns many pupils to believing that they are 'not good enough'.

An overemphasis on statutory testing and examinations and the high stakes attached to the data has a negative impact on the breadth of the curriculum, the teacher and pupil experience of teaching and learning and can be harmful to the health of children and young people.

But the school accountability system doesn't allow for schools to choose which assessments work best for their pupils. Nor does it allow space for the broad and balanced curriculum that fully prepares our children and young people for life after school. You would have to be a very brave school leader to completely ignore the accountability measures on which so much depends.

Therefore, we believe the profession needs to continue to speak the truth to government, so that it can address these challenges; 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.

6. How can the workload associated with assessment be reduced?

What are the problems and challenges?	What are the ways forward?
<p>Schools often have inefficient assessment practices, and in particular an over-reliance on:</p> <ul style="list-style-type: none">• Heavy marking, and;• Mock tests and exams.	<p>Improved access to quality training and resources could shift schools towards more efficient assessment practices, specifically:</p> <ul style="list-style-type: none">• Reduced use of marking;• Keeping the number of mock tests and exams to a minimum, and;• Using technology effectively.



6.1 Challenge: Schools often have inefficient assessment practices

Assessment frequently has a profoundly negative impact on teachers' workload, respondents to our online consultation told us. While assessment is a fundamental part of teaching and learning and therefore deserving of time and attention, respondents argued that teachers spend too long on the 'wrong' sorts of assessments, namely "over-assessing children in order to have 'current data'". This view was emphasised by teachers, governors, and parents, and aligns with other surveys exploring how assessments impact teachers' working lives.^{xcvi} Ultimately, "feedback should be more work for the recipient than the donor", although few teachers feel their pupils spend as much time processing feedback as they spend creating it.^{xcvii}

Practices that can be particularly draining for schools include an over-reliance on:

- Heavy marking, and;
- Mock tests and exams.

Heavy marking

Several studies and reports highlight how marking and planning of assessment impacts on English teachers' workloads,^{xcviii} and the Education Endowment Foundation suggests that schools can rely disproportionately on forms of marking that only have a limited impact on learning.^{xcix}

Evidence suggests marking can be useful, but that there is a practical limit to its utility.^c As Professor Coe explains "a lot of marking is pretty ineffective in terms of promoting student learning:"

"We need to better understand as a profession how to invest our time in providing meaningful feedback, and I'm not convinced that narrative marking is necessarily the most effective way of doing that."

Simon Knight

"At the moment marking, in the way that it's required in these schools, is structurally impossible. Even just marking one set of Year 8 RE or geography or history essays is going to take up half a teacher's weekend. And it's simply not sustainable."

Christine Counsell

Mock tests and exams

Schools undertake mock tests and exams for a variety of reasons, including giving pupils experience of taking assessments under test conditions. They also undertake such assessments to track pupils' progress over time.^{ci} Testing can be hugely beneficial for pupils and teachers, for example supporting retrieval practice, and identifying gaps in existing knowledge.^{cii} However, running regular, full-form tests (as opposed to conducting more focused, formative quizzes, for example) is time consuming both for pupils and teachers, and can be an inefficient distraction. Quigley explained why, saying:

"It jumps straight to the big game, straight to the mock exams, straight to the big essays, straight to the full pieces of work, and what that has created is an excessive workload..., and at the root of that is a misunderstanding about how children best learn in preparation for terminal exams."

Alex Quigley

Increasing assessment efficiency and impact – St Matthias Primary School

St Matthias Primary School's approach to assessment and feedback, and in particular its marking, was historically seen as strong. However, the headteacher Clare Sealy said that it created an unsustainable workload for teachers who had taken to carrying children's books around in wheely suitcases. In 2016 Clare decided to overhaul the school's entire approach to assessment, seeking to maintain its efficacy while improving its efficiency.

In terms of marking, Clare told teachers in the 2016 summer term to stop marking altogether. She specified no marking whatsoever, as she felt otherwise teachers would have still carried on. Instead, teachers looked at pupils' work, made notes about common misconceptions and mistakes, and revisited these during the next lesson. Teachers discussed how they were finding the process and, in general, they were extremely positive; the approach cut-back on their marking by around two thirds, without reducing the quality of teaching. If

anything, teachers felt able to put more time and energy into planning and teaching. Overall, marking now generates "more work for the pupil than the teacher."

Clare has now introduced further changes to the schools' assessment practices in order to continue reducing teachers' workloads, while ensuring teachers and pupils get prompt and useful feedback. The school's approach to assessment varies by subject. However, common strategies include:

- 'Do nows', which take place daily, and test pupils' knowledge of recently taught content, and;
- 'Check its', which test pupils' knowledge of particular topics three weeks after the content was first taught.

In maths and writing, pupils take termly summative assessments, with the results at the end of the year reported to parents. In humanities subjects and science, pupils take multiple-choice quizzes at the end of units of work, and then larger quizzes at the end of the year.

Pupils also sit Rising Stars standardised reading and maths assessments each term, because "you do need some sort of external benchmark, otherwise you could be living in this fool's paradise."

Harminder Dhanjal is a year 3 teacher, and said the school's approach to assessment (especially marking) had been very positive:

"In literacy, you'd spend a lot of time marking and giving appropriate feedback to children, but we've really turned that around. There's less focus on the teacher spending hours and hours marking books and pinpointing for the children where their mistakes are. Now, children are in the driving seat, so they're the ones taking control of their learning. They are putting in the work to identify mistakes. They learn and become more enriched by the process."

Far from reducing children's progress, Harminder believes that, "if anything, children are more able now ...because they're in control of their learning."

6.1.2 Ways forward

Improved access to quality training and resources could shift schools towards more efficient assessment practices

Improved access to quality training could raise awareness of more efficient assessment practices amongst school leaders and class teachers. In turn, this would help them implement structural changes leading to more sustainable workloads.

Reducing reliance on marking

In its review of the evidence on marking, the EEF suggested that:

- Awarding grades for everything can distract pupils from teachers' formative feedback;
- Pupils must be given time to reflect upon and respond to marking, and;
- Some marking, such as 'acknowledgement marking', is unlikely to enhance pupils' progress.

The EEF explains "schools should mark less in terms of the number of pieces of work marked, but mark better." Furthermore, pupils need to be involved heavily in the process, and in responding to feedback. Without this, Alex Quigley describes marking as representing "a huge workload factor with little impact on learning."

Using technology to support daily classroom assessment – Chesterton Community College

All pupils have iPads at Chesterton Community College, which is part of the Cambridgeshire Educational Trust. Richard Auffret is the Trust's Director of Curriculum and Technology, and said the school decided pupils should have iPads because of their potential to support learning, and in particular in relation to assessment.

Teachers can quickly collate and review pupils' answers to short quizzes and tests during lessons, determining the next steps in teaching. Teachers use a range of online assessment tools to assess pupils, many using multiple-choice quizzes through apps such as Kahoot! and Quizlet. Pupils also use the app SeeSaw to upload evidence of their work, including photos, and to share this with their parents. Technology also supports students in the build up to their assessments, as teachers can share revision resources online.

The main benefit of all this is in closing the learning and assessment 'feedback loop'. The apps also give teachers the flexibility to, for example:

- Make results public or private;
- Have pupils sit quizzes individually, or in groups, or;

- Record pupils' presentations, which is useful both for the purposes of gathering evidence of pupils' achievements, but also if pupils are not in school and need to submit presentations remotely.

Overall, the technology means pupils 'cannot hide', so teachers know how all pupils in their classes are doing, not just the more vocal ones. It is possible to use technology badly, though, and this can result from:

- Teachers using technology in ways in which they are not comfortable (such as attempting to use apps with which they are unfamiliar);
- Teachers using technology for the sake of it, rather than because it will support and extend a learning activity, or;
- Infrastructural issues, such as poor WiFi connection, or iPads losing charge.

To counteract these risks, teachers receive training including 'TeachMeet' style meetings. These involve teachers giving short presentations about ways in which they are using the iPads to conduct assessment, and then discussing these ideas with their colleagues.

Counsell argued that teachers are better off making notes on common errors that arise when they mark a group's essays, "and doing some darn good whole class teaching afterwards." Counsell argued this is not only effective pedagogy, but massively reduces the workload for the teacher, because "the work of reading those essays is an hour to an hour and half max, as opposed to six or seven hours of ploughing through every single child's essay and writing feedback."

The Independent Teacher Workload Review Group argued that:^{ciii}

- Senior leaders and governors should review school policies to ensure class teachers do not spend longer than necessary marking.
- Schools should make the most of other forms of feedback including self- and peer-assessment, and verbal feedback from teachers.
- Senior leaders should take notice of Ofsted's clarification materials. These specify that inspectors do not expect to see feedback of any specific form, frequency or volume.
- Schools should support universal expectations with regards to marking while giving individual subject areas freedom to use the approaches they deem most appropriate.

Keeping the number of mock tests and exams sat by pupils to a minimum

Schools should reduce their reliance on mock exams by ensuring pupils only sit one or two full mock exams in the build up to their exams.

Automating marking to give students more practice while freeing up teachers – STACK

STACK provides an automated means of assessing mathematical procedures. It has been developed by Chris Sangwin, Professor of Mathematics Education at the University of Edinburgh. It provides quiz questions, and can give students feedback on their responses, showing them how and where they can improve, and providing teachers with marks and underlying statistics about students' performance. STACK can assess mathematically-based procedures, and is used by maths, engineering and physics departments in universities in the UK and across Europe.

Chris explained that lots of electronic assessments are multiple-choice and can be very helpful, but ultimately require students to make selections from options presented to them. STACK is instead about students being completely responsible for providing the right answer. The software helps identify where mistakes have been made, and can now mark some answers line-by-line.

This is exciting because to improve at maths or maths-based subjects students need lots of practise. "No one wants to mark this kind of

stuff", though, because it would be incredibly time-consuming, "so technology is ideally suited" to these sorts of assessments. STACK therefore offers a practical means of giving students plenty of opportunities to apply mathematical skills and procedures, while not putting teachers under immense pressure to mark it all; "we automate the things that can be automated, and devote staff time to the things that really require a human marker."

STACK can provide feedback to students in real time, or after the event, although Chris explained the evidence is unclear on which approach is more helpful.

Chris believes the system has enormous potential, and within five years the technology may be reliable enough for marking A-level maths and physics papers. However, in the meantime he identifies several challenges:

1. Knowing when to provide feedback.
2. Managing the user interface. Students could either write answers into computers, or hand write them and scan them in. Both have advantages and draw backs.
3. Even with the technology in place, writing good questions remains challenging.

Using technology effectively

Technology already plays an important role in assessment in many school settings, and there is scope for it to further reduce teachers' workload.^{civ} It can support assessment by, for example:

- Improving the questions teachers ask;
- Sharing knowledge and information between teachers, between teachers and their pupils, and between schools, parents and governors;
- Ensuring pupils are assessed 'responsively', so that questions get harder as they answer them correctly, and easier if they answer incorrectly;
- Automating marking and in some cases automating feedback, and;
- Streamlining and improving the reliability of moderation processes.

Professor Wiliam said that "there's no doubt in my mind that technology is going to become really, really good, really, really quickly," offering the potential "for a complete panopticon of students' work."

On the other hand, while technology can produce procedural efficiencies for teachers,^{cv} Professor Torrance explained "the data that goes in still has to be as solid as possible." There also remain significant barriers to technology becoming embedded in teaching and learning, some of which are attitudinal, some practical.^{cv} Attitudinal concerns include overly optimistic or wary views of the ways in which technology can support classroom teaching and assessment. Practical concerns include:

- Up front and ongoing costs of ensuring technology is accessible and ready to use, and;
- Teachers' ability to use technology smoothly, as part of their teaching, and as opposed to a 'bolt on'.

Furthermore, technology offers the potential to capture all of pupils' attempts and achievements, but this does not mean it *should*. Professor Wiliam expressed concern that pupils still need the space to take risks and make mistakes. Capturing these too soon could do the young person a disservice, if it is not their latest and best effort.



Improving the reliability of teachers' judgements – No More Marking and comparative judgement

No More Marking is an organisation developing comparative judgement. Comparative judgement is based on the principle that people are better at making comparisons between pieces of work than at making absolute judgements about quality. No More Marking has developed software that enables teachers to compare two pupils' work (for example, two essays exploring the question: 'What were the causes of the Second World War?'). Each teacher simply judges which response he or she thinks best addresses the question. Once enough teachers have judged enough pieces of work, it is possible to see where an individual pupil's work fits into the broader distribution of achievement, and standardise pupils' achievements to see how they compare with other pupils, nationally.

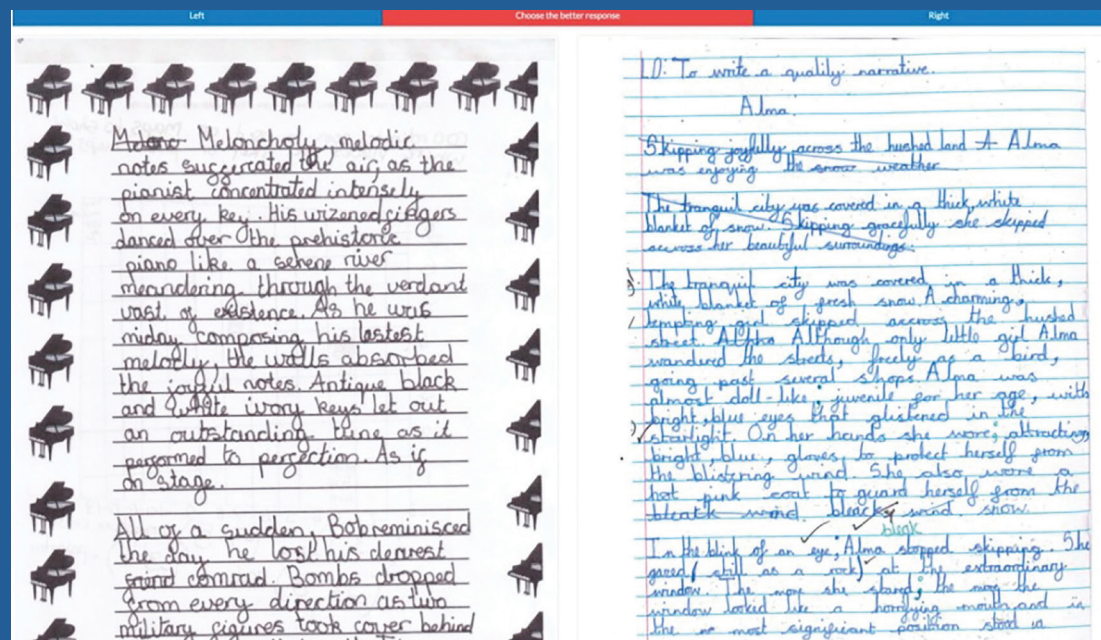
The government said it is going to explore the potential for comparative judgement in supporting assessment in primary schools in its assessment consultation response, published in September 2017.

Daisy Christodoulou is No More Marking's Director of Education, and said she feels the approach has three key benefits:

1. Reliability: Pilots of the comparative judgement approach suggest it is more reliable than methods traditionally used to assess pupils' written work.
2. Validity: The assessment of written work using marking criteria has historically proven problematic, leading to pupils jumping through hoops (for example, using clunky or nonsensical frontal adverbial phrases). Comparative judgement involves making a holistic evaluation of a piece of work. In Christodoulou's words, "you're actually assessing the thing you want to assess."

3. Efficiency: Comparative judgement enables markers to assess pupils' written work more quickly. Christodoulou explained that the average time it takes users of comparative judgement to make a judgement is 38 seconds, far quicker than the time taken to grade an essay using marking criteria.

Image taken from the No More Marking website of pupils' work, as displayed to markers on screen:



Christodoulou also explained that, while in its relative infancy, comparative judgement could be used for peer assessment by pupils too. The approach enables evaluation of the sorts of features judges identify and reward when assessing work, "so if you set up pupils as judges you can see to what extent they're agreeing with the teachers."

While the evidence supporting comparative judgement is encouraging, the approach represents an entirely different way of evaluating pupils' work:

"If you're used to marking with criteria and having those moderation meetings it's really different and it can feel very subjective. The weird thing is it's much more rigorous than traditional marking but it doesn't feel it."

Daisy Christodoulou

Other challenges are practical, such as inputting pupils' work before conducting the assessment.

Focusing on effort and progress, and putting the onus on pupils (and not just their teachers) to work hard – Huntington School

Prompted by the removal of levels in 2014, staff at Huntington School in York reviewed the school's approach to assessment. As a result, the school has introduced a number of changes to ensure assessments consistently incentivise pupils to work hard while making progress:

- Teachers know what pupils' academic targets are, but the school no longer shares these with students or their parents. Assistant headteacher Garry Littlewood explained "we had too many students coming into lessons and saying 'I've got my 'C'", and then having no aspiration to improve this. Equally, pupils with high targets could become disillusioned if they did not achieve them. Garry feels the change is "encouraging more aspiration amongst the students."
- On entry in year 7, pupils are designated as low, middle or high starters, based on their KS2 scores and baseline CATs results. At KS3 and the majority of KS4 the school does not set or stream pupils meaning most classes are mixed

attainment. Students' academic progress is then described as 'exceeding expectations', 'meeting expectations', 'working towards', or 'underperforming', relative to their starting point.

- There are two points in the year, normally towards the end of the autumn and spring term, where teachers report pupils' current progress and effort. Subjects either conduct their own summative assessment, or report teachers' holistic judgements. Alongside reporting pupils' progress, teachers report on pupils' effort as 'Excellent', 'Good', 'Insufficient' (meaning the pupil is coasting), and 'Poor' (meaning the pupil takes little or no responsibility for his or her progress). Pupils said they often grudgingly agreed with teachers where their effort was highlighted as a concern, even in subjects they are 'good at'.
- Pupils sit a final summative assessment in each subject at the end of the academic year, designed and set by subject leaders in collaboration with class teachers. In KS4, the tests may be based on past exam questions, although Jane Elsworth, Assistant Director of the Research School, explained "we don't just want to move the GCSE questions down to year 9 but use KS3 to scaffold some of the skills needed for success at GCSE". Pupils'

marks are uploaded, and subject leaders then create mark distributions for each learning group. In collaboration with other subject teachers and their line managers, subject leaders then decide where boundaries lie for the academic judgements ('exceeding expectations' compared to starting points, and so on). Garry then receives each department's data, and can moderate to ensure expectations across subjects are comparable.

- Students receive ongoing feedback on their work throughout the year. Departments highlight, on schemes of learning, when teachers are expected to give detailed feedback. Garry and Jane explained that this approach cuts back on teachers' workload by reducing the amount of marking they are expected to do. After the detailed feedback has taken place, directed improvement and reflection time ('DIRT') is set aside so pupils can respond to teachers' feedback.

Garry and Jane both said that these changes had placed far greater onus on pupils to work hard consistently throughout their time at the school, irrespective of their starting points. Furthermore, it has reduced teachers' workloads because the practices themselves are more targeted.

7. How can statutory assessments and tests help all young people demonstrate their academic abilities, while providing trustworthy results?

What are the problems and challenges?	What are the ways forward?
<p>Statutory assessments and qualifications do not adequately capture all pupils' achievements, because:</p> <ul style="list-style-type: none">• They are based upon age-related expectations.• The accountability system incentivises schools to select certain qualifications over others.	<p>The Department for Education should stop reporting schools' performance using the 'EBacc' performance measure, and instead focus on Attainment 8 and Progress 8.</p>



7.1 Challenge: Statutory assessments and qualifications do not adequately capture all of pupils' achievements

National tests and qualifications are based on age-related expectations

One of the most prominent concerns expressed as part of our online consultation, focus groups, and by experts during interviews was that statutory assessments at present do not capture all children and young people's achievements.

Because accountability uses assessment to determine 'expected' performance at different stages, Simon Knight asked:

"Children with exceptional educational needs are often going to learn at an alternative rate to the majority of their peers. How do we accommodate that in a system that judges schools by how much children have learnt by certain ages?"

Simon Knight

This is a challenging question, because while it is important all children's achievements are celebrated, it is also vital that expectations for young people with exceptional needs are not lowered:

Meeting the needs of statutory assessments while supporting pupils' rounded development – Limpsfield Grange School

Limpsfield Grange is a special school in Oxted for girls aged 11 to 16 with social communication and interaction difficulties, the majority of whom have forms of autism. The school offers a mainstream curriculum, and most of the pupils take GCSEs at the end of year 11. Sarah Wild is the school's headteacher, and said that teachers need to juggle 'high and challenging' academic expectations with the realities of working with pupils with a diverse range of needs:

"[Pupils] don't always have a lot of head-space for doing heaps of learning. Part of what we've got to do is enable them to be calm enough for a sustained period of time so that they can access the curriculum and make great progress."

Sarah Wild, Headteacher

The school adopts many strategies that are common to mainstream settings. For example, students are grouped into 'low', 'middle' and 'high' prior attainment bands, which shape academic targets and teachers' evaluations of pupil progress. The school conducts end-of-unit assessments across the curriculum, and pupils sit mock exams so that they familiarise themselves with the

process. Furthermore, the school uses assessments from CEM such as 'MidYIS' and 'YELLIS' to triangulate its decisions. Sarah explained, though, that classroom and practice assessments have to be pitched just right, because if they are too hard, the students – who are prone to feeling anxious – may write off entire tasks.

While developing any young person's confidence and communication is vital, Sarah feels it is especially important for the students in her school. Pupils take GL Assessment's 'Pupil Attitudes to Self and School' (PASS) assessments, which gives Sarah and her team a sense of how the girls feel about themselves and their environment.

While academic grades matter hugely, Sarah does not prioritise these over pupils' personal development, as "there's absolutely no point in having a GCSE in RE if you can't talk to your colleagues at work." Consequently, teachers integrate activities that are designed to boost the pupils' communication, resilience, and self-confidence into the full curriculum, for example asking the girls to work in groups to discuss questions before feeding back to the class.

Pupils themselves find qualitative feedback helpful, "because you can actually say what you can improve on, rather than just 'here is a number'." They also talked about finding the reformed GCSEs anxiety-inducing, as the goalposts can feel like the keep moving.

"Because we already know that far too few go on to lead the lives that they potentially could, whether that be finding employment, achieving independent living, and living healthy and secure lives. If we make changes to the education system to accommodate that difference, we need to make sure that we do it in a way that enhances their outcomes not reduces them."

"A significant number of these children are capable of attaining a very high level, they just need additional support and maybe a slightly different pathway to get there."

Simon Knight

Assessments are getting harder

A related concern was around the higher standards set by the reformed assessments at Key Stages 2 and 4, meaning some pupils have been unable to demonstrate what they could do because these tests and exams do not provide a broad enough range of questions for them to do so.

The accountability system incentivises schools to select certain qualifications over others

Despite numerous efforts over several decades, academic and vocational qualifications have yet to gain parity of esteem. This issue is well documented, with a review in 2011 referring to many vocational qualifications as "dead end",^{cvii} offering little to no value to those taking them. This resulted in 2012 with the government removing over 3,000 vocational qualifications from league table measures.

In 2010, the government announced the introduction of the 'EBacc' (English Baccalaureate) performance

measure, indicating the proportion of a school's pupils who take GCSEs in English, maths, science, a language, and history or geography. Many of the effects of the EBacc on pupils and schools are still working through the system.^{cviii} However, focus group participants – and in particular, school leaders – expressed concerns that the EBacc:

- Is not suitable for all pupils, and that the measure may incentivise schools to enter pupils for these subjects even where they are not appropriate, and;^{cix}

- Could reduce schools' willingness to offer vocational courses that are not acknowledged as part of the EBacc measure.

The issue is not so much that high quality vocational qualifications are not available, but rather how the accountability system reflects the achievements of pupils who take them:

"A problem with performance tables is that children may be getting pushed into qualifications they cannot achieve and aren't suitable for them. But that is not a problem with the qualification, it is a problem with the accountability system."

Becky Allen



If I could wave a magic wand...

Conduct fewer statutory tests: Professor Harry Torrance, Director of the Education and Social Research Institute, Manchester Metropolitan University

Fortunately this is not a difficult test question. The answer is very straightforward. The one thing that would improve assessment in English schools is to do less of it – less formal, summative, test-based national assessment. We over-assess young children compared to school systems elsewhere and we can manage with far less. The government has recently announced the end to testing at age 7, though the actual finish to this is some years away. Some testing at age 11 should be retained. It is important to have some sort of 'staging post' in a child's school career whereby teachers and parents can review evidence of progress and identify any problems before transfer to secondary school.

However, GCSEs should be dropped as they are unnecessary. Young people have to stay on in full time education and training, beyond the formal school leaving age of 16, and these programmes and training courses have their own assessments – A-level, BTEC, NVQs, and so forth. There is absolutely no need for all secondary school students to sit all GCSEs at the same time, at age 16, like some mass ritual induction into the cult of academic success and failure. Some GCSEs could be retained for specific courses and purposes, to be taken when students are ready and need them to demonstrate achievement in specific fields.

In short, we have to break the grip of assessment on English schools and put it in its proper place – supporting the curriculum, underpinning learning and demonstrating the full range of student achievement; not dominating the curriculum, undermining learning, and reducing the definition of achievement to passing tests.

7.1.2 Ways forward

The Department for Education should stop reporting schools' performance using the 'EBacc' performance measure, and instead focus on Attainment and Progress 8

Attainment and Progress 8 scores are headline performance measures for secondary schools, respectively measuring a year group's achievement and progress across eight subjects, in comparison with pupils nationally with similar KS2 results. There is significant cross-over between the Attainment and Progress 8 measures, and the EBacc, as the boxes show, below.^{cx}

Qualifications that currently count towards Attainment and Progress 8:

- Maths (double weighted);
- English (double weighted, if both English language and English literature are entered);
- Three qualifications that count in the EBacc, and;
- Three further qualifications that can be GCSE qualifications (including EBacc subjects) or technical awards from the Department for Education's approved list.

The Department from Education should scale back the emphasis it places on the EBacc and, eventually, stop reporting this measure altogether, instead emphasising schools' performance under Attainment and Progress 8 (both of which essentially incorporate the EBacc). This would help schools ensure pupils can access a wide range of subjects, including a range of arts subjects, as well as the technical and vocational qualifications recognised in the Department for Education's approved list.

Qualifications that currently count towards the EBacc:

- English literature and English language, plus;
- Maths, plus;
- Science (either core and additional science, or three out of biology, chemistry, physics, and computer science), plus;
- A language (either a modern or ancient language), plus;
- A humanities subject (either history or geography).



If I could wave a magic wand...

Defining assessment broadly: Geoff Barton,
General Secretary, ASCL

I would shift our use of language so we see 'assessment' as something incredibly broad that can be applied in different ways at different times. Assessment should be a tool for learning, and less an instrument of accountability, enabling teachers to focus on pupils' learning and whether they have understood and retained key knowledge and skills, address misconceptions and, ultimately, motivate pupils to study harder.

As part of this, we should be discussing what helps pupils learn best, and how assessment supports this. For example, what are we doing every time we ask a good question? We are doing more than assessing their knowledge and understanding; we are making pupils think, prompting them to retrieve pieces of key knowledge. The right question will give teachers valuable information as to how well pupils have retained and understood key information, clues as to how well they can apply this learning in different contexts, and help pupils' develop their confidence and understanding.

Let us therefore expand how we construe assessment, unpicking what questioning, retrieval practice, and testing can mean and how these practices can best support pupils' learning.



If I could wave a magic wand...

Valuing all pupils' achievements through a wider array of qualifications: Adam Boddison, Chief Executive, nasen

It would be to have a more holistic approach that values both academic and non-academic achievements.

Currently, there is too much emphasis on academic progress and academic outcomes. Our notion of outcomes as a society must be broader, and we must have an approach to assessment that reflects this. A parent's dying wish for their children is unlikely to be that they achieve five good GCSEs including maths and English, or that they are pivotal in helping the school to secure a good Progress 8 score. Parents would want to know that their children left school feeling prepared to take on the demands of life and work, so they could lead happy and fulfilled lives.

To be clear, this is far more than simply valuing vocational subjects, although that is certainly a step in the right direction. It is about recognising that good progress and good outcomes are different for every individual and may not always be academic. For example, it may be that learning to live independently is a significant outcome for one young person, whilst for another success is about securing sufficiently strong academic results to get a place at a Russell Group university. Our current assessment system values one of these outcomes, but not the other. This is both sad and ironic because the personal characteristics required to achieve either of these outcomes, such as perseverance, courage and problem solving, may be evident in either case. These are skills that employers often say they are looking for, yet our assessment system immediately devalues non-academic progress and in the process categorises swathes of learners as failures.

I would like to see an approach to assessment that values both academic and non-academic achievements as this will help to create a more inclusive society in which everybody is able to positively contribute.

Providing and assessing vocationally-focused opportunities – Real World Learning Projects – School 21

At School 21 in Newham, east London, year 10 and 12 students spend 16 and 11 half-days respectively each term working with employers. These 'Real World Learning Projects' (RWLP) are the school's version of work experience, explained Hannah Barnett, a Senior Programme Officer and the programme lead.

The assessment and feedback process for RWLPs has several stages. Students begin with mock interviews with the employer that are held in school with two other students in their 'triad'. The interviewee will receive feedback from the employer and peers after completing the interview.

During their placements, students are assessed twice by the employer, once mid-way through and once again at the end. These assessments are based on the RWLP Student Performance Management frameworks, which Hannah said match the success criteria with which School 21 evaluates its staff. The criteria include whether a person is a:

- 'Pioneer', generating ideas and perspectives.
- 'Multiplier', committed to growing themselves and others.
- 'Craftsperson', dedicated to improving his or her own practice.
- Person of humanity and integrity.

Each of these overarching areas are further broken down (for example, being a 'multiplier' includes being a team player, and showing leadership). The employers score pupils from '0' (no evidence of the skill) to '3' (clear and competent use of the skill).

The mid-way review is a one-to-one conversation between the student and employer, in which the employer cites evidence of how and where the young person has demonstrated various skills and attributes. The student, equally, can explain how and where he or she has displayed certain skills. This process repeats at the end of the placement, and the employers tell students whether they would 'hire or fire' them. Students go on to present at the RWLP Exhibition, talking through what they learnt on the placements. Employers also provide students with a real employer reference that students can use if they wish.

Hannah said it is challenging to moderate employers' judgements, but that in many ways this is besides the point, because "the conversation is the most important part, over and above the scores."

8. How can unnecessary stress about assessment be reduced for young people and their teachers?

What are the problems and challenges?	What are the ways forward?
<p>Statutory assessments result in significant and unhelpfully high levels of stress.</p> <p>Linking test results to teachers' performance management can be counter-productive.</p>	<p>Schools should use more low stakes assessments.</p> <p>Schools must decouple pupils' test results from teachers' performance evaluations.</p>

Shifting teachers' conceptions of 'progress' to help all pupils access the curriculum – Margaret McMillan Primary School

Margaret McMillan Primary School in Bradford has given its approach to assessment a "massive overhaul", according to class teacher and literacy coordinator, Kayleigh Simmonds. This was prompted following reforms to the National Curriculum and primary assessment in 2014.

The school initially attempted to use levels to track pupils under the new curriculum, although teachers found that while children were working hard, the increase in academic standards under the reformed curriculum meant the academic gaps between

some children "just got bigger." Furthermore, levels were complicated to track and demoralising for the children explained Kayleigh and her colleague, SENCo Sabina Iqbal.

Now, the school sets age-related expectations for all pupils, except those with forms of SEND who cannot access National Curriculum content. Pupils begin the year as 'developing', irrespective of their performance in the previous academic year. Gradually, children move from 'developing' to a 'secure' category as and when this is appropriate in each specific area of the curriculum. Kayleigh said the shift had been hugely positive both for the pupils and their teachers:

"We've noticed a good impact on children's work ethics in school. Those children who find things difficult aren't giving up quite as easily. [They say] 'I'm the same level as so-and-so, who I know is really clever.'"

Kayleigh Simmonds

In some ways, the assessment system now in use across the school is more akin to that used in the Early Years, explained the school's coordinator of Early Years Provision, Adam Bagherian, where children's achievements are judged in relation to their age.

8.1 Challenge: Testing can create considerable stress

Statutory assessments result in significant and unhelpfully high levels of stress

Over 80% of school leaders worry more about pupils' mental health during assessment periods now than they did two years ago, and 79% have noticed an increase in stress, anxiety and panic attacks among their pupils during the same timeframe.^{cx} Spontaneous associations with assessment by parents tend to focus on stress and pressure.^{cxi} The stress and pressure of assessment for teachers and young people was a major concern flagged by respondents to our online consultation and during focus groups. Parents in particular worried about statutory tests and exams' negative impact on pupils' mental health.

One expert noted that tests are stressful for young people where they believe the results will have major consequences:

"The thing that is very stressful for children is going through the experience of sitting a high stakes test for the first time that has incredibly high consequences."

Becky Allen

Additionally, teachers do not all experience these pressures evenly, with upper KS2, GCSE and A-level teachers often bearing the brunt.

Linking test results to performance management can be counter productive

Teachers are often very stressed about their pupils' performance in tests. This is partly because teachers rightly care about how their pupils do, and want them to do well. However, it is also because their schools' performance and assessments of their own skill as teachers depend on pupils securing good results. Yet Becky Allen explained that teachers "know that regardless of whether they have done a good job in teaching their own class, the test results for whatever reason might not reflect that." This is because the tests capture the influence of teaching in part, but also 'unobservables' such as children's background:

"The tests don't reflect how good teaching is, they reflect a whole bunch of things, of which that is one. And yet they then get judged on whether or not they have done a good job of teaching."

Becky Allen

The lack of reliability intrinsic in assessment, including the effect of 'unobservable' factors, means linking teachers' performance to pupils' test results is problematic at best and, at worst, entirely unfair. Furthermore, identifying a single teacher's influence on a pupil's grade is extremely difficult (if not impossible), when that pupil may have been taught by any number of teachers throughout his or her schooling.

Using frequent low stakes assessment to improve feedback – West London Free School

Pupils at West London Free School take frequent and low-stakes quizzes in most of their subjects, the school's assistant headteacher Wade Nottingham explained. This is because doing so provides high quality information promptly to teachers about what knowledge pupils have grasped. In turn this enables teachers to quickly recap key concepts. The school undertakes a wide variety of other assessments, including termly summative assessments.

Many teachers use 'Show My Homework' to create quizzes, and quickly and easily analyse results. Quizzes differ considerably between subjects; for example, Ruth Fonseka-McFarlane is the school's Head of Biology. She explained that terminology-heavy subject quizzes often target pupils' spelling and definitions of key words. Pupils also take regular vocabulary quizzes in Latin, explained Head of Classics, Natasha Crook.

Quizzes help pupils and teachers build towards larger pieces of work because they ensure pupils get the facts in place, before learning how to deploy them, explained history teacher Oliver Bell. This is true across Key Stages 3, 4 and 5, suggested geography teacher Romy Bartram, as pupils across all age groups need to build up the basic knowledge before being able to tackle longer, more extended pieces of work.

8.1.2 Ways forward

Schools should use more low stakes assessments

While all assessments have consequences of some form or another, the results of low stakes assessments do not have a significant material consequence for pupils (or their teachers).

In addition to their educational benefits (discussed, above), taking a greater number of low stakes assessments could:

- Reduce teachers' workload (and therefore stress) as a result of the regular use of low stakes assessments that do not require intensive marking (such as multiple-choice quizzes);
- Over time, help pupils feel increasingly confident with different forms of assessment;
- Show pupils how different forms of assessment are of benefit to them, and;
- Reduce the stress of higher stakes assessments, as pupils will be more used to sitting tests.

Ultimately this could help pupils, teachers and parents understand, Tim Oates suggested, that "a test is just an opportunity for them to demonstrate what they know, understand and can do."

Schools must decouple pupils' test results and teachers' performance evaluations

It is extremely difficult for schools to reliably link one teacher to a class's performance in statutory tests or exams, and tests and exam results reflect many factors besides the quality of teaching pupils

received. Schools should therefore decouple pupils' test results and formal evaluations of teachers' performance. Instead test and exam data should be used as a prompt for conversations between teachers and line managers as one source of information about the areas in which a teacher might need additional support.



If I could wave a magic wand...

Remove the fear of assessments by using them to help pupils develop: Allana Gay,
Deputy Headteacher, Lea Valley Primary School

I would remove fear of assessment. Despite being a regular part of teaching practice within classrooms, when it becomes standardised testing, crippling dread takes hold and rationality and perspective are lost.

Schools fear judgements and ratings based on the attainment of their pupils. These judgements can then become evaluations of teachers' performance. Teachers, in turn, can project their stress onto students and parents. As such, rather than being part of the learning cycle, helping determine knowledge and understanding, assessment is castigated as the evil stepmother of education and treated with a mixture of trepidation and contempt.

So there are improvements to be made. Firstly, the format of assessment requires change from a single window system to a more flexible approach that allows for at least two assessment points through years 2, 4 and 6. In this way students will have multiple opportunities to demonstrate they have met the minimum expected standard.

Secondly, assessment data needs to be used as a developmental tool for individual children rather than providing a judgement about institutions.

There are governmental pressures on schools to deliver academic results at a given standard, regardless of circumstance. Since this is unlikely to change, it is the attitude a school takes in relation to assessment that will determine the atmosphere within that school. We must remove the obstacle of fear by seeing assessment as a means of helping pupils develop, rather than unfairly judging teachers.

9. Testing the Water – Pearson's Response



Rod Bristow

President of Pearson in the UK

Education must be an essentially collaborative endeavour if it is to achieve the best outcomes for learners. That's perhaps one reason it attracts such controversy and debate, and it's also why we sponsored this report, on which we are proud to have collaborated with colleagues at LKMco. Testing and assessment are controversial, but often stakeholders seem to be talking at cross-purposes and with varying levels of understanding about what is after all a highly technical, but incredibly high-impact activity. Given the undoubted impact of assessment on learning and on teachers, we wanted to provide a platform for a collaborative debate; a debate which would inform our own actions as well as perhaps, the actions of others. The report makes a number of recommendations and suggestions for how assessment in the UK can be improved. Some may be directed more toward some education stakeholders than others, but all are relevant to us at Pearson as one of the UK's exam boards and a worldwide education company. We take our role in assessment seriously and so we are responding to this report with energy and commitment.

At a system level, we need to reassess the purpose and nature of assessment, to separate out what is there to enable good teaching and learning and what is there to hold schools to account.

Accountability measures are important, but we expect too many different things of individual assessments – a student's performance in a GCSE exam for example, is expected to communicate different kinds of information to students, parents, teachers, schools, regulators, government, universities and employers. If we're to continuously improve the quality of assessments (as we should) each needs designing for a singular purpose. And if we're to reduce distracting pressures on teachers (as we surely must) it should be clear how that purpose serves the higher goal of improving teaching and learning.

To meet that higher goal we need much more than clarity of purpose. Teachers should also be provided with better support, training and resources founded on well-researched pedagogy, equipping them with a deeper and more technical understanding of assessment, as well as the tools to plan their curriculum, deliver high quality formative assessments, mark them, and use the insights to inform their teaching.

Our commitment, as part of what we hope will be a collaborative effort, is to do our utmost to help provide that support for deeper understanding. The only question, is how?

Increasing transparency demystifies assessment. At an SSAT meeting of head teachers a few years ago, I was asked why awarding bodies are 'Kafkaesque' when it comes to answering questions about assessment. One of the reasons I think, is a fear of being either misunderstood, or mis-represented – or perhaps inducing shock or indignation with the 'revelation' that no assessment is 100% reliable, as Dylan Wiliam points out. We've reached a point today however, where given the inevitable weight placed on assessment, we need to do better. We want to help demystify assessment by providing greater insight about what makes a good assessment – including how we match up, and how we'll continuously improve assessment quality. We'll act in four areas:

1. Improve transparency and training about assessment and its relationship to effective teaching and learning – The Pearson Assessment Charter

We will provide coherence and transparency in our work across formative, summative and high stakes assessments.

In 2018 we will be inviting students, parents, teachers and governors to work with us in the creation of a Pearson Assessment Charter. The Pearson Assessment Charter will lay out the requirements of good assessment – validity, reliability, comparability, manageability and minimising bias. We will be transparent in how our assessments meet

these requirements and commit to continuous improvement in the quality and design of our assessments across all our 'products' and services.

We'll offer more training on assessment to teachers and we'll make that training freely available to all those who work for us as markers.

2. Establish an assessment bank with supporting tools for building and using assessments that encourage and enable teaching beyond the requirements of high stakes exams

'Testing the Water' calls for teachers to have access to the best in class assessment materials and an understanding of how to design and use assessment effectively as a teaching tool. We have been talking to schools about the provision of a "build your own assessment" service, and will accelerate those plans. This will enable teachers to pick from a bank of assessment items with guidance about how to ensure the assessment provides appropriate differentiation and accessibility; for instance, for lower ability students or for undertaking the different kinds of assessments often necessary for creative subjects. The assessment items in the assessment bank will be a free resource to which we ask others to contribute should they wish. We will charge for the use of associated tools and services.

Technology can play an important role in addressing some of the workload challenges which can bedevil the use of good, formative and summative

assessment. Where valid and reliable our digital formative and summative assessments are auto-marked, and we are also investigating the provision of marking services for our printed formative assessments. This will allow teachers to tailor their teaching to support issues identified within assessments, rather than spending large amounts of time creating and marking assessments, not all of which may test the skills required.

We will link the data from formative and summative assessments that enable and encourage learning of individual building blocks of knowledge and empower teachers to teach more than just what might be tested in high stakes exams, so as to minimise 'teaching to the test'. This will give teachers information about student performance in prerequisite curriculum areas as an integral part of their lesson planning and preparation. As part of our efforts to minimise teaching to the test, we will only give supporting textbooks for our high stakes exams our endorsement if they are authored to teach beyond what might be expected in the exam. Furthermore, to improve the experience that students have of assessment, and with better use of data to inform assessment construction, we will introduce adaptive testing over time.



3. Free Access To Scripts: we will extend indefinitely our popular (free) access to scripts service and add analytical tools to support better teaching

In the summer 2017 exam series we made every exam script available to teachers, at no cost to schools. Over 400,000 student scripts were accessed from the summer exams series. We want teachers to have a greater understanding of how the mark scheme is applied and used in conjunction with resources, such as the Examiner's Report.

The next stage for this work will be to bring together the 'micro' script access with 'macro' data available on nationwide pupil performance on exam scripts and individual questions. We are finding ways to help teachers understand where their students performed well in exams, where they performed less well, and how to apply these insights to future teaching. Data can be a very useful tool for teachers if its purpose is to improve teaching, but teachers should never feel they are recording large volumes of data for its own sake.

We need to support teachers, and provide assessment opportunities; however, we appreciate that formative and summative assessments are different and our work with teachers needs to be clear about their forms and purposes.

4. Publish an Annual Monitoring Review on the reliability of assessments and comparability across subjects

Given the extent of our work globally, we produce and deliver internationally a very wide range of high quality assessments. We have carried out extensive international comparative studies to learn from best practice. We have conducted longitudinal randomised controlled trial research whereby we have observed students interacting with both our learning resources and formative assessments. Doing so has helped us to understand how our services impact student motivation, engagement, learning and, ultimately, outcomes.

The reliability of summative assessments is a key factor in ensuring confidence in the results that are issued. This is ever more important during a period of curriculum reform and the introduction of a new grading scale. Therefore we are carrying out research using knowledge and intelligence from the awarding of GCSE 9-1 Mathematics in summer 2017, for example, to understand comparability across the Foundation and Higher tiers for GCSE Science, French, Spanish and German in summer 2018.

As part of the programme of curriculum reform, we have committed to an annual Assessment Monitoring Review for all components of reformed qualifications. This annual review will use validity and reliability measures to continually improve our assessments over time, whilst ensuring that the tests remain true to the specification and do not represent a surprise for students and their teachers.

We will publish the Annual Monitoring Review, and unpack the meaning of it for teachers and their future teaching. We have already begun to do so following the first assessment of GCSE 9-1 Maths this summer, which has been well received.



Appendices

Appendix 1 – Phase 1 workshops

We ran the following workshops in collaboration with the following organisations.

Who was the partner organisation?	Who were the participants?	Where did it take place?
Ambition School Leadership	Middle leaders	Southwark, London
ASCL	Senior leaders	Birmingham
British Youth Council (at the Hounslow Youth Centre)	Young people, aged 13 to 19	Hounslow, London
Challenge Partners	Early Years teachers	Lambeth, London
Challenge Partners	Special school leaders	Islington, London
The Essex School Governors' Association	Governors and academy trustees	Chelmsford, Essex
NAHT	Secondary school senior leaders	Haywards Heath, Sussex
The National Association for Special Educational Needs (nasen)	Special school teachers	Webinar
NASUWT	Post-16 teachers	Birmingham
The National Governors Association	Governors and academy trustees	Birmingham
Pearson	Researchers, teachers and school and MAT leaders	City of London, London
Pearson	Researchers	City of Westminster, London
The Portsmouth Teaching School Alliance (through Challenge Partners)	Senior leaders	Portsmouth, Hampshire
Rescue Our Schools	Parents	City of Westminster, London
ResearchEd	Teachers and researchers	Hackney, London
SSAT	Senior leaders	Islington, London
Teach First	Class teachers	Bristol

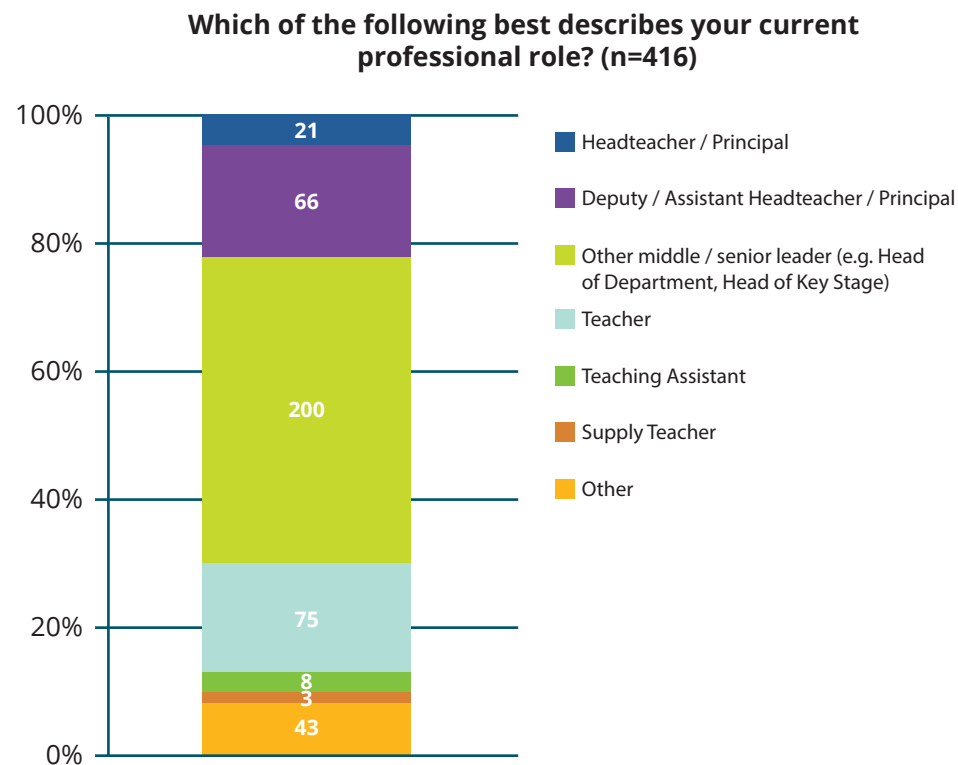
Appendix 2 – Online consultation responses

We asked respondents to our online consultation to tell us about themselves. The following chart gives a breakdown of their responses.

Question: What best describes you? (Please select all that apply):

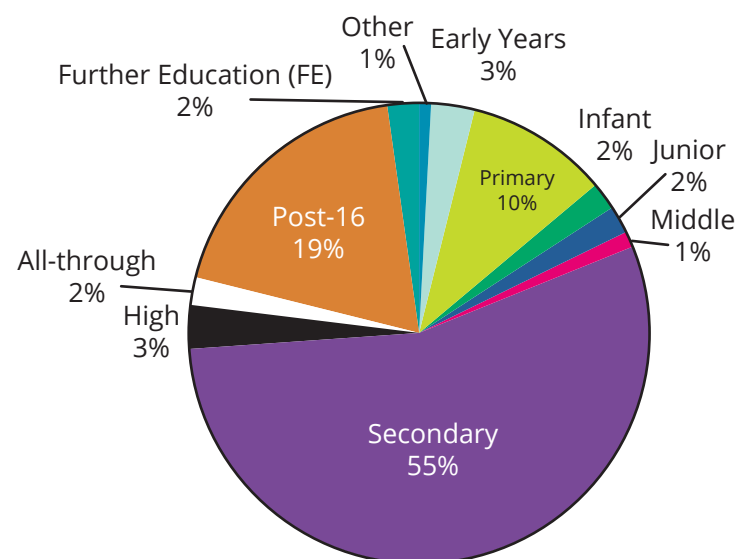
Response	Number
I work in a school	422
I am a school/academy governor/trustee	208
I am a parent or guardian of a child currently in school	186
I am a parent or guardian of a child, but they are not in school	39
I work for the government/Civil Service	13
I am a researcher/academic	29
I work for an organisation with a specific educational focus (e.g. special educational needs, literacy, sport)	39
Other	94

We asked school-based practitioners about their current roles:

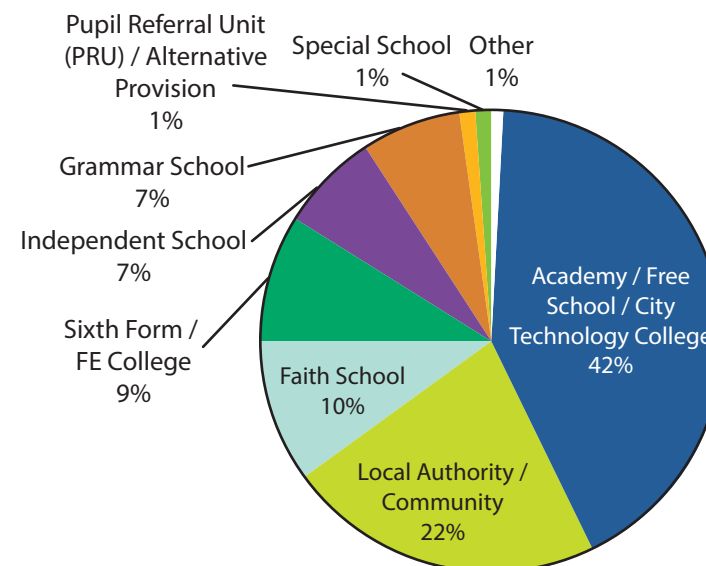


School-based practitioners were asked what phase and type of setting they work in.

What phase is your school / setting? (Select all that apply)



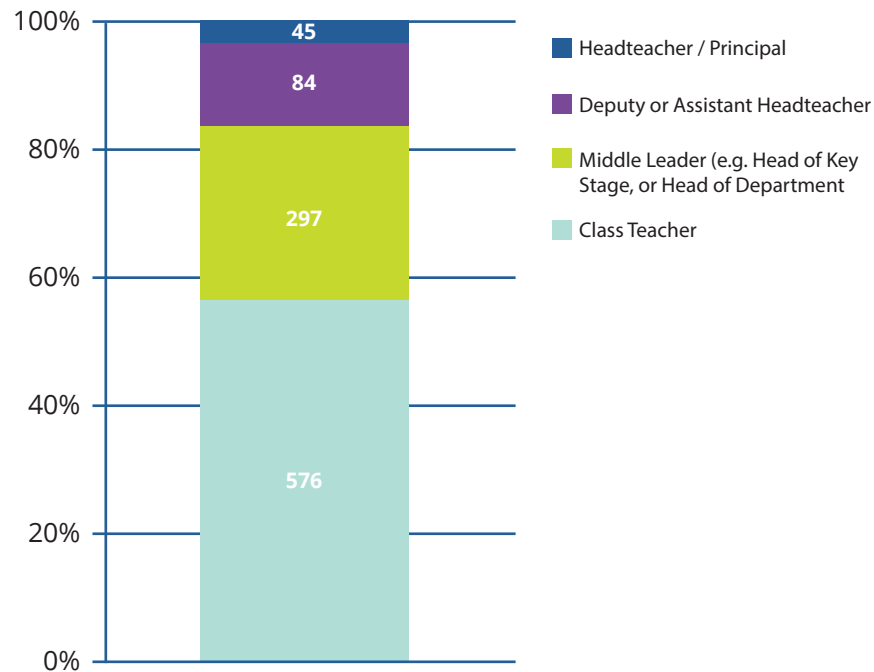
Which best describes the type of school you currently work in? (n=453)



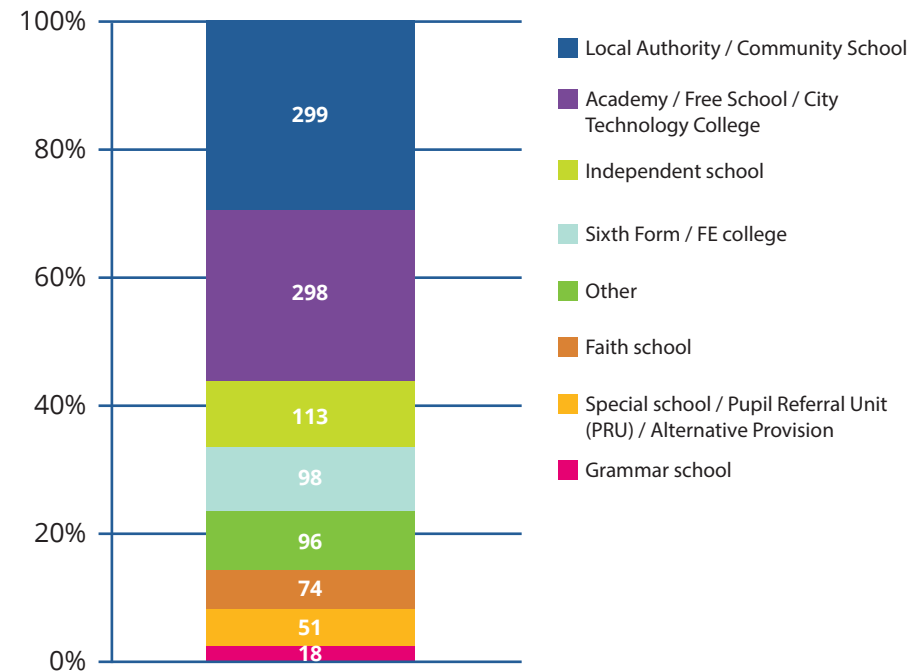
Appendix 3 – YouGov poll

We asked respondents to our poll a series of questions about themselves.

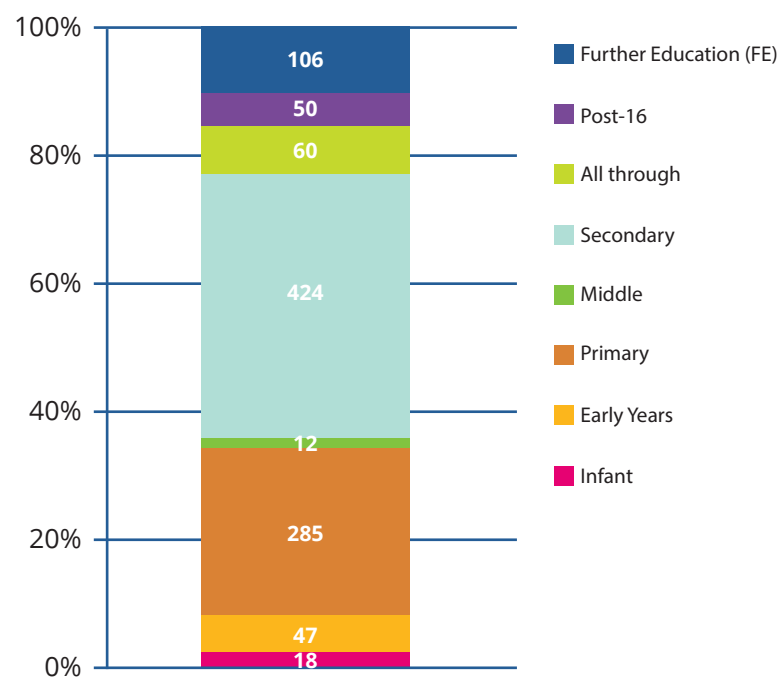
Which one of the following best describes your current professional role? (n=1,002)



Which, if any, of the following describe the type of school you currently work in? (Please select all that apply)



Which one of the following best describes the phase of education you work in?
(n=1,002)



Appendix 4 – Case studies

Case studies were conducted in settings across England, and in Finland, Japan, and Ontario, Canada.

Interviewee name	Interviewee role and organisation
Jasper Green	Head of Curriculum and Assessment, Ark Schools
Jack Deverson	Managing Director, Assessment Academy
Jamie Scott	Head of Partnerships, Assessment Academy
Simon Yates	Headteacher, Chailey Heritage Foundation, East Sussex
Julie Tilbury	Class teacher, Chailey Heritage Foundation, East Sussex
Chris Sangwin	Professor of Technology Enhanced Science Education, University of Edinburgh, and creator of STACK
Heidi Dennison	Deputy Headteacher, Frank Wise School, Oxfordshire
Garry Littlewood	Assistant Headteacher, Huntington School, York
Jane Elsworth	Assistant Director of Research School, York
Mr Iwamoto	Pearson's Country Manager for Japan and Korea
Mr Goto San	Pearson consultant, and education journalist
Mr Kobayashi	Vice Principal, Ryogoku High School, Sumida, Tokyo
Ms Oki	Class teacher, Ryogoku High School, Sumida, Tokyo
Mr Ohno	Class teacher, Kunitachi High School, Kunitachi
Mr Inoue	Class teacher, Kunitachi Junior High School, Kunitachi
Mr Yoshino	Principal, Ohyu Gakuen Girls High School
Mr Hirose	Director, Hosei University, Tokyo
Mr Kondo	Director, Hosei University, Tokyo
Mr Sato	Admissions Department, Hosei University, Tokyo
Mr Yoshida	Principal, Senzoku Gakuen Elementary School, Kawasaki

Interviewee name	Interviewee role and organisation
Richard Auffret Pupil focus group	Director of Curriculum and Technology, Chesterton College, Cambridgeshire Educational Trust
Minna Welin Pekka Peura Sonny Johnson Mikaela Sumeli	Vice Principal, Saunalahti School, Espoo, Finland Principal, StartUp High School, class teacher, and education blogger, Finland Elementary school teacher, Finland and the UK Elementary school teacher, Helsinki, Finland
Sarah Wild Pupil focus group	Headteacher, Limpsfield Grange School, Surrey
Sabina Iqbal Kayleigh Simmonds Adam Bagherian	SENCo, Margaret McMillan Primary School, Bradford Class teacher and literacy coordinator, Margaret McMillan Primary School, Bradford Early Years, Reception and Year 1 Phase Leader, Margaret McMillan Primary School, Bradford
Katie Ashford	Deputy Headteacher, Michaela Community School, Brent
Andrea Gillespie Jenn Clark Karen Dobbie Mark Cobham Beverley Buxton	Superintendent of Learning, Trillium Lakelands District School Board, Ontario School Administrator, Algonquin Park, Ontario Student Achievement Office, Ministry of Education, Ontario Pearson VP for K-12 Product Development, Ontario Pearson General Manager, Pearson Canada School Division
Sacha Nelson Natalie Gregory Pupil focus group	Assistant Headteacher, Parkwood Primary School, Bradford Class teacher, Parkwood Primary School, Bradford
Richard Slade Dave Witham Annette Green Julia Holmes	Headteacher, Plumcroft Primary School, Greenwich Upper Key Stage 2 Phase Leader, Plumcroft Primary School, Greenwich Lower Key Stage 2 Phase Leader, Plumcroft Primary School, Greenwich Year 2 Leader, Plumcroft Primary School, Greenwich

Interviewee name	Interviewee role and organisation
Rob Webster	Director, Maximising the Impact of Teaching Assistants (MITA) Project, at the UCL IOE
Hannah Barnett	Senior Programme Officer, School 21
Clare Sealy	Headteacher, St Matthias Church of England Primary School, Tower Hamlets
Harminder Dhanjal	Class teacher, St Matthias Church of England Primary School, Tower Hamlets
Wade Nottingham	Assistant Headteacher, West London Free School, Hammersmith and Fulham
Ruth Fonseka-McFarlane	Head of Biology, West London Free School, Hammersmith and Fulham
Oliver Bell	History teacher, West London Free School, Hammersmith and Fulham
Natasha Crook	Head of Classics, West London Free School, Hammersmith and Fulham
Romy Bartram	Geography teacher, West London Free School, Hammersmith and Fulham
Pupil focus group	

References

- ⁱ Pearson (2016) *Assessments Attitudes of UK Teachers and Parents*. In June 2016 Pearson commissioned BritainThinks to conduct a survey of 506 teachers, headteachers and parents. The random sample had good spread across the country and comprised: 256 parents, all with children aged between 4 and 18; 200 fulltime teachers, and; 50 headteachers.
- ⁱⁱ Millard, W. and Menzies, L. (2017) *Testing the Water: Exploring the Role of Assessment in Teaching – Interim Report*, London: Pearson/LKMco.
- ⁱⁱⁱ DeLuca, C. and Johnson, S. (2017) 'Developing assessment capable teachers in this age of accountability', *Assessment in Education: Principles, Policy & Practice*, 24:2, 121-126.
- ^{iv} Hill, P. (2016) *Stimulus Paper* (internal document for Pearson); Pearson (2016).
- ^v Wiliam, D. (2017) 'Evaluating education and improving it', *Impact: Assessment in practice – Journal of the Chartered College of Teaching*, 1.
- ^{vi} Wiliam, D. (2011) 'What assessment can – and cannot do', *Pedagogiska Magasinet*, available at: https://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/Pedagogiska%20magasinet%20article.docx.
- ^{vii} Millard, W. and Menzies, L. (2017).
- ^{viii} Department for Education (2010) *The Important of Teaching*, London: TSO, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175429/CM-7980.pdf.
- ^{ix} Lord Bew (2011) *Independent Review of Key Stage 2 testing, assessment and accountability – Final Report*, available at: www.educationengland.org.uk/documents/pdfs/2011-bew-report-ks2tests.pdf.
- ^x Ofqual (2014) *Review of Quality of Marking in Exams in A Levels, GCSEs and Other Academic Qualifications – Final Report*, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/393832/2014-02-14-review-of-quality-of-marking-in-exams-in-a-levels-gcses-and-other-academic-qualifications-final-report.pdf.
- ^{xi} McIntosh, J. (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.
- ^{xii} Department for Education (2017) *Reducing teacher workload*, available at: <https://www.gov.uk/government/publications/reducing-teachers-workload/reducing-teachers-workload>.
- ^{xiii} Standards and Testing Agency (2016) *The Rochford Review: final report*, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/561411/Rochford_Review_Report_v5_PFDA.pdf.
- ^{xiv} Department for Education (2016) *Standards and Testing Agency Review: Final report*, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/566914/STA_review_-_Final_report_FINAL_.pdf.
- ^{xv} Ofsted (2016) *The Annual Report of Her Majesty's Chief Inspector of Education, Children's Services and Skills 2015/16*, London: Ofsted, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/574186/Ofsted_annual_report_education_and_skills_201516_web-ready.pdf.
- ^{xvi} Commons Select Committee (2017) 'High-stakes' testing harming teaching and learning in primary schools, available at: <https://www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/news-parliament-2015/primary-assessment-report-published-16-17/http://www.parliament.uk/business/committees/committees-a-z/commons-select/education-committee/news-parliament-2015/primary-assessment-report-published-16-17/>.
- ^{xvii} Pearson (2017) *Understanding GCSE 9-1 marks and grades*, available at: <https://qualifications.pearson.com/en/qualifications/edexcel-gcses/understanding-9-1.html>.
- ^{xviii} Ofqual (2017) *National Reference Test 2018*, available at: <https://www.gov.uk/government/news/national-reference-test-2018>; NFER (2018), available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/645287/6279_-_NRT2018_Information_for_Schools.pdf.

- ^{xix} Department for Education (2017) *Primary assessment in England: Government consultation response, September 2017*, London: Department for Education, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/644871/Primary_assessment_consultation_response.pdf.
- ^{xx} Department for Education (2017) *Primary school pupil assessment: Rochford Review Recommendations – Government consultation response, September 2017*, London: Department for Education, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/644729/Rochford_consultation_response.pdf.
- ^{xxi} Ofsted (2017) *Ofsted strategy 2017-22*, Manchester: Ofsted, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/648212/Ofsted_strategy_2017-22.pdf.
- ^{xxii} Spielman, A. (2017) *HMC's commentary: recent primary and secondary curriculum research*, available at: <https://www.gov.uk/government/speeches/hmcis-commentary-october-2017>.
- ^{xxiii} Teacher Development Trust (2017) *How Does Your School's Spending Compare?*, available at: <http://tdtrust.org/benchmarking>.
- ^{xxiv} Sellen, P. (2016) *Teacher workload and professional development in England's secondary schools: insights from TALIS*, London: Education Policy Institute, available at: https://epi.org.uk/wp-content/uploads/2016/10/TeacherWorkload_EPI.pdf.
- ^{xxv} Sellen, P. (2016).
- ^{xxvi} Teacher Development Trust (2014) *Developing Great Teaching: Lessons from the international reviews into effective professional development*, London: Teacher Development Trust, available at: <http://tdtrust.org/wp-content/uploads/2015/10/DGT-Summary.pdf>.
- ^{xxvii} DeLuca, C. and Johnson, S. (2017) 'Developing assessment capable teachers in this age of accountability', *Assessment in Education: Principles, Policy & Practice*, 24:2, 121-126; Popham, W. J. (2009) 'Assessment literacy for teachers: Faddish or fundamental?', *Theory Into Practice*, 48, 4-11; Xu, Y., and Brown, G. (2016) 'Teacher assessment literacy in practice: A reconceptualization', *Teaching and Teacher Education*, 58, 149-162; Coe, R., Aloisi, C., Higgins, S. and Elliot Major, L. (2014) *What makes great teaching?: Review of the underpinning research*, London: Sutton Trust, available at: <https://www.suttontrust.com/wp-content/uploads/2014/10/What-makes-great-teaching-FINAL-4.11.14-1.pdf>; Dunlosky, J., Rawson, K. A., Marsh, E. J., Nathan, M. J., and Willingham, D. T. (2013) 'Improving Students' Learning with Effective Learning Techniques: Promising Directions From Cognitive and Educational Psychology', *Psychological Science in the Public Interest*, 14:1, 4-58; Kersting, N. B., Givvin, K. B., Sotelo, F. L., and Stigler, J. W. (2010) 'Teachers' Analyses of Classroom Video Predict Student Learning of Mathematics: Further Explorations of a Novel Measure of Teacher Knowledge', *Journal of Teacher Education*, 61:1-2, 172-181.
- ^{xxviii} NAHT (2014) Report of the NAHT Commission on Assessment, Haywards Heath: NAHT, p. 17, available at: <http://www.naht.org.uk/assets/assessment-commission-report.pdf>.
- ^{xxix} Straw, S., Tattersall, J. and Sims, D. (2016) *NFER Teacher voice omnibus: Research report*, London: Department for Education, p. 8, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/535269/DFE-RR532-Teacher-voice-omnibus-November-2015-responses.pdf.
- ^{xxx} Department for Education (2015) *Carter review of initial teacher training*, London: Department for Education, p. 9, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/399957/Carter_Review.pdf.
- ^{xxxi} Diagnostic Questions (2017), available at: <https://diagnosticquestions.com/>.
- ^{xxxii} Rising Stars (2017) *Assessment Bank*, available at: https://www.risingstars-uk.com/Series/RS-Assessment-Bank?utm_source=AB_DM_BROCHURE_917&utm_medium=direct_mail&utm_campaign=AB_DM_BROCHURE_917; Pearson (2017) Pearson Formative Item Bank, available at: <https://www.pearsonassessments.com/learningassessments/products/100001102/pearson-formative-item-bank.html#tab-details>; Cambridge Assessment (2017) Past exam material, available at: <http://www.cambridgeassessment.org.uk/our-research/archives-service/past-exam-material/>.
- ^{xxxiii} Menzies, L. (2015) *The future for examining*, available at: <https://cdn.lkmco.org/wp-content/uploads/2015/12/The-Future-for-Examining-Final.pdf>.
- ^{xxxiv} For example, see: Pearson (2016).
- ^{xxxv} Standards and Testing Agency (2017) *Information for parents: 2017 national curriculum tests at the end of key stages 1 and 2*, available at: <https://www.gov.uk/government/uploads/system/uploads/>

- attachment_data/file/603590/Information_for_parents_-_2017_NCTs_at_the_end_of_key_stages_1_and_2_v4_PDFA.pdf.
- xxxvi Millard, W. and Menzies, L. (2017); Pearson (2016).
- xxxvii Christodoulou, D. (2016) Making Good Progress? *The future of Assessment for Learning*, Oxford: Oxford University Press; Wiliam, D. and Black, P. (1996) 'Meanings and Consequences: a basis for distinguishing formative and summative functions of assessment?', *British Educational Research Journal*, 22:5, pp. 537-548.
- xxxviii Extensive research explores this. Christodoulou (2016) provides an overview. Sweller examines the issue in relation to problem solving in: Sweller, J. (1988) 'Cognitive Load During Problem Solving: Effects on Learning', *Cognitive Science*, 12, 257-285;
- xxxix Fordham, M. (2017) *Decoupling summative and formative feedback*, Clio et cetera website, available at: <https://clioetcetera.com/2017/11/02/decoupling-summative-and-formative-assessment/>.
- xl Straw, S., Tattersall, J. and Sims, D. (2016).
- xli Millard, W. and Menzies, L. (2016) *Oracy: The State of Speaking in Our Schools*, London: LKMco / Voice 21, available at: https://docs.wixstatic.com/ugd/2c80ff_c167ea8db0c845b887113b1c8a62ebef.pdf.
- xlvi Ofsted (2015) *Key Stage 3: the wasted years?*, Manchester: Ofsted, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/459830/Key_Stage_3_the_wasted_years.pdf.
- xliv Lubinski, D. and Humphreys, L. G. (1997) 'Incorporating general intelligence into epidemiology and the social sciences', *Intelligence*, 24:1, 159-202; Hirsch, E. D. (2001) 'Seeking Breadth and Depth in the Curriculum', *Educational Leadership*, 59:2, 22-25, available at: http://www.ascd.org/ASCD/pdf/journals/ed_lead/el200110_hirsch.pdf.
- xliv These are available commercially, and include resources from GL Assessment, the NFER, CEM at the University of Durham, Renaissance Learning, and Rising Stars.
- xliv Straw, S., Tattersall, J. and Sims, D. (2016), p. 39.
- xlvi Pearson (2016).
- xlvi Christodoulou, D. (2016) pp. 60-76.
- xlvi Ladd, H. F. and Zelli, A. (2002) 'School-Based Accountability in North Carolina: The Responses of School Principals', *Educational Administration Quarterly*, 38: 4, 494-529; Stecher, B. and Chun, T. (2001) *School and Classroom Practices During Two Years of Education Reform in Washington State - CSE Technical Report 550*, Los Angeles, CA: National Centre for Research on Evaluation, Standards, and Student Testing, available at: https://www.researchgate.net/profile/Brian_Stecher/publication/234658739_School_and_Classroom_Practices_during_Two_Years_of_Education_Reform_in_Washington_State_CSE_Technical_Report/links/0deec53a8403f586b1000000/School-and-Classroom-Practices-during-Two-Years-of-Education-Reform-in-Washington-State-CSE-Technical-Report.pdf.
- xlvi Spielman, A. (2017).
- l Koretz, D. (2008) *Measuring Up: What Educational Testing Really Tells Us*, Cambridge, MA: Harvard University Press, p. 158.
- li For example, Wiliam, D. (2009) *The reliability of educational assessments*, Ofqual Annual Lecture, 07.05.2009, available at: https://www.dylanwiliam.org/Dylan_Wiliams_website/Presentations_files/Ofqual%20talk.ppt; Wiliam, D. (2001) 'Reliability, validity, and all that jazz', *Education*, 3, 17-21, available at: <http://eprints.ioe.ac.uk/1156/2/Wiliam2001Reliability3long.pdf>.
- lii Wiliam, D. (2001).
- lii OECD (2016) *PISA 2015 Results (Volume I): Excellence and Equity in Education*, Paris: PISA, OECD Publishing.
- lii Crehan, L. (2016) *Cleverlands*, London: Unbound; OECD (2011) *Lessons from PISA for the United States, Strong Performers and Successful Reformers in Education*, Paris: PISA, OECD Publishing.
- lii Crehan, L. (2016).
- lii Treadaway, M. (2017) *Don't try to forecast Progress 8!*, Education Datalab blog, available at: <https://educationdatalab.org.uk/2016/02/dont-try-to-forecast-progress-8/>.
- lii The Headteachers' Roundtable (2016) *The Alternative Green Paper - Schools That Enable All To Thrive and Flourish*, available at: <https://www.schoolsweek.co.uk/wp-content/uploads/2016/09/The-Alternative-Green-Paper.pdf>; Green, S. and Oates, T. (2007) *How to promote educational quality through national assessment systems*, a paper presented at the International Association for Educational Assessment Annual Conference,

- Baku, Azerbaijan, September 2007, available at: <http://www.cambridgeassessment.org.uk/Images/109749-how-to-promote-educational-quality-through-national-assessment-systems.pdf>.
- lviii Torrance, H. (2017) 'The Return to Final Paper Examining in English National Curriculum Assessment and School Examinations: Issues of Validity, Accountability and Politics', *British Journal of Educational Studies*, 1-25.
- lix NAHT (2017) *Report of the Assessment Review Group: Redressing the Balance*, Haywards Heath: NAHT, available at: http://www.naht.org.uk/_resources/assets/attachment/full/0/65197.pdf.
- lx Brooks, R. and Tough, S. (2006) *Assessment and Testing: Making space for teaching and learning*, London: IPPR, available at: https://www.ippr.org/files/images/media/files/publication/2011/05/ASSESSMENT_1554.pdf.
- lxi OECD (2016) p. 231.
- lxii Cambridge Assessment (2017); OECD (2016).
- lxiii Hill, P. (2016) *Stimulus Paper* (internal document for Pearson); Crehan, L. (2016); Acquah, D. (2013) *Examining school accountability*, Centre for Education Research and Practice and AQA, available at: https://cerp.aqa.org.uk/sites/default/files/pdf_uploads/CERP_IP_DA_06062013.pdf.
- lxiv Coe, R. and Sahlgren, G. H. (2014) 'Incentives and Ignorance in Qualifications, Assessment, and Accountability', in G. H. Sahlgren (ed.) *Tests worth teaching to: Incentivising quality in qualifications and accountability*, London: The Centre for Market Reform of Education, pp. 74-95.
- lxv Allen, R. (2015) *Is an Ofsted judgement a lagging or leading indicator of school performance?*, Education Datalab, available at: <https://educationdatalab.org.uk/2015/11/is-an-ofsted-judgement-a-lagging-or-leading-indicator-of-school-performance/>.
- lxvi Crawford, C. and Benton, T. (2017) *Volatility happens: Understanding variation in schools' GCSE results – Research Report*, Cambridge: Cambridge Assessment, available at: <http://www.cambridgeassessment.org.uk/Images/372751-volatility-happens-understanding-variation-in-schools-gcse-results.pdf>; Allen, R. (2016) *Submission to the House of Commons Education Select Committee Inquiry on Primary Assessment*, available at: <https://educationdatalab.org.uk/wp-content/uploads/2016/12/Primary-assessment-Education-Datalab-submission.pdf>; Department for Education (2016) SFR 39/2016 - *National curriculum assessments at Key Stage 2 in England*, 2016 (provisional), available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/549432/SFR39_2016_text.pdf; Kane, T. J. and Staiger, D. O. (2002) 'The Promise and Pitfalls of Using Imprecise School Accountability Measures', *Journal of Economic Perspectives*, 16: 4, 91-114.
- lxvii For example, see Thomson, D. (2015) *Significance tests for school performance indicators: Why they're OK really (but we should probably call them something else)*, Education Datalab Blog, available at: <https://educationdatalab.org.uk/2015/07/significance-tests-for-school-performance-indicators-why-theyre-ok-really-but-we-should-probably-call-them-something-else/>.
- lxviii Allen, B. (2017) *Ofsted inspections are unreliable by design*, Schools Week, available at: <https://schoolsweek.co.uk/ofsted-inspections-are-unreliable-by-design/>.
- lxix Menzies, L. (2017) *The end of 'Outstanding': Why we no longer need Ofsted's gold star*, LKMco blog, available at: <https://www.lkmco.org/end-ofsted-outstanding-no-longer-need-ofsteds-gold-star/>; Ofsted (2017) 'Do two inspectors inspecting the same school make consistent decisions?' A study of the reliability of Ofsted's new short inspections, Manchester: Ofsted, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/596708/Reliability_study_-_final.pdf.
- lxx Coe, R. (2007) *Changes in standards at GCSE and A-level: Evidence from ALIS and YELLIS*, available at: <http://www.cem.org/attachments/ONS%20report%20on%20changes%20at%20GCSE%20and%20A-level.pdf>; Tymms, P. and Morrell, C. (2007) 'Standards and Quality in English Primary Schools Over Time: The National Evidence' (Research Survey 4/1), *The Primary Review*, Cambridge: University of Cambridge Faculty of Education.
- lxxi Oates, T. (2014) 'The 'Qualifications Sledgehammer': Why Assessment-Led Reform Has Dominated the Education Landscape', in G. H. Sahlgren (ed.) *Tests worth teaching to: Incentivising quality in qualifications and accountability*, London: The Centre for Market Reform of Education, pp. 28-43.
- lxxii Oates, T. (2013) *Radical solutions in demanding times: alternative approaches for appropriate placing of 'coursework components' in GCSE examinations*,

- Cambridge: Cambridge Assessment, available at: <http://www.cambridgeassessment.org.uk/Images/145830-coursework-radical-solutions-in-demanding-times-.pdf>.
- lxxiii This is well documented in the literature. For example, see: Campbell, T. (2015) 'Stereotyped at seven: Biases in teachers' judgements of pupils' ability and attainment', *Journal of Social Policy*, 44:3, 517-547; Zavodny, M. (2013) 'Does weight affect children's test scores and teacher assessments directly?', *Economics of Education Review*, 34, 135-145; Burgess, S. and Greaves, E. (2013) 'Test scores, subjective assessment, and stereotyping of ethnic minorities', *Journal of Labor Economics*, 31:3, 535-576; Ritts, V., Patterson, M. L., and Tubbs, M. E. (1992) 'Expectations, Impressions, and Judgements of Physically Attractive Students: A Review', *Review of Educational Research*, 62:4, 413-426.
- lxxiv Allen, R. (2016); Christodoulou, D. (2016).
- lxxv Wiliam, D. (2011) *Assessing achievement at the end of Key Stage 2*, submission to Lord Bew's Independent Review of Key Stage 2 testing, assessment and accountability, available at: https://webcache.googleusercontent.com/search?q=cache:n8g38_reaXYJ:https://www.dylanwiliam.org/Dylan_Wiliams_website/Papers_files/Evidence%2520to%2520Bew%2520Review.doc+&cd=1&hl=en&ct=clnk&gl=uk.
- lxxvi Torrance, H. (2017).
- lxxvii See, for example: Brown, G. T. L., Lake, R. and Matters, G. (2011) 'Queensland Teachers' Conceptions of Assessment: The Impact of Policy Priorities of Teacher Attitudes', *Teaching and Teacher Education*, 27:1, 210-220.
- lxxviii Finnish National Agency for Education (2017) *Education system – Equal opportunities to high quality education*, available at: http://www.opf.fi/english/education_system.
- lxxix Oates, T. (2015) *Finnish Fairy Stories*, Cambridge: Cambridge Assessment, available at: <http://www.cambridgeassessment.org.uk/Images/207376-finnish-fairy-stories-tim-oates.pdf>.
- lxxx Heller Sahlgren, G. (2015) *Real Finnish Lessons: The true story of an education superpower*, London: Centre for Policy Studies.
- lxxxi Coe, R. and Sahlgren, G. H. (2014).
- lxxxii Coe, R. (2007); Tymms, P. and Morrell, C. (2007) 'Standards and Quality in English Primary Schools Over Time: The National Evidence' (Research Survey 4/1), *The Primary Review*, Cambridge: University of Cambridge Faculty of Education.
- lxxxiii Tymms, P. and Morrell, C. (2007).
- lxxxiv See Commons Select Committee (2017); NAHT (2017); The Headteachers' Roundtable (2016).
- lxxxv Wiliam, D. (2011).
- lxxxvi Cambridge Assessment (2017) *A Cambridge Approach to Improving Education: Using international insights to manage complexity*, available at: <http://www.cambridgeassessment.org.uk/Images/cambridge-approach-to-improving-education.pdf>; PISA OECD 2015 (2016) *PISA 2015 Results: Policies and Practices for Successful Schools, Volume II*, OECD Publishing: Paris, available at: <http://www.oecd-ilibrary.org/docserver/download/9816071e>. Schmidt, W. H. and Prawat, R. S. (2006); Schmidt, W. H. and Prawat, R. S. (2006) 'Curriculum coherence and national control of education: issue or non-issue?', *Journal of Curriculum Studies*, 38:6, 641-658.
- lxxxvii Department for Education (2012) *Statistical First Release – Phonics Screening Check and National Curriculum Assessments At Key Stage 1 in England, 2011/12 – SFR 21/2012*, p. 4, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/219208/main_20text_20_20sfr21-2012.pdf
- lxxxviii Adams, R. (2014) *Did primary school teachers in England game the phonics check?*, article in The Guardian, 26.11.14, available at: <https://www.theguardian.com/education/datablog/2014/sep/26/primary-school-teachers-game-phonics-check>.
- lxxxix Department for Education (2017) *Primary school accountability in 2017*, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/641197/Primary_school_accountability_in_2017.pdf
- xc Standards and Testing Agency (2016) *Key Stage 1 2017 Assessment and Reporting Arrangements 2017*, https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/561720/2017_KS1_ARA_v1.0.pdf
- xc Ofqual (2015) *Our announcement on speaking and listening assessments*, The Ofqual Blog, available at: <https://ofqual.blog.gov.uk/2013/09/04/our-announcement-on-speaking-and-listening-assessments/>.
- xcii OECD (2016); Hill, P. (2016).
- xciii OECD (2016) (p. 133).

- ^{xciv} Ofsted (2016) *Ofsted inspection: myths*, available at: <https://www.gov.uk/government/publications/school-inspection-handbook-from-september-2015/ofsted-inspections-mythbusting>.
- ^{xcv} For example, see: Ofsted (2017) *Ofsted – developments in education inspection*, available at: <https://educationinspection.blog.gov.uk/author/sean-harford/>.
- ^{xcvi} Straw, S., Tattersall, J. and Sims, D. (2016).
- ^{xcvii} Wiliam, D. (2015) 'Feedback for Learning: Make Time to Save Time', *Dylan Wiliam Centre*, available at: <http://www.dylanwiliamcenter.com/feedback-for-learning-make-time-to-save-time/>.
- ^{xcviii} Department for Education (2016) *Eliminating unnecessary workload around marking: Report of the Independent Teacher Workload Review Group*, London: Department for Education, available at: [https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/511256/](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/511256/Eliminating-unnecessary-workload-around-marking.pdf) Eliminating-unnecessary-workload-around-marking.pdf; Sellen, P. (2016).
- ^{xcix} Elliott, V., Baird, J., Hopfenbeck, T. N., Ingram, J., Thompson, I., Usher, N., Zantout, M., Richardson, J. and Coleman, R. (2016) *A marked improvement? A review of the evidence of written marking*, London: Education Endowment Foundation, available at: https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Marking_Review_April_2016.pdf.
- ^c Elliott, V. et al. (2016).
- ^{ci} Straw, S., Tattersall, J. and Sims, D. (2016); Quigley, A. (2017) 'The Problem with Past Exam Papers', *The Confident Teacher (blog)*, available at: <https://www.theconfidentteacher.com/2017/04/the-problem-with-past-exam-papers/>.
- ^{cii} Dunlosky, J., et al. (2013); Roediger, H. L., Putnam, A. L. and Smith, M. A. (2011) 'Ten Benefits of Testing and Their Application to Educational Practice', *Psychology and Learning and Motivation*, 55, 1-36.
- ^{ciii} Department for Education (2016).
- ^{civ} Roberts, J. (2016), *Can technology genuinely reduce teacher workload?* (Datchet: Advanced Learning).
- ^{cv} Haswell, R., Donnelly, W., Hester, V., O'Neill, P., and Schendel, E. (2012) 'An Annotated Bibliography of Writing Assessment: Machine Scoring and Evaluation of Essay-length Writing', *The Journal of Writing Assessment*, 5: 1; Tisi, J., Whitehouse, G., Maughan, S. and Burdett, N. (2013) *A review of literature on marking reliability research: Report for Ofqual*, Slough: NFER.
- ^{cvi} BESA (2017) *Schools highlight urgent need for teacher CPD in EdTech in major BESA report*, available at: <https://www.besa.org.uk/news/schools-highlight-urgent-need-teacher-cpd-edtech-major-besa-report/>; Bennett, T. (2017), *Creating a Culture: How school leaders can optimise behaviour – Independent review of behaviour in schools* (London: Department for Education, 2017); Roberts, J. (2016); Bennett, T. (2015), *Anti-socially mobile? What I really think about classroom tech*, *TES*, 31.08.15, available at: <https://www.tes.com/news/blog/anti-socially-mobile-what-i-really-think-about-classroom-tech>.
- ^{cvi} Wolf, A. (2011) *Review of Vocational Education – The Wolf Report*, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/180504/DFE-00031-2011.pdf.
- ^{cvi} Department for Education (2017) *Trends in arts subjects in schools where English Baccalaureate entry has increased*, London: Department for Education, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/630223/Trends_in_arts_subjects_in_schools_where_English_Baccalaureate_entry_has_increased.pdf; Allen, R. and Thompson, D. (2016) 'Changing the subject: How are the EBacc and Attainment 8 reforms changing results?', Sutton Trust Research Brief: Edition 13, available at: https://educationdatalab.org.uk/wp-content/uploads/2016/07/Changing-the-subject_FINAL.pdf.
- ^{cix} Allen and Thompson, 2016.
- ^{cx} See Department for Education (2017) *Secondary accountability measures*, available at: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/651158/Secondary_accountability_measures-Guide.pdf.
- ^{cx} The Key (2017) *State of Education Survey Report 2017*, London: The Key, available at: <https://stateofed.thekeysupport.com/#intro>.
- ^{cxii} Pearson (2016).

“Society should ensure that all young people receive the support they need in order to make a fulfilling transition to adulthood.”



This report was written by the education and youth development ‘think and action tank’ LKMco. LKMco is a social enterprise – we believe that society has a duty to ensure children and young people receive the support they need in order to make a fulfilling transition to adulthood.

We work towards this vision by helping education and youth organisations develop, evaluate and improve their work with young people. We then carry out academic and policy research and advocacy that is grounded in our experience.

www.lkmco.org.uk / [@LKMco](https://twitter.com/LKMco) / info@lkmco.org