



Sense and Accountability: Holding our primary schools to account for what matters most

Final report of the ASCL Primary Accountability Review
February 2018

Contents

Executive summary	4
Recommendations	5
Section 1: Introduction	6
Section 2: How are primary schools in England currently held to account?	7
Section 3: What alternative approaches are there?	8
Section 4: Measuring what matters	9
Section 5: What should an effective and fair accountability system look like?	10
Section 6: How does the current system measure up against these principles?	11
Section 7: Conclusion	22
Acknowledgements	23

Executive summary

Many of the negative effects of assessment are in fact caused by the use of results in the accountability system rather than the assessment system itself.

So stated the Education Select Committee in its hard-hitting report on primary assessment last year. The committee identified a range of problems with the way in which primary pupils are assessed, including:

- the rushed way in which the new assessments were implemented
- an excessive focus on specific grammatical techniques in the assessment of writing
- insufficient training and support for teachers in implementing assessment without levels

However, it was the way in which these assessments are then used to hold schools to account that particularly exercised the committee. The impact of this, according to witnesses providing evidence to the committee, includes:

- a narrowing of the curriculum experienced by primary pupils
- excessive pressure on both children and teachers
- incentives to 'game' the system

The role of accountability in primary education in England, it seems, deserves further scrutiny.

ASCL's ***Blueprint for a Self-Improving School System*** (Blueprint), our 2015 inquiry into the leadership of England's education system, set out a vision for accountability in a self-improving system as follows:

Accountability is the obligation of an individual and organisation to account for its activities, accept responsibility for them, and to disclose the results in a transparent manner. The highest form of accountability is the individual's professional accountability for the quality of his or her own work and to the people who the profession serves. In a self-improving system, we believe that teachers and school leaders are agents of their own accountability.

The Blueprint outlined the role of government in accountability as "*defining a slim, smart and stable public accountability framework with a small number of ambitious goals*". This framework should "*incentivise schools, trusts and federations to implement policies and behaviours that contribute to a high-quality education for all*".

This report takes this definition of accountability as its starting point. It focuses mainly on what a "*slim, smart and stable*" (and fair and effective) public accountability

framework might look like in a primary context, proposing a set of principles for such a framework. It also begins to explore how school leaders and teachers can be encouraged and supported to become "*agents of their own accountability*".

We propose seven principles of an effective and fair accountability system, and hold the approach in England up against these principles.

Principles

An effective and fair accountability system should:

- 1 start from a shared understanding of what outcomes we, as a society, want for our children and young people
- 2 be based around a set of measures which incentivise schools to deliver on these outcomes, seeking ways to recognise and reward aspects which are important but difficult to measure, as well as those which are more easily quantifiable
- 3 drive positive behaviour
- 4 be based on information which is as accurate as possible, and not try to read too much into a small, unrepresentative amount of data
- 5 be fair to schools in different circumstances and contexts, while recognising the importance of enabling every child to reach their potential
- 6 lead to fair, proportionate, transparent and constructive consequences for schools which fall short of its desired outcomes, aligned with the best current evidence of what is most likely to lead to improvements
- 7 be relentlessly self-critical, regularly evaluating impact and modifying as necessary

We make 15 recommendations. Some are for government, some for Ofsted, and some for school leaders and leadership organisations.

Implementing these recommendations would, we believe, take us closer to the fair and effective accountability system we need, and which our pupils, parents, teachers, and school leaders deserve.

Recommendations

Holding schools to account for a broader range of measures

- 1 Government should work with a broad range of stakeholders to co-construct a clear set of aims for primary education and a shared, long-term view of the desired outcomes for children as a result of that education.
- 2 Both government and Ofsted should engage with research around different ways to measure outcomes which are important but less tangible, and consider how these might be incorporated into the accountability system.
- 3 Ofsted should commit to commenting more frequently in its inspection reports on subjects other than English and maths, and ensure its focus on published performance figures is proportionate.

Improving the accuracy of the current accountability measures

- 4 The Standards and Testing Agency (STA) should seriously consider how writing might be assessed more reliably as part of the accountability system. If this is not possible, writing should be excluded from the performance measures.

Promoting ethical leadership and effective curriculum design

- 5 School leaders and teachers should engage closely with educational research, particularly around effective curriculum design. They should be confident that their school curriculum is informed by the best available evidence on how to enable children to succeed, both against the current performance measures and in the broader outcomes they value.
- 6 Providers of school leadership programmes should ensure they include a strong focus on both ethical leadership and curriculum design. School leaders should ensure teachers in their schools are given the encouragement and opportunity to develop their own expertise in these areas.

Using performance measures in a proportionate way

- 7 The primary school performance tables should be based on data from a three-year rolling period, rather than on results from a single year's assessments.
- 8 School leaders and governors should ensure they understand their school's performance data intimately, and that they are sufficiently skilled in analysing and interpreting data more generally. School leadership

organisations, and providers of leadership programmes, should support them more effectively with this.

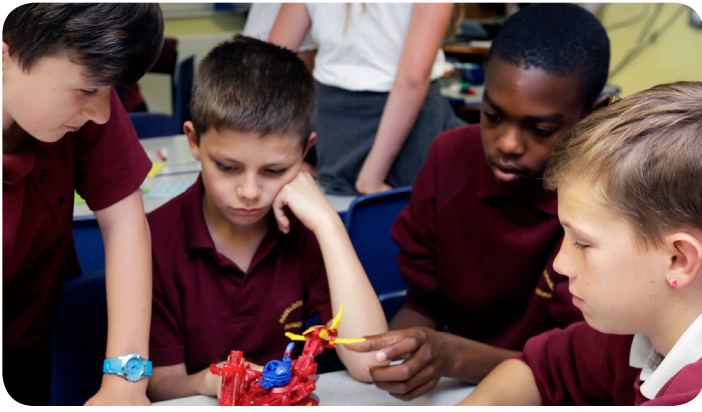
- 9 The government must be cautious about over-interpreting the data provided by the new Reception baseline assessment.
- 10 The government should continue to hold schools to account for both progress and attainment, but no judgements about a school should be based on the floor standard alone.
- 11 Schools should no longer be required to tell parents that their child has or has not 'met the expected standard'. Instead, parents should simply be told their child's scaled scores on the Key Stage 2 tests, alongside their teachers' broader assessment of their attainment and achievements.

Employing the most effective responses to under-performance

- 12 The government must urgently commission research into the success of compulsory academisation as a school improvement mechanism, investigating both the desired and the unintended consequences of this approach. It must also urgently seek to learn, and share, lessons from multi-academy trusts (MATs) which have succeeded in improving underperforming schools, as well as from those which have not.
- 13 In the current absence of evidence for the benefits of academisation as a driver of school improvement, the government should consider the wisdom of this being the only available action for schools in certain circumstances (ie being judged inadequate by Ofsted). Instead, they should permit local authorities and Regional School Commissioners (RSCs) to use greater discretion in the actions they can take with schools deemed as underperforming, and commit to more effectively tracking the impact of different approaches.

Ensuring we continue to build our collective understanding of how accountability works

- 14 The government should investigate in detail the likely impact of possible changes to the current accountability system. They should also commit to regularly monitoring both the positive and negative impact of the way in which they hold schools to account, and to finding ways to minimise unintended and undesirable consequences on both individual schools and on the education system as a whole.
- 15 The government should commit to piloting different approaches to accountability in order to explore potentially more effective long-term solutions.



Section 1: Introduction

A system under scrutiny

The way in which children are assessed in English primary schools has recently come under extensive scrutiny.

The Commission on Assessment Without Levels, which **published its final report** in September 2015, examined how schools were responding to the opportunity to develop their own approaches to in-school assessment after the removal of National Curriculum levels.

The NAHT's 2017 report, *Redressing the balance*, examined the current context for primary assessment and set out a series of principles for effective assessment systems.

The Education Select Committee published, in April 2017, **the results of a wide-ranging inquiry** into primary assessment and its impact on children and schools.

And a recent report from LKMco and Pearson, *Testing the Water*, was the result of a year-long consultation on the future of assessment in English schools.

Collectively, these reports painted a worrying picture of a troubled assessment system.

The Commission on Assessment Without Levels described a period of uncertainty in which *"many schools [were] just beginning the journey towards assessment without levels"*. The report commented that *"to ensure the success of these journeys and to embed and share effective new practices, further support may be needed"*.

The NAHT report lamented the lack of a *"vision for a stable, proportionate and coherent approach to acknowledging children's achievements and measuring school performance"*.

The Select Committee echoed the worries of the Commission on Assessment Without Levels about the lack of training and support for teachers in developing their own approaches to assessment, condemned the rushed implementation of recent changes, and voiced concerns about the distorting effect of high stakes assessments on the primary curriculum and on pupil and staff wellbeing.

And the LKMco report found that *"only one third of classroom teachers feel 'very confident' conducting assessment as part of their day-to-day teaching"*, and that *"far too much of classroom teachers' work is geared towards summative assessment"*.

Steps in the right direction

There are positive signs that the government is willing to listen and respond to these concerns. The STA, which develops and administers the primary statutory assessments, now holds regular meetings with representatives of the school leadership and teaching unions, at which our members' views are aired and discussed.

In October 2016, the then Secretary of State Justine Greening, made a number of commitments which helped to address some of the most glaring problems with the current system. These included:

- making the Key Stage 1 grammar, punctuation and spelling test non-statutory
- improving the moderation of teacher assessment
- dropping the proposal to require children who didn't meet the expected standard in their Key Stage 2 tests to resit them in Year 7

The recent government consultation on primary assessment, which closed in June 2017, also included some thoughtful and sensible proposals which could lead to further improvements. The government confirmed, in its September 2017 **response to the consultation**, that it will go ahead with most of these proposals. This includes moving the input to the primary progress measure from the end of Key Stage 1 to the beginning of Reception, making the Key Stage 1 assessments non-statutory, and **considering better approaches to assessing writing**.

The pernicious problem

These changes, however, will have little impact on the most pernicious aspect of the current system: the way in which the results of statutory assessments are used. The Select Committee report pulled no punches on this, stating that *"many of the negative effects of assessment are in fact caused by the **use of results in the accountability system** [our emphasis] rather than the assessment system itself"*. The report recognised the importance of holding schools to account, but concluded that *"this high-stakes system does not improve teaching and learning at primary school alone"*.

ASCL agrees. While the increased focus on progress in primary accountability is welcome, the current system still places far too much weight on a single set of tests in English

and maths, taken in one week in May by 11 year-olds. The effects of this, as we explore further in Section 6, include:

- a narrowing of the curriculum experienced by many children (particularly at the upper end of primary school)
- the de-professionalising of teachers and school leaders
- a tendency to over-analyse (or wrongly interpret) often flimsy data

Furthermore, performance in these assessments can trigger a set of punitive consequences for schools which fall below certain thresholds.

The Select Committee proposed a number of solutions to this problem, including publishing a rolling three-year average of Key Stage 2 results instead of results from a single cohort, and requiring Ofsted to report on the extent to which every primary school it inspects offers a broad and balanced curriculum. We are heartened that Ofsted is similarly concerned about the distorting effect of assessment and accountability on the school curriculum, and welcome Ofsted's focus on this issue in its current thematic review of the curriculum.

We believe, however, that the broader impact of the current approach to primary accountability, and possible ways of ameliorating its negative effects, merit further scrutiny. We hope that the findings and recommendations in this report will prove a helpful contribution to the debate, and provide a useful steer in tackling this pernicious problem.



Section 2: How are primary schools in England currently held to account?

The government publishes a number of benchmarks, designed to enable parents and pupils to see how their school is performing in comparison with other schools. Despite the fact that more and more schools now operate as part of MATs, rather than as individual institutions, accountability still largely operates at the level of the individual school.

The current benchmarks for primary schools in England are as follows:

1 **A set of headline, data-driven measures, which appear in performance tables.**

These currently consist of:

- the percentage of pupils achieving the 'expected standard' in reading, writing and maths at the end of Key Stage 2
- pupils' average scaled scores in reading and maths at the end of Key Stage 2
- the percentage of pupils who achieve at a higher standard in reading, writing and maths
- pupils' average progress in reading, writing and maths

2 **Performance against the floor standard:** the minimum standard for pupil attainment and/or progress that the government expects most schools to meet¹. Currently, a school will be above the floor if, in its latest annual results:

- at least 65% of pupils meet the expected standard in reading, writing and maths; or
- the school achieves sufficient progress scores in all three subjects².

3 **Performance against the coasting definition:** intended to identify schools in which, over time, pupils do not fulfil their potential. Currently, a school³ falls within the coasting definition if:

- in 2015, less than 85% of pupils achieved Level 4 in reading, writing and maths and below the national

1 Floor standards do not apply to infant schools, special schools, independent schools, pupil referral units, alternative provision, hospital schools, schools with very small cohorts or schools which have recently closed and reopened as sponsored academies.

2 Details of how 'sufficient progress' is determined can be found here: <https://www.gov.uk/government/publications/primary-school-accountability>

3 Again, some schools are excluded from the coasting measure, including PRUs, special schools, alternative provision academies, maintained nursery schools and schools with very small cohorts.

median percentage of pupils achieved expected progress in all three subjects; **and**

- in 2016, less than 85% of pupils achieved the expected standard at the end of primary school and average progress made by pupils was less than -2.5 in reading, -2.5 in maths or -3.5 in writing; **and**
- in 2017, less than 85% of pupils achieved the expected standard at the end of primary school and average progress made by pupils was less than -2.5 in reading, -2.5 in maths or -3.5 in writing

Various public bodies use these benchmarks as a starting point to check, on behalf of parents and pupils, that schools are performing well for their pupils and to intervene when schools are perceived as underperforming. These include:

- Ofsted, which uses a publicly-available **common inspection framework** and handbook to consider the effectiveness of a school's leadership and management; the quality of teaching, learning and assessment; pupils' personal development, behaviour and welfare; and the outcomes for children and other learners
- Local authorities and RSCs, who use the various benchmarks to determine where to focus support.

In addition, MATs may set their own benchmarks against which to assess the quality of education their schools are providing.

The consequences of underperforming against some of these measures can be severe:

- A maintained school judged inadequate by Ofsted will be issued with an academy order and required to become a sponsored academy.
- An academy judged inadequate can be moved to a different trust.
- Schools which fall below the floor standard may be issued with a warning notice and, if they are unable or unwilling to comply with the requirements of the notice, may become eligible for further intervention (including, again, being issued with an academy order).
- Schools that fall within the coasting definition may be subject to formal action by their RSC (yet again, including potential forced conversion).

The stakes are high.



Section 3: What alternative approaches are there?

The purpose of accountability

There are many reasons why a society might wish to hold its schools to account. These include:

- to help parents to choose (or at least 'express a preference for') a school for their child
- to help parents keep an eye on the standard of education being provided by their child's school
- to incentivise school leaders and teachers to have ambitious aims for their pupils
- to help governments to decide where to focus limited resources
- to enable governments to incentivise particular behaviours or a focus on particular aspects of learning in schools – and to disincentive others
- to help governments to understand how the system is performing as a whole – and how this has changed over time

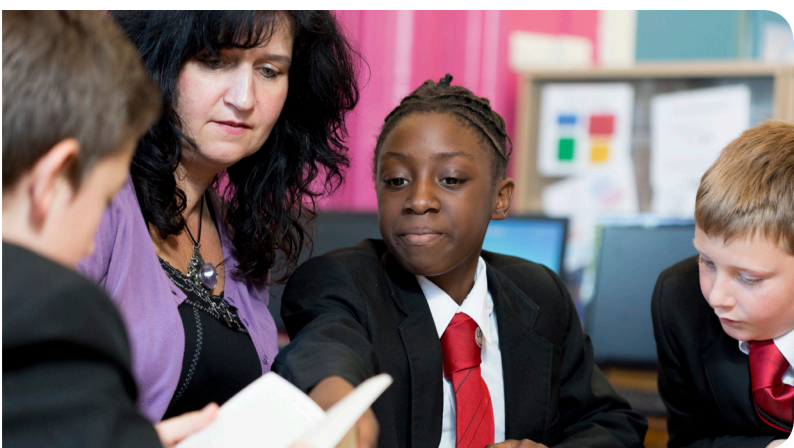
The relative importance a particular jurisdiction places on each of these purposes will influence its approach to accountability. Systems focused on enabling informed parental choice, for example, are likely to prioritise accountability measures which are relatively simple to understand. Those that are most interested in the performance of the system as a whole may choose to sample attainment in a representative proportion of schools, rather than requiring all pupils to undergo the same assessments.

The unit of accountability

Different jurisdictions also take different approaches to the level at which accountability is exercised. Some

policymakers favour a focus on accountability for individual teachers – for example, through merit or performance-based pay – rather than for schools. Others prefer to exercise accountability at a regional, rather than individual school level.

There are pros and cons to all of these approaches. This report seeks primarily to suggest improvements which are achievable within the current broad approach to school accountability in England, ie one based mainly on individual school performance against a prescribed set of measures. Our focus is largely, therefore, on what those measures should be, and on how government might most effectively use them to drive improvement.



Section 4: Measuring what matters

Standards-based accountability systems, such as in England, largely involve evaluating a school's performance on the basis of pupil performance measures. This approach emerged out of a desire, particularly seen in the US and the UK from the 1980s onwards, to measure performance in the public and non-profit sectors⁴. The objective was to identify a set of clear, measurable and ambitious performance standards for pupils across a number of core subject areas, to align the curriculum to these standards, and to expect pupils to meet these high standards.

Such systems generally have the following characteristics:

- Pupils undertake standardised assessments at set points in their school career, to ensure they are meeting the expectations set out for them.

- Schools with pupils who are relatively successful (or unsuccessful) in meeting these expectations are publicly identified.
- Schools are explicitly or implicitly rewarded or sanctioned on the basis of aggregate pupil performance on these assessments.

Data gathered in this way should, in theory, provide policymakers and other stakeholders with independent information about how well schools are performing in comparison both with their peers and against externally set expectations. Attaching positive and negative consequences to meeting or failing to meet particular targets incentivises school leaders and teachers to concentrate on the subjects, knowledge and skills being measured, and is a powerful lever by which policymakers can influence the behaviour of schools.

Any accountability system can only measure so much, and policymakers must make trade-offs in order to obtain a manageable and meaningful view of a school, teacher, or region's performance. These decisions include:

How broad-based the approach should be

Systems that align accountability with a smaller set of outcomes enable policymakers to focus on results that they believe are particularly important, and to measure those relatively accurately. The disadvantage of such approaches is that they tend to narrow the scope of the education provided to pupils. There is strong evidence that schools tend to:

- concentrate their attention on the subjects tested and on the year groups that take high-stakes tests⁵
- shift their teaching emphasis from non-tested to tested subjects⁶
- focus more on tested content areas within specific subjects⁷
- concentrate their energies on the most easily-improved areas of instruction within tested subjects⁸

Many school leaders, teachers, parents, and other stakeholders are keen to see a broader set of outcomes valued via the accountability system. There is evidence⁹ that some non-test-based indicators of school performance (including the drop-out rate, rate of suspension and pupil mobility measures) correlate relatively closely with pupil

4 See Figlio and Loeb, 2011 for more on standards-based accountability systems.

5 Deere and Strayer, 2001; Ladd and Zelli 2002; Stecher et al., 2000

6 Hamilton et al., 2005; Jones et al., 1999; Koretz and Hamilton, 2003; Linn, 2000; Stecher et al., 1998; and Stecher et al., 2000

7 Shepard and Dougherty, 1991; Romberg et al., 1989

8 Chakrabarti, 2005

9 Hanushek and Raymond, 2003

attainment. This suggests that it may be possible to use such indicators to replace, or at least supplement, pupil performance measures in accountability systems. It also suggests that, if schools are trading off these outcomes in order to increase measured and incentivised outcomes, the accountability system may be counterproductive.

However, blunt attempts to measure broader outcomes by, for example, introducing subjective measures, or tests in more subjects, may simply create a different set of problems, such as an increase in the cost of the assessment and accountability system, and the temptation to ‘teach to the test’ in more subjects.

Attainment or progress?

There are two main approaches to measuring school performance in a standards-based system: focusing on attainment or focusing on progress. These two approaches measure different things, tend to generate different objectives and incentives for schools, and can lead to different rankings. Many schools deemed ineffective based on their aggregate attainment levels may actually have quite high ‘value added’ scores, and vice versa.

Which pupils to include and which to exclude

The obvious answer may be that all pupils should be included in measures designed to hold schools to account. But this risks missing some of the complexities of accountability systems. Should schools be held to account to the same extent for children who have been on roll for six years and those who only joined six months ago? Should the attainment or progress of children with special educational needs and disabilities be included in accountability measures in the same way as those of other pupils? Different answers to these questions can have a significant impact on the perceived performance of different schools.

Whether to place particular emphasis on certain groups of pupils, and if so, which

Policymakers can choose to design accountability systems to deliberately focus attention on traditionally underperforming groups of pupils, such as those from disadvantaged backgrounds or those from particular ethnic groups. This can be a powerful policy lever, but can also lead to judgements being made about schools based on disproportionately small numbers of pupils.

How long a time period to consider

Finally, designers of accountability systems need to consider the time period they wish to employ. Accountability systems based on a single year of data are far more likely to

misjudge the performance of schools than those based on multi-year moving averages¹⁰. The disadvantage of the latter approach, of course, is that it requires more years of data to spot indications of improvement or decline.

We don’t have all the answers

As the summary above suggests, this is a complex area which raises myriad questions, many of which we don’t yet know the answer to. We still have no clear understanding of how the many factors involved in designing an accountability system work together, and what consequences, both intended and unintended, result.

It is important, then, that we keep this uncertainty in mind as we consider ways in which England’s primary accountability system might be improved. There are aspects of the current system which seem problematic, and appear to be driving adverse behaviours; we explore these further in Section 6, and recommend some actions which evidence suggests may help with these undesirable consequences.

It is imperative, though, that we don’t make any kneejerk changes which may solve one problem while creating another, that we set up systems to properly monitor the impact of the current approach and any changes, and that we, as a country, are willing to make further adjustments in future if evidence suggests we should.



Section 5: What should an effective and fair accountability system look like?

As already outlined, we still have much to learn about how accountability systems work and how the different variables interrelate. However, we believe that it is both possible and helpful to set out the principles of what an effective and fair accountability system should look like. These will enable us to shine a light more closely on the current system, and see how well it aligns with these principles.

10 Kane and Staiger, 2002

In our view, the principles of an effective and fair accountability system should include the following:

- 1 start from a shared understanding of what outcomes we, as a society, want for our children and young people
- 2 be based around a set of measures which incentivise schools to deliver on these outcomes, seeking ways to recognise and reward aspects which are important but difficult to measure, as well as those which are more easily quantifiable
- 3 drive positive behaviour
- 4 be based on information which is as accurate as possible, and not try to read too much into a small, unrepresentative amount of data
- 5 be fair to schools in different circumstances and contexts, while recognising the importance of enabling every child to reach their potential
- 6 lead to fair, proportionate, transparent and constructive consequences for schools which fall short of its desired outcomes, aligned with the best current evidence of what is most likely to lead to improvements
- 7 be relentlessly self-critical, regularly evaluating impact and modifying as necessary



Section 6: How does the current system measure up against these principles?

Here, we take each of the principles in Section 5 in turn, explore how well the current primary accountability system in England measures up to them, and suggest some ways in which it might be brought into closer alignment.

Principle 1: Start from a shared understanding of what outcomes we, as a society, want for our children and young people

The historical view

The question of what education is for has exercised the finest minds for millennia. Statements on the purpose of primary education over the last 150 years or so offer wildly different visions, from the instrumental and utilitarian to the expansive and flowery.

At the instrumental end of the spectrum, such statements tend to focus on the primacy of the '3Rs'. Back in 1861, the Newcastle Commission Report made clear its view that *"the duty of a state in public education ... is to obtain the greatest possible quantity of reading, writing and arithmetic for the greatest number"*.

The Dearing Report of 1993 picked up the refrain, asserting that *"the principal task of the teacher ... is to ensure that pupils master the basics skills of reading, writing and number"*. The 1997 government white paper *Excellence in Schools* concurred, stating that *"the first task of the education service is to ensure that every child is taught to read, write and add up"*.

In contrast, other commentators have emphasised the development of the whole child and its place in society. The 1931 Haddon Report's statement on the aims of primary education said:

The primary school should not ... be regarded merely as a preparatory department for the subsequent stage, and the courses should be planned and conditioned, not mainly by the supposed requirements of the secondary stage, nor by the exigencies of an examination at the age of eleven, but by the needs of the child at that particular stage in his physical and mental development. The primary school should ... arouse in the pupil a keen interest in all the things of the mind and in general culture, fix certain habits, and develop a reasonable degree of self-confidence.

The 1967 Plowden Report took the 'child-centred' theme a stage further:

A school is not merely a teaching shop. It is a community in which children learn to live first and foremost as children and not as future adults ... The school sets out deliberately to devise the right environment for children, to allow them to be themselves and to develop in the way and at the pace appropriate to them ... It lays special stress on individual discovery, on first hand experience and on opportunities for creative work. It insists that knowledge does not fall into neatly separate compartments and that work and play are not opposite but complementary. A child brought up in such an

atmosphere at all stages of his education has some hope of becoming a balanced and mature adult and of being able to live in, to contribute to, and to look critically at the society of which he forms a part.

The current position

We have no such grand view of the purpose of education at present. The closest we come is the stated aims of the National Curriculum for both the primary and secondary stage:

- 1 *The National Curriculum provides pupils with an introduction to the essential knowledge that they need to be educated citizens. It introduces pupils to the best that has been thought and said, and helps engender an appreciation of human creativity and achievement.*
- 2 *The National Curriculum is just one element in the education of every child. There is time and space in the school day and in each week, term and year to range beyond the National Curriculum specifications. The National Curriculum provides an outline of core knowledge around which teachers can develop exciting and stimulating lessons to promote the development of pupils' knowledge, understanding and skills as part of the wider school curriculum.*

While these aims demonstrate lofty ambition, they make little attempt to define the outcomes we aspire to, as a nation, for our children (beyond an ability to appreciate human creativity and achievement, and the somewhat tautological suggestion that education should result in educated citizens). Without a clear, shared vision for what we want our schools to achieve, how can we even begin to hold them to account for the outcomes they deliver?

A possible way forward

There is no shortage of existing educational aims that could form a starting-point for the development of such a shared vision. Many successful jurisdictions have already clearly articulated their aims.

Singapore, for example, has a set of four Desired Outcomes for Education, attributes that educators aspire for every Singaporean to have by the completion of their formal education. Someone schooled in the Singapore education system should be:

- a **confident person** who has a strong sense of right and wrong, is adaptable and resilient, knows himself, is discerning in judgement, thinks independently and critically, and communicates effectively
- a **self-directed learner** who takes responsibility for his own learning, who questions, reflects and perseveres in the pursuit of learning

- an **active contributor** who is able to work effectively in teams, exercises initiative, takes calculated risks, is innovative and strives for excellence
- a **concerned citizen** who is rooted to Singapore, has a strong civic consciousness, is informed, and takes an active role in bettering the lives of others around him

These outcomes, according to Singapore's Ministry for Education, "*establish a common purpose for educators, drive our policies and programmes, and allow us to determine how well our education system is doing*". They are further translated into a set of developmental outcomes for each key stage, with each educational level building on the previous stages and laying the foundation for subsequent ones.

Ontario has a clear mission statement for its education system:

Ontario is committed to the success and well-being of every student and child. Learners in the province's education system will develop the knowledge, skills and characteristics that will lead them to become personally successful, economically productive and actively engaged citizens.

It has recently, following extensive consultation, defined four renewed goals for education:

- **Achieving Excellence:** Children and students of all ages will achieve high levels of academic performance, acquire valuable skills and demonstrate good citizenship. Educators will be supported in learning continuously and will be recognised as among the best in the world.
- **Ensuring Equity:** All children and students will be inspired to reach their full potential, with access to rich learning experiences that begin at birth and continue into adulthood.
- **Promoting Wellbeing:** All children and students will develop enhanced mental and physical health, a positive sense of self and belonging, and the skills to make positive choices.
- **Enhancing Public Confidence:** Ontarians will continue to have confidence in a publicly funded education system that helps develop new generations of confident, capable and caring citizens.

Closer to home, Scotland's Curriculum for Excellence includes a clear statement of aims:

The purposes of the curriculum 3–18 are to provide the structure and support in learning which will enable [all children and young people] to develop these four capacities:

- Successful learners.

- Confident individuals.
- Responsible citizens.
- Effective contributors.

In an English context, the Cambridge Primary Review, the most comprehensive inquiry into English primary education for half a century, and which published its final report in 2010, proposed 12 aims for primary education, under three headings. In summary, these are:

The individual

- wellbeing
- engagement
- empowerment
- autonomy

Self, others and the wider world

- encouraging respect and reciprocity
- promoting independence and sustainability
- empowering local, national and global citizenship
- celebrating culture and community

Learning, knowing and doing

- exploring, knowing, understanding and making sense
- fostering skill
- exciting the imagination
- enacting dialogue

How different might a system which sought to hold schools to account against a set of aims such as any of these, look from the one we currently have?

Recommendation 1

Government should work with a broad range of stakeholders to co-construct a clear set of aims for primary education and a shared, long-term view of the desired outcomes for children as a result of that education.

Principle 2: Be based around a set of measures which incentivise schools to deliver on these outcomes, seeking ways to recognise and reward aspects which are important but difficult to measure, as well as those which are more easily quantifiable

As we explored earlier, no school accountability system can measure everything that a society deems important. Compromises must always be struck. But, as John Macbeath put it: “We should measure what we value, rather than valuing what we can measure”.

The quality of the current measures

The approach in England currently leans very much towards the ‘instrumental and utilitarian’ end of the spectrum discussed above. Most of our accountability measures focus on the performance of an often very small group of children, in a single year, in reading, writing and maths – and on the subset of what can be relatively easily tested within those subjects.

Reading, writing and maths are assessed in different ways, which impacts on the reliability of the outcomes. The new reading and maths tests, introduced in 2016, are perceived as being relatively high-quality assessment instruments. This view was reinforced by a recent evaluation of the tests by **Ofqual**, whose regulatory role includes the promotion of standards and confidence in the primary National Curriculum assessments. Ofqual concluded that:

Although we have identified potential areas for improvement and further research, our findings provide support for the robustness of STA’s approach to domain sampling for the new suite of National Curriculum tests. Their approach compares favourably with approaches adopted for similar tests, internationally.

There is much less confidence, however, in the results of the third assessed subject: writing. Writing is currently teacher-assessed at both key stages. Although teachers’ judgements are scaffolded by a **set of assessment frameworks**, there is substantial evidence, both anecdotal and more objective, that different teachers interpret these frameworks in very different ways, and that the moderation of these assessments by local authorities also varies significantly¹¹.

STA have acknowledged concerns about the assessment of writing and have put in place measures designed to improve arrangements, including national training and standardisation of all Local Authority moderators, and improved guidance to clarify expectations of the writing frameworks and address misconceptions.

These changes, as well as the fact teachers and schools are adjusting to the transition from National Curriculum levels, appear to be having an impact. Data released in August 2017 suggests that there was less variation in writing results between 2016 and 2017 than between 2015 and 2016.

Teachers, however, remain sceptical about the accuracy of the writing results, with fewer than one in five, according to a **recent poll**, believing that this year’s results will be honest and accurate.

11 See, for example, Education Datalab’s 2016 blog post on comparing LA performance in writing compared to reading: <https://educationdatalab.org.uk/2016/09/consistency-in-key-stage-2-writing-across-local-authorities-appears-to-be-poor/>

Ofsted: a broader view?

The principal means by which we currently hold primary schools in England to account, then, are both narrow and of variable quality. But the National Curriculum assessments, while paramount, are not the only accountability measure. Ofsted inspections are designed to judge a school against a broader set of criteria:

- Leadership and management.
- The quality of teaching, learning and assessment.
- Pupils' personal development, behaviour and welfare.
- The outcomes for children and other learners.

Ofsted inspections should, then, provide a broader, more balanced view of the effectiveness of a school, to balance the narrowness of the performance measures. And, to a certain extent, this is the case. However, there are limits to how far Ofsted can meaningfully perform this role, particularly in schools previously judged outstanding, which are exempt through government regulations from further inspection unless in certain specific circumstances.

There are also concerns that Ofsted too often focuses mainly on English and maths, and that their judgements are overly driven by a school's performance in the National Curriculum tests. A **recent TES article** reported that subjects other than English and maths were rarely mentioned in Ofsted reports, with references to science appearing in only 4% of primary reports, and languages and RE in only 3%. History and geography appeared even less frequently. By contrast, maths was mentioned in 74% of reports, reading in 64% and writing in 67%. In the same report, a former inspector described the extent to which, in his view, Ofsted inspections are driven by the same data as the performance tables:

The inspection itself is only two days, or one day. I think the nature of inspections has focused quite narrowly on measurable data indicators ... Inspectors prepare for school visits by setting up a series of hypotheses, based on the data they can see ... This tends to lead Ofsted down the track that the government selected.

This is, of course, an anecdotal view. And Ofsted has, to its credit, begun to recognise and attempt to address this issue. Speaking at ASCL Annual Conference 2017, Her Majesty's Chief Inspector, Amanda Spielman, spoke about her commitment to encouraging schools to provide a broad, rich, deep curriculum, and acknowledged the conflict the current accountability system sets up between school leaders' "desire to give children the right education and the pressure to maintain [their] league table position". Announcing that the curriculum would be the focus of Ofsted's first big thematic review under her tenure,

Spielman committed Ofsted to "look[ing] at how schools are interpreting the National Curriculum or using their academy freedoms to build new curricula of their own and what this means for children's school experience. We will look at what makes a really good curriculum. And we will also look at the problems, such as curriculum narrowing, and what we can do to tackle them."

ASCL welcomes this focus. We look forward to working closely with Ofsted as its work on the curriculum review continues, and we hope that this report will be a useful contribution to their thinking in this area.

The bigger question

We believe, however, that we also need to ask some bigger questions, not only about what we are holding schools to account for, but how we know how successful they are against these outcomes. What would a school that is really succeeding in delivering against our co-constructed aims for primary education look like? How could we tell? And for those outcomes that are difficult to measure, are there reliable proxies that we could focus on instead?

There is some interesting work being done in this area. Dr Rebecca Allen and Sam Sims at Education Datalab, for example, are currently investigating the relationship between teacher working conditions and school performance, exploring the extent to which staff turnover could operate as a proxy for school effectiveness.

It is important that we continue to consider deeply the most effective and reliable ways to evaluate schools against the outcomes we believe are important, including those that may not be immediately obvious.

Recommendation 2

Both government and Ofsted should engage with research around different ways to measure outcomes which are important but less tangible, and consider how these might be incorporated into the accountability system.

Recommendation 3

Ofsted should commit to commenting more frequently in its inspection reports on subjects other than English and maths, and ensure its focus on published performance figures is proportionate.

Recommendation 4

The STA should seriously consider how writing might be assessed more reliably as part of the accountability system. If this is not possible, writing should be excluded from the performance measures.

Principle 3: Drive positive behaviour

Accountability is a powerful lever, and one which can result in both desired and unintended consequences.

Are standards increasing? At what cost?

There is significant evidence that standards-based accountability systems can lead to improvements in the outcomes being assessed. A 2015 study¹² of the relationship between the use of test results and US students' performance in PISA, for example, found a significant positive correlation between using test results to hold schools to account and pupils' performance in the tested subjects. The research found no statistically significant relationship between pupil performance and other uses of test scores, such as informing parents of their child's performance, providing information for instructional purposes and evaluating teachers and principals.

The same study, however, warned of the dangers of too great a focus on test results in accountability. The authors cautioned that:

Accountability systems of this type may also cause schools to shift resources away from important but non-tested subjects and to focus more on subjects heavily represented in state tests. In some cases, teachers may "teach to the test" by narrowing the content and add extra test preparation activities. Therefore, when higher student performance is observed on large-scale tests presumably as a result of accountability policies, it is important to know whether the increased test performance also reflects meaningful gains in a broad range of cognitive skills.

This raises two key questions when considering the impact of standards-based accountability systems:

- 1 Do improvements in the measured outcomes represent a genuine increase in pupil attainment – both in the knowledge and skills specifically tested and in cognitive ability more widely?
- 2 What is lost as a result of a disproportionate focus on a narrow set of outcomes?

The sawtooth effect

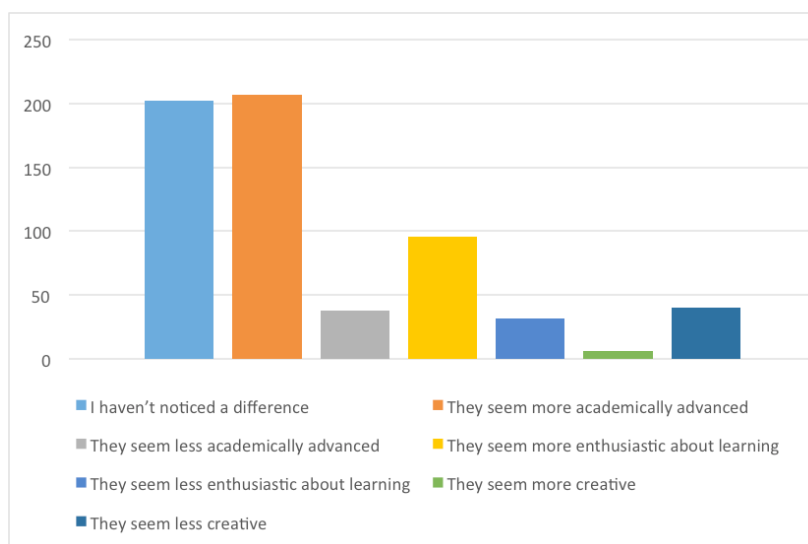
It is too early to say whether or not the new, more challenging, primary National Curriculum tests in England are leading to meaningful improvements in the domains they are designed to assess. The 2017 results, in all assessed subjects and at both key stages, were significantly higher than those in 2016. This could represent genuine improvements in cognitive ability, or it could simply be

a result of the recognised 'sawtooth effect', with teachers being more confident in preparing children for the tests, and children having followed the new National Curriculum for a year longer.

One, admittedly unscientific, way to get a sense of what difference the new primary curriculum and assessments may be having is to ask secondary schools what difference they are seeing in pupils now in Years 7 and 8, compared to their peers under the previous system. ASCL did exactly that as part of a series of conferences held around the country in autumn 2017.

At these conferences, we asked 621 secondary school leaders to choose from seven statements to best describe their current Year 7 and 8s compared to their predecessors. The limitations of the software we used meant they could only choose one statement, so if they felt that more than one were true, we asked them to choose the one they felt most strongly about.

The results were as follows:



- Some hadn't noticed a difference, which may partly be due to a lack of direct involvement with the relevant pupils.
- The majority, however, did perceive a difference, and that difference was largely positive.
- 207 of the 621 thought the most striking difference was that the pupils who had followed the new primary curriculum and taken the new assessments were more academically advanced, with only 38 opting to say they were less academically advanced.
- Perhaps more surprisingly, 96 chose 'They seem more enthusiastic about learning' as their top response (with 32 saying the opposite). The only measure on which secondary leaders felt less positive was on creativity, with

12 Li, Fortner & Lei, 2015

40 opting for 'They seem less creative' (and 6 choosing the opposite).

We make no claims that this data proves anything conclusive, but it gives an interesting informal snapshot which provides tentative grounds for optimism that the new primary curriculum and assessments may be leading to genuine improvements in children's attainment in the domains tested.

An impoverished curriculum?

However, the effect of the new national assessments on children's broader abilities, not to mention less tangible outcomes such as their wellbeing, is very difficult to evaluate. There is growing concern that an accountability system that appears to incentivise schools to concentrate excessively on a subset of English and maths, and perhaps to value performance against a narrow set of measures over deep, sustained learning, is leading to an impoverished curriculum and educational experience for children.

Back in 2010, even before the recent ratcheting up of expectations, the Cambridge Primary Review found that *"many teachers felt impelled, because they considered they were being judged on the SATs results, to spend a good deal of time in Year 6 and sometimes earlier in revision and practice tests"*, cautioning that *"disproportionate time was spent on the subjects tested at the expense of creativity and personal and social development"*. The National Association for Primary Education's submission to the review went even further, claiming that *"in a great many schools coaching for test performance has replaced education"*.

These concerns have only grown in the intervening years. Launching the Ofsted review of the curriculum last year, Amanda Spielman talked of a *"corrosive pattern"* in which many schools feel they have little choice but to narrow their curriculum or teach to the test in order to get *"the league table pay off"*. Describing the behaviours of some school leaders as *"nothing short of a scandal"*, she urged leaders and teachers to remember that:

Childhood isn't deferrable: young people get one opportunity to learn in school and we owe it to them to make sure they all get an education that is broad, rich and deep ... Vitally important though a school's examination results are, we must not allow curricula to be driven just by SATs, GCSEs and A levels. It is the substance of education that ultimately creates and changes life chances, not grade stickers from exams.

Amanda Spielman warmed to this theme in her second commentary as HMCI, published in October 2017, suggesting that it should *"not be taken as read that higher scores for the school always means a better deal for pupils"*.

Discussing the emerging findings of the curriculum review, she reported that, from Ofsted's research with primary schools so far:

- Leaders of eleven of the 14 schools visited were explicit that they carried out some form of preparation for SATs.
- Preparation time for the tests varied between a few weeks in the lead up to the exams and a longer sustained period, typically from the end of the Easter holidays, but sometimes from Christmas.
- The leaders of one school informed inspectors that their pupils sat test papers every week in Years 5 and 6.
- Around half of the 163 parents who responded to Ofsted's questionnaire believed that test preparation had reduced the teaching time available for the other foundation subjects or for reading for pleasure.
- Some leaders, in order to cope with staff workload issues, had chosen to push curriculum development down their list of priorities, indicating that preparing staff to teach to the tougher assessment criteria for the new SATs was more pressing.

These findings may only have come from a small sample of schools so far, but they are worrying, and align all too closely with the messages coming from other studies¹³, and from schools themselves.

Principled leadership

It's easy, and right, to call for Ofsted to put its own house in order here. If it believes schools should be providing a broad, rich, deep curriculum, why does Ofsted so rarely report on any subject but English and maths? If it wants to discourage schools from focusing too much on SATs, perhaps it needs to explore more closely how inspectors use the results of these assessments to form judgements. School leaders may feel justified in expecting a slightly more self-reflective commentary from HMCI.

However, school leaders (and governors) also need to accept some responsibility for the situation in which we find ourselves. It is right that schools ensure children are able to show what they can do when they sit their SATs, and that they aren't fazed by either the content or the style

13 Spielman's commentary references **Sir Michael Wilshaw's earlier commentary** on the downgrading of science and foreign languages in primary schools; Ofsted's 2001 report on the **adverse impact of the National Literacy and Numeracy Strategies**, along with increasingly demanding performance targets, on the breadth of the primary curriculum; and their subject reports on **art and design** and **history** which raised similar concerns.

of the assessments – but is sitting test papers every week for two years the best way to achieve this? Few people would dispute that success in English and maths is crucial to children's success – but how many school leaders and teachers are fully *au fait* with the extensive research that suggests the best way to achieve this is through offering a broad and balanced curriculum, not an over-concentration on 'the basics'?

Strong, principled, ethical leadership is as important now as it ever was. The **ASCL Ethical Leadership Commission**, which is running in parallel with our primary accountability review, is considering this issue in depth, and its findings and recommendations will have resonance in this area. It is crucial that school leaders hold firm to their beliefs about what is important for the children for whom they are responsible, immerse themselves in what research tells us about the best ways to achieve those outcomes, and encourage their staff to always do what is in the best interests of their pupils.

Ethical leadership, then, is part of the solution. It remains perverse, though, that we have a system that appears actively to disincentivise many school leaders and teachers from doing what they (and, indeed, most policymakers) believe is right for their pupils. School leaders need to think carefully about their own response to these incentives, but those responsible for creating accountability measures also need to consider the impact they are having on children's education, both positive and negative, and to be prepared to adjust them if necessary (see Principle 7 for more on this).

Recommendation 5

School leaders and teachers should engage closely with educational research, particularly around effective curriculum design. They should be confident that their school curriculum is informed by the best available evidence on how to enable children to succeed both against the current performance measures and in the broader outcomes they value.

Recommendation 6

Providers of school leadership programmes should ensure they include a strong focus on both ethical leadership and curriculum design. School leaders should ensure teachers in their schools are given the encouragement and opportunity to develop their own expertise in these areas.

Principle 4: Be based on information which is as accurate as possible, and not try to read too much into a small, unrepresentative amount of data

We have already mentioned our concerns, in the discussion of Principle 2, about the validity of the current approach to assessing writing. It is unacceptable that an assessment in which people have so little confidence should be used as part of a high-stakes accountability system. The government must commit to either improving the way in which writing is assessed, or excluding it from the accountability measures.

While there is more confidence in the reading and maths tests, they still only provide a snapshot of a child's attainment, at a particular moment, in a relatively unfamiliar environment, in a subset of the knowledge and skills children possess in those subjects. A child's results in these tests undoubtedly tell us something useful, but we should be wary of any attempt to place more weight on them than such assessments can realistically bear.

The problem of small cohorts

We should be even more cautious about making assumptions about a school based on the performance of a single cohort in these assessments.

A third of primary schools in England (more than 5,000) have fewer than 200 pupils, with 12% having fewer than 100. A school with 200 pupils on roll may have around 25 children a year sitting the Key Stage 2 SATs; a school with 100 pupils may have only 12. Given the likely year-on-year variation between, for example, the prior attainment of these children, or the number of children in the cohort with significant SEND, it is patently ridiculous to suggest that a single year's SATs results can be a reliable indicator of a school's performance.

And yet this is what our accountability system too often does. Looking back at the performance measures outlined in Section 2, the majority are based on a single year's performance in the Key Stage 2 SATs. The progress part of the floor standard does at least attempt to set this in the context of the cohort's prior attainment (see our discussion of Principle 5 for more on this), but it is still based on the performance of a small number of children in a small number of subjects. Only the relatively new coasting measure attempts to provide a longer term view of a school's performance, but, again, based on only a few subjects.

A proportionate response

The government has taken some welcome steps recently towards treating the data from a single year in a more proportionate manner. The DfE's guidance for local authorities and RSCs on intervening in schools causing concern used to include the power to issue a warning notice to a school in which the current year's performance data puts it below the floor standard. If the school failed to comply with the warning notice to the satisfaction of the RSC or local authority within the specified period, it became 'eligible for intervention'. The RSC or local authority then had the power to intervene in a number of ways, up to and including (in the case of the RSC) making an academy order and appointing a sponsor to take over the running of the school. This could have major consequences for both the school as a whole and the individuals leading it.

A recent revision to this guidance still includes a school's most recent performance data as an indicator that local authorities and RSCs should take into account when considering whether to issue a warning notice. The new guidance, however, states that:

No single piece of performance data or inspection outcome will determine any decision on intervention. Before deciding whether a warning notice is necessary, local authorities and RSCs will consider the school in the round, and consider a range of data and other evidence of the school's performance and capacity to improve.

This recognition of the volatility of a single cohort's results, particularly in a primary context, is welcome. We would, however, encourage the DfE to go further. Echoing the recommendations of both the **NAHT assessment review group** and the **Education Select Committee**, we strongly encourage the DfE to commit to basing the primary performance tables on a school's performance over a three year period, rather than on the performance of a single cohort.

We all need to understand data

While we strongly believe the system itself needs to change here, we also believe there are actions school leaders and governors can take to help themselves, and that leadership organisations such as ASCL can support them. Some school leaders and governors are extremely skilled in interpreting and analysing data, but many are under-confident in this area. If they believe that the way in which data is being used to judge their school does not represent an accurate picture of its effectiveness, they need to be prepared to argue a different case.

It is easy to see a red box in a table and accept that it represents low performance; it is much harder to

get a handle on whether or not that apparent under-performance is statistically significant. It is easy to get swept along by a narrative of poor performance based on a narrow set of targets; it is much harder to halt that tide with a well-argued vision of high performance against a broader set of measures that you think are important for your pupils.

Holding fast to your school's vision and ethos, and standing up to ill-informed or inaccurate portrayals of its performance against often flimsy data, isn't easy. And we are not naïve enough to think that attempting to disguise poor performance in reading by talking about how well the children did in the local swimming gala will work – and nor should it. But school leaders and governors do need to ensure they understand the data being used to judge their schools intimately, to be prepared to question inappropriate interpretations of that data, and to remember that it's not the only thing that matters.

Recommendation 7

The primary school performance tables should be based on data from a three year rolling period, rather than on results from a single year's assessments.

Recommendation 8

School leaders and governors should ensure they understand their school's performance data intimately, and that they are sufficiently skilled in analysing and interpreting data more generally. School leadership organisations, and providers of leadership programmes, should support them more effectively with this.

Principle 5: Be fair to schools in different circumstances and contexts, while recognising the importance of enabling every child to reach their potential

The focus on progress

The government has taken some significant, and welcome, steps over the last few years to recognise the very different contexts in which schools operate, and to attempt to level the playing field in how they hold schools to account. The **secondary floor standard** is now entirely determined by the progress pupils make from the Key Stage 2 SATs to their performance in GCSEs (or equivalent qualifications). At primary, the floor standard gives equal weight to progress and attainment. The Ofsted **Common inspection framework** now also gives greater weight than previously to the current progress of pupils, and relies less on historical published data.

Progress measures are far from perfect, but we believe they are an important way of attempting to recognise the value

different schools add. The fact that the current primary progress measure starts from almost halfway through a child's time at primary school, using as it does the Key Stage 1 SATs results as its starting point, is an oddity which will be addressed through the introduction of the new Reception baseline assessment, planned for September 2020.

This was a controversial decision, and developing a baseline assessment which is robust, reliable, a strong indicator of success at Key Stage 2, and appropriate and proportionate to the age of the children involved will be challenging. ASCL supports the introduction of the Reception baseline in principle, but it is essential it is thoroughly trialled before being rolled out, to ensure it is as valid and reliable as possible. It is also crucial to remember that no assessment is ever 100% reliable, and no single measure can tell us everything about a school's performance.

Particular care needs to be taken with a measure which:

- a) is based on an assessment of very young children
- b) spans such a long time period. During the seven years between a cohort starting Reception and finishing Year 6, there may have been significant turnover in both the pupil population and the leadership and teaching teams. The average length of a primary headship, for example, is less than seven years¹⁴, so in many cases the head who was in place when a cohort started school will no longer be there by the time they leave

It is essential, therefore, that the caution we urged above about reading too much into a single piece of data is also applied to the results of the new Reception to Year 6 progress measure. Data, including that from this measure, should always be the start of the conversation, not the conversation itself.

The challenge of threshold measures

It's also important to remember that progress forms only part of the primary floor standard. The other part is based on pupils' attainment in the Key Stage 2 SATs. Schools need to reach a threshold of at least 65% of pupils achieving the expected standard in reading, writing and maths in order to be above the attainment part of the floor standard.

The fact that a school can be above the floor as a result of their performance on either the progress measure or the attainment measure should, in theory, mean we avoid the worst of the issues associated with threshold measures. These include an excessive focus on pupils at the borderline, inappropriate treatment of pupils or schools that fall just either side of the line, the demotivation of children

who will never reach the threshold, and children being labelled (or labelling themselves) as 'failures'.

However, uncertainty around both the current accuracy and the future shape of the progress measure, combined with potentially punitive consequences for schools that fall below the floor, means that many school leaders do see the attainment measure as a cliff-edge. And this too often leads to the undesirable behaviours described above.

Anecdotally, children likely to fall just short of achieving the expected standard are often the subject of disproportionate scrutiny and intervention in many schools, just as those working just below Level 4 were previously. Perhaps this is what policymakers intended, but it appears to go against the spirit of reforms designed to encourage schools to focus on enabling all children to achieve as highly as they possibly can.

More importantly, perhaps, the binary nature of the 'met / not met expected standard' judgement can have a profound impact on children's perceptions of themselves as learners. The **DfE's Assessment and Reporting Arrangements** for the KS2 SATs require schools to include in their reports to parents, in as many words, whether or not their child has met the expected standard. While many children were all too aware of the perceptions of a 'Level 3 child' compared to a 'Level 4 child', even that wasn't as stark as being told, at the end of seven years of primary school, that you have not lived up to the expectations of your teachers, and indeed of society as a whole. How does that make you feel, as you prepare to leave behind the familiarity of your primary school for the uncharted waters of secondary education?

ASCL is not opposed to the dual nature of the headline primary accountability measures. While enabling children to make as much progress as they possibly can is vital, it is also important that as many children as possible leave primary school ready to engage fully with the demands of the secondary curriculum. Including both progress and attainment-based accountability measures encourages schools to focus on both of these aims.

Neither are we entirely opposed to a threshold-based standard in principle. In a system with finite resources, we need a mechanism to determine, as accurately as possible, which schools are most in need of support to deliver the quality of education all children deserve. The important thing is that the floor standard is kept in perspective, that it is seen as just one of the measures by which schools are held to account, and that there are no automatic consequences for falling below it.

14 <http://webarchive.nationalarchives.gov.uk/20130402120754/https://www.education.gov.uk/publications/eOrderingDownload/RR336.pdf>

We do, however, have grave concerns about a threshold-based approach being applied to individual pupils. Children of primary age should be encouraged and supported to work hard and effectively for the love of learning and for their own future success – not to enable their school to jump through externally-imposed hoops. Eleven year-olds should not be told that they have failed in the first stage of their education, just as they are about to move on to the next. We can do better than that.

What about infant, junior and middle schools?

Finally, it is important to recognise that not all schools fit the 4-11 followed by 11-19 model on which the current approach to accountability is based. In regions which retain three-tier systems, for example, the Key Stage 1 SATs take place partway through first school, and the Key Stage 2 SATs partway through middle school. Treating the results of such assessments as these schools' key attainment measure is far from ideal.

In addition, where primary education is split across separate infant and junior schools, a number of challenges are created. Key Stage 1 results tend to be higher in infant schools than in all-through primary schools (probably due to differing incentives: it is in primaries' interests to depress Key Stage 1 results as they form the input to their progress measure, while it is in infant schools' interests to maximise them, as they form their key attainment measure¹⁵). This makes it extremely difficult for junior schools to perform well against the progress measure.

And the difficulty of holding both infant and junior schools to account will be exacerbated by the government's proposal to make the Key Stage 1 SATs non-statutory from 2023. This is a welcome move overall, but raises further questions about the right approach when children move school at age seven.

ASCL's response to the 2017 government consultation on primary assessment, which asked about the most effective accountability measures for infant, junior and middle schools, suggested a number of different models. Options include requiring all schools to administer progress assessments on entry and exit (and in intermediate years if they wish); and holding all schools to account for the progress they make from the Reception baseline, whatever age they are when they leave the school.

There is no simple answer to this issue, but we are pleased that the government has acknowledged the problem and is seeking a better solution.

Recommendation 9

The government must be cautious about over-interpreting the data provided by the new Reception baseline assessment.

Recommendation 10

The government should continue to hold schools to account for both progress and attainment, but no judgements about a school should be based on the floor standard alone.

Recommendation 11

Schools should no longer be required to tell parents that their child has or has not 'met the expected standard'. Instead, parents should simply be told their child's scaled scores on the Key Stage 2 tests, alongside their teachers' broader assessment of their attainment and achievements.

Principle 6: Lead to fair, proportionate, transparent, and constructive consequences for schools which fall short of its desired outcomes, aligned with the best current evidence of what is most likely to lead to improvements

So far, we have touched several times on the consequences of falling short of the current accountability measures, commenting that these can be both far-reaching and, worryingly, potentially based on insubstantial data.

What we haven't yet explored is whether or not they are effective. If major, and potentially hugely disruptive changes are being imposed on schools which are perceived as underperforming, it's vital that we are as confident as we possibly can be that this will lead to meaningful, sustained improvement in those schools.

Powers of intervention

The DfE's *Schools causing concern* guidance includes a long list of ways in which local authorities and RSCs can support or intervene in schools perceived as underperforming. These include:

- leaving the school to implement its own improvement plan
- brokering support for the school from a Teaching School Alliance, high performing local school or National Leader of Education
- requiring the school to make changes to its governing board
- appointing additional governors

15 See Education Datalab's blog for more analysis of this issue: <https://educationdatalab.org.uk/2015/03/we-worry-about-teachers-inflating-results-we-should-worry-more-about-depression-of-baseline-assessments/>

- appointing an interim executive board (IEB)
- suspending the school's delegated budget
- directing the closure of a school
- taking over responsibility for an IEB
- requiring the school to become a sponsored academy
- terminating the funding requirement for an academy and moving it to a new trust

Different actions will be taken in different circumstances, and some are obviously only relevant to either maintained schools or academies. In some circumstances (such as falling within the definition of coasting), legislation allows the local authority or RSC considerable discretion in how to respond. In others (a maintained school being judged inadequate by Ofsted), there is no choice but to go straight for the nuclear option of forced academisation.

It is not unreasonable to expect that the agencies making such important decisions about schools' futures do so in a way which is fair and transparent, and that their decisions are based on an evidence-informed view of the actions most likely to lead to long-term improvements in each case.

The missing evidence

Worryingly, this does not appear to be the reality. Minutes of meetings between the RSCs and their headteacher boards, the engine room of the decision-making process, are often only released months later, and then severely redacted¹⁶. We lack any real evidence for the benefits of academisation as a mechanism for school improvement¹⁷. Stories are emerging on a worryingly frequent basis of MATs that are failing to adequately support, let alone improve, the schools for which they are responsible.

ASCL has no issue with academisation in itself. Many ASCL members run highly successful academies and MATs. We are convinced of the benefits of formal, long-term partnerships of all types between schools¹⁸, and work tirelessly to support schools that want to join or form MATs¹⁹. We recognise that many of the conversations the RSCs need to have about schools in their area are sensitive, and need to be conducted discreetly. And we understand that school improvement is not an exact science, and there is no silver bullet that is guaranteed to work in every case.

However, the current situation remains a very long way from our vision for a fair and effective accountability system. Not only is there scant evidence for the benefits of the current approach to school improvement, there are also many examples of concerning, unintended consequences, such as the disincentive for school leaders to work in challenging schools.

The government owes it to school leaders, teachers, parents, and young people to properly research its chosen approaches to schools deemed underperforming, and to demonstrate that they are indeed leading to meaningful, sustained improvement.

Recommendation 12

The government must urgently commission research into the success of compulsory academisation as a school improvement mechanism, investigating both the desired and the unintended consequences of this approach. It must also urgently seek to learn, and share, lessons from MATs which have succeeded in improving underperforming schools, as well as from those which have not.

Recommendation 13

In the current absence of evidence for the benefits of academisation as a driver of school improvement, the government should consider the wisdom of this being the only available action for schools in certain circumstances (ie being judged inadequate by Ofsted). Instead, they should permit local authorities and RSCs to use greater discretion in the actions they can take with schools deemed as underperforming, and commit to more effectively tracking the impact of different approaches.

Principle 7: Be relentlessly self-critical, regularly evaluating impact and modifying as necessary

Finally (and to expand on a point we made back in Section 2), not only do we lack evidence for the benefits of academisation as a school improvement mechanism, we also lack a clear understanding of how the many factors involved in designing an accountability system work together overall.

We have explored in this paper some of the apparent consequences, intended or otherwise, of the current

16 See TES analysis: <https://www.tes.com/news/school-news/breaking-news/exclusive-extent-academy-secrecy-revealed-dfe-publishes-only-3-papers>

17 See the Education Select Committee's report into academies and free schools for more on this: <https://publications.parliament.uk/pa/cm201415/cmselect/cmeduc/258/258.pdf>

18 See the Education Select Committee's report into school partnerships and collaboration for more on this: <https://publications.parliament.uk/pa/cm201314/cmselect/cmeduc/269/269.pdf>

19 See, for example, our suite of 'Staying in control of your school's destiny' guidance papers for schools considering forming or joining a multi-academy trust: <https://www.ascl.org.uk/help-and-advice/guidance-papers/>

approach in English primary education, and suggested some possible improvements. However, as Baroness Onora O'Neill pithily cautions in an influential article on intelligent accountability in education:

Every time one performance indicator is shown to be inaccurate, or misleading, or likely to produce perverse results, some people claim that they can devise a better one that has no perverse effects. Experience suggests that they may well be as wrong as those who invented the last lot of indicators.²⁰

It is important, therefore, that we are committed, as a society, to exploring different accountability models, to evaluating their impact, and to using the knowledge we gain to improve the approach we take. Professor Robert Coe and Gabriel Heller Sahlgren put forward a compelling case for such an experimental methodology in 2014²¹, proposing a range of pilot programmes to trial different approaches to accountability and to find out what works and what doesn't.

Coe and Sahlgren use the example of different approaches to teacher assessment to make their point, but an experimental methodology could be used to explore radically different ideas. What might be the impact, for example, of only testing a (randomly selected) sample of children at the end of primary school, rather than every Year 6 pupil? What if we replaced the Key Stage 2 SATs with a test that could assess children on any aspect of the National Curriculum? What if we moved away from judging schools based on pupil performance at all, and instead used other measures, such as staff turnover?

The point is, of course, that we simply don't know. If the government is serious about implementing evidence-based policy in education, it needs to commit to properly investigating the impact of the current approach to accountability, to exploring alternative models, to trialling any proposed changes, and to monitoring the system on a long-term basis. This analysis must include the impact of accountability measures both on individual schools and on the system as a whole.

Recommendation 14

The government should investigate in detail the likely impact of possible changes to the current accountability system. They should also commit to regularly monitoring both the positive and negative impact of the way in which they hold schools to account, and to finding ways to minimise unintended and undesirable consequences on both individual schools and on the education system as a whole.

Recommendation 15

The government should commit to piloting different approaches to accountability in order to explore potentially more effective long-term solutions.



Section 7: Conclusion

In conclusion, then, we believe there are a number of changes that could lead to an improvement in the way in which we hold our primary schools to account.

These involve:

- holding schools to account for a broader range of measures
- improving the accuracy of the current accountability measures
- promoting ethical leadership and effective curriculum design
- using performance measures in a proportionate way
- employing the most effective responses to under-performance
- crucially, ensuring we continue to build our collective understanding of how accountability works

Enacting the recommendations in this report would, we believe, help us to focus on the things that really matter in our primary schools, and bring some much needed sense to the accountability system.

We look forward to working with the DfE and others to take these recommendations forward.

²⁰ O'Neill, O, 2013

²¹ Coe, R & Sahlgren GH, 2014

Acknowledgements

ASCL would like to thank the many people who have shared their thoughts on primary accountability with us during the course of the review, as well as those who read and commented on drafts of this final report.

Particular thanks go to the review's expert panel, whose commitment, expertise and good humour was invaluable in shaping our thinking. While the final report inevitably includes views and recommendations which are not shared unanimously across the panel, we hope it reflects some of the depth of our discussions over the last few months.

The expert panel consisted of:

Katharine Bailey, Director of Applied Research, Centre for Evaluation and Monitoring, Durham University

Professor Robert Coe, Director, Centre for Evaluation and Monitoring, Durham University

Dame Reena Keeble, former primary headteacher, author of the recent Teaching Schools Council report on effective primary teaching practice

Catherine Kirkup, Research Director for Assessment, National Foundation for Educational Research

Will Millard, Senior Associate, education 'think and action tank' LKMco

Lee Owston, HMI, Ofsted's Specialist Adviser for Early Education

Dame Alison Peacock, CEO, Chartered College of Teaching

James Pembroke, Data Analyst and TES columnist

David Reedy, former director, Cambridge Primary Review Trust

Richard Selfridge, Primary Teacher and blogger

Michael Tidd, Primary Headteacher and TES columnist

Greg Watson, Chief Executive, GL Assessment

Julie McCulloch, Interim Director of Policy, ASCL and report author