Education Policy Institute & More Than A Score's Event Series – Spring 2023

Re-defining standards in English primary education

To date, standards in primary schools have been measured through two instruments; SATs scores and Ofsted inspection outcomes. Both measures have been critiqued by some in the sector who argue that they have a negative impact on teacher workload and a lack of recognition of pupil demographics (e.g. levels of deprivation or additional educational needs). Both measures are also high stakes for schools; and SATs results are (in normal years) published and used to create "league tables" and can trigger Ofsted inspections.

Some stakeholders, including school leaders, teachers and parents, also have concerns about the efficacy of SATs and the extent to which the continued focus on test scores as a measure of school standards risks displacing a broader and more balanced evaluation of both pupil and school achievement. There are further concerns that SATs, though intended to be low stakes for pupils, have unintended consequences including an effect on wellbeing and impact on breadth of curriculum through a heavy focus on preparing for tests.

The government's 'Levelling Up' white paper, published in February 2022, set out its ambition for 90% of pupils to achieve the expected standard in reading, writing and maths by the end of Key Stage 2, by 2030. This target was reached by just 1% of schools in 2022.[1] The 2030 cohort of primary pupils will be starting Reception in September 2023, meaning that it now seems timely to ask: what do high standards really mean in practice?

In spring 2023, the Education Policy Institute partnered with More Than A Score to conduct two roundtables to discuss whether the current use of test scores to evaluate primary school standards effectively meets the needs of pupils, schools and the system overall and if not, what reforms to the system might look like. This summary paper outlines the key topics that were discussed across the two events, including the current approach to standards setting in English primary education, the impact on pupils and education staff, the role of Ofsted and whether further reform is needed.

We are grateful to More Than A Score for their support and to all participants for their contributions.

How are standards in English Primary Education currently measured?

The current system uses two methods to measure standards in English primary schools: data from statutory assessments and Ofsted inspections. Most well-known among the assessments are the SATs sat by pupils in year 2, and by those in year 6 over a week, in the summer term. Alongside these tests, there are also statutory assessments in three other years meaning that

[1] https://ffteducationdatalab.org.uk/2022/07/key-stage-2-2022-how-things-have-changed-for-schools-and-regions/

children are tested in five out of the seven years at primary school.[2] The government announced in 2017 that year 2 SATs would be non-statutory from September 2023, so the choice to carry out testing in this year will be down to schools.

Data gathered during these assessments are used to inform school leaders, the Department for Education and parents about not only the performance and progress of a child, but also the performance of the school and its teachers. Ofsted also carries out regular inspections in all state schools across the country to measure standards.

There was some debate during our roundtable discussions about how the system of standards setting in England compares internationally. International comparisons were highlighted as challenging to ascertain; one participant suggested that we should not be thinking in terms of good and bad education systems globally, highlighting that this is too simplistic. Rather, even in high-performing systems, there are children for whom the system is not designed or well-suited. They suggested that any discussion around what we can learn from international examples should start by aiming to understand the varying experiences of children within the system, as different children experience education in unique ways. This point was raised again in response to a comment that Finland is often lauded as a "fairly egalitarian" education system and has no primary assessment. Rather, the move from primary to secondary is facilitated by teacher-assessed grades. In response, another participant noted that context is very important, and disadvantage or inequality looks different in Finland than England, so we should be careful to consider context when implementing best practice from other countries.

England was noted by one participant however as having both strong data analysis and, on the part of policy makers at least, a strong desire for this data, while other countries are more reticent to collect evidence.

What are the impacts of the current system?

There was much discussion across both events about the efficacy of statutory assessments and Ofsted inspections, and their impacts on pupils, teachers and the curriculum.

Some participants felt that the current way we measure pupils through assessment is an inexpensive way for the Department for Education and Ofsted to collect data, suggesting that the system is set up to provide easy-to-use data to assess standards, rather than considering what is best for children. Current year 6 SATs test a child's ability in reading, writing and maths, measuring technical abilities, rather than the more complex skills of creativity, teamwork and critical thinking. Given that SATs data is used to measure school performance and set standards, participants shared concerns that this narrow breadth of assessment can lead to "teaching to the test" and narrowing of the curriculum. The key question arose of whether our standards setting system is "the tail that wags the dog" and drives what is taught in schools, rather than the other way around. One attendee felt that this is deeply problematic and that the curriculum should be the point of departure for any assessment design rather

[2] Other statutory assessments include the reception baseline assessment (age 4/5), the phonics screening check in Year 1 (age 5/6) and the multiplication tables check in Year 4 (age 8/9).

than the inverse. Building on this, it was also felt high stakes assessments which have a pass/fail boundary embed the idea of failure in children from an early age, preventing opportunities to encourage learning from mistakes and basing success on a select set of metrics, that do not take into account the wealth of other skills and knowledge a child may possess that cannot be demonstrated through the current assessment system.

Others also described SATs data as two dimensional and insufficient, as it does not take into account external contributing factors which may impact performance such as poverty, housing, mental health, cultural and socio-economic backgrounds. Rather it captures a snapshot on a single day without the opportunity to contextualise and can, some suggested, place a considerable amount of pressure on children's mental health.

Similarly, participants also noted that Ofsted inspections and the need for children to perform well in statutory assessments can have washback effects on teacher mental health too, placing further burdens on the profession. One attendee added that Ofsted should not be held solely responsible for teacher wellbeing and that schools and trusts also have a responsibility to create a supportive environment that fosters wellbeing.

Finally, a key concern shared was the fear that SATs and other assessments undermine the role of the teacher by removing their autonomy and affecting trust between teachers, senior leadership teams and parents. One participant highlighted strongly that we must return to trusting teacher judgement of their pupils' progress. It was noted that teaching standards have been raised in the past thirty years. However, in response, another participant felt that it is not enough to simply say 'trust teachers'; this places a high burden on teachers and a better accountability system is needed.

Moving away from the impacts listed above, one participant highlighted Ofsted's 2019 Education Inspection Framework (EIF) as a "counterbalance" and "shift in perspective" which lessens the reliance on SATs data and looks broadly across the curriculum. When Ofsted conducted research to inform the current framework, the importance of being alert to the purpose of assessment and its current role in the education system was centred, recognising the "perverse consequences" that assessment can create such as the narrowing of the curriculum and "teaching to the test" mentioned above.

Proponents of the new Ofsted framework argue that it aims to assess the extent to which a school instils the joy of what all subjects have to offer and encourages children's ability to think. They also argue that it is designed to hold schools to account to ensure they deliver a curriculum that encourages a love of learning.

However, the extent to which it is feasible to assess in a one-or two-day inspection was questioned. It was felt that this snapshot judgement risked being reductive and would be unable to capture a school in the round. In response, another participant noted that it is important to recognise that one-day inspections are not graded and that to some extent, inspections are limited by the amount of funding that Ofsted is provided with.

Another critique of the EIF was its purported drive to "know more, remember more", encouraging children to become "learned" rather than "learners" and eradicating their

curiosity and love of learning; one participant cited that children ask 80% less questions when they leave primary school than when they enter. In rebuttal, another participant felt that skills are knowledge-based and the ability to think critically and solve problems come from drawing on knowledge schema.

The role of parents was also discussed. One participant felt that there is a move towards treating parents as consumers within the education system and, therefore, that role as consumer within a market has the potential to drive up standards. But there was some debate over the extent to which parents are equipped with the right information to make informed choices. One participant felt that parents know whether schools are good without Ofsted reports, so to test children continually to provide this information is redundant. This view caused pushback, with other participants stating that they do not agree that parents just know a good school when they see it. A 2021 University of Exeter study found that just under two-thirds of parents do not consider Ofsted reports when deciding where to send their children to school.[3] Moreover, children's SATs results are supposed to come out in annual reports but often parents are merely told whether their child is below, meeting or above the expected standard. Participants tended to agree that the current information provided by Ofsted is not enough to equip parents with the knowledge to make informed choices about where to send their child, and to what extent they are succeeding in school. At the same time, it was noted that the amount that teachers and parents communicate is facilitated by a strong school culture, rather than through Ofsted. One attendee felt that "schools are terrible at listening to parents" so to involve parents more in their child's learning requires building this culture, outside of any reliance on Ofsted judgements.

Ultimately, the debate returned often to ask whether and to what extent the current system offers the skills and knowledge children need to succeed. There was consensus that, although standards setting and accountability are very important, the current system has resulted in unintended but negative consequences; that the multi-purpose nature of assessment can have undesirable impacts on teachers and pupils alike; and that often our current system does not encourage the development of a broad range of skills, including 'soft' skills such as self-confidence and agency over learning.

How could we reform the system to improve?

When considering how to reform the standard-setting system, participants agreed that the underlying purposes of assessment are important to consider. As noted above, currently SATs data is mainly used for accountability purposes to inform external bodies including Ofsted and the Department of Education rather than improving learning processes. Looking back to when the national curriculum was first introduced in 1988, it was suggested that at that time there was a need for greater school accountability, as there were children in disadvantaged areas who were leaving school without developing basic skills such as being able to read and write, primarily due to poor quality teaching and leadership. Since then, however, accountability has

[3] https://education.exeter.ac.uk/news/articles/starkdividesinparentingat.html

ramped up, and the participants agreed that we should be thinking more broadly in terms of the role that assessment now plays in the education system.

Participants agreed this should start with the needs of children and young people. Attendees discussed the importance of considering not just the breadth of the curriculum, but also the skills we want young people to develop, including their sense of self. It was agreed that it can be challenging to think strategically beyond the current assessment system, as it is highly engrained, however attendees suggested we could start by considering a set of questions to formulate new ideas. Topical questions included thinking about the range of knowledges, skills and abilities pupils should learn at different stages, how these can be assessed and how assessment can support teachers to do their jobs effectively. With these questions in mind several alternative possibilities were discussed in relation to how the assessment system could be reformed.

One method discussed was based on the work of the BERA expert panel. Their aim was to address the key question of how we can measure standards without using SATs results and how we could design an intelligent and democratic system for school accountability. Looking back to the first report published by Ofsted that used SATs data, attendees recalled that the highest performing schools had an advantaged intake, sparking interest in the concept of value-added. This is a measure of the progress that pupils make over time that considers the pupils starting point and compares their progress relative to similar pupils. If you start with an advantaged intake, the participants agree that there will be a different set of pedagogic problems that need to be addressed than if you start with a disadvantaged intake, and in order to address highly varied needs it helps to have an understanding of the demographic.

To achieve this, the BERA panel considered the impact of removing whole population testing in favour of using a national sample. The advantage of a national sample is twofold; firstly, fewer students have to participate in testing, and this decreases the pressure on schools to spend a disproportionate amount of time teaching to the test. Secondly, the panel envisioned using survey data on pupils, parents and teachers and for this combination to gives us a better grip on the ethos of the school and the whole community. A metric system could be used to collect more sophisticated contextual data that would be compared with attainment data to reveal where investment is needed in different areas. One of the things that the COVID-19 pandemic made starkly clear is that children don't have equal access to quality study conditions when they are working at home, including a quiet study space and access to technology. The BERA model aims to support understanding for how poverty intersects with learning and influences what children need to thrive.

By zeroing in on specific problems that relate to different localities and making the data available in the public domain, some attendees suggested that the BERA model could support research being commissioned to focus on what different children need, and results can then be used to steer investment. Decisions could also be made democratically or locally to reopen a conversation about how children are progressing to the whole community, rather than having decisions made in government without external contribution. An accountability body such as Ofsted could then report on the data being produced so that parents can still have an insight into their child's school quality, according to key themes that are marked as areas of

interest for each individual school, such as student wellbeing, a broad, rich curriculum and high-quality teaching, which were regarded as important measures of standards by parents and senior leaders in a recent report published by More Than a Score.[4]

On the other hand, one attendee argued that overhauling the system could potentially be damaging. Schools are already under a great amount of pressure that would be exacerbated by such an extensive reform programme. Moreover, they highlighted that changing the primary accountability system would create a domino effect as it would have a knock-on effect on how we measure progress in secondary schools. Most notably, the BERA model does not provide individual pupils with a baseline score from which progress can be measured at the end of Key Stage 4. Instead, this participant suggested that there is a middle ground where the system can be reformed to address the main concerns with SATs without a significant overhaul of the system. More focus should be given to how we can address the breadth of the curriculum, the narrow focus on English and mathematics, and the length of SATs testing which is done over four days.

These concerns could be addressed by creating a smarter assessment design to broaden the scope of the assessment so that questions on multiple subjects are included. In doing so, it would not be possible to anticipate questions, and this would decrease teaching to the test and curriculum narrowing. The participant that favoured this model also supported moving towards using computer-based assessments to enable us to test a broader range of skills, and the combination of testing multiple subjects at once. Using computer-based assessments could also decrease the amount of time dedicated to testing. It was further suggested that using an average from several cohorts to inform league tables would decrease the amount of pressure placed on students and teachers from high stakes assessments.

Some attendees were in favour of moving towards using e-portfolios, in order to build an understanding of children's learning over time and consider a broader range of knowledge and skills, encouraging creativity and agency over one's own learning. It was suggested that this model could be used to assess in a way that is comparable to how teachers assess in the early years when practitioners cannot rely on standardised tests. Supplementing teacher assessment with e-portfolios made by the children themselves could support teachers to feel empowered to make decisions in the moment and in the best interests of pupils, which could also support bringing trust back into the teaching profession. This deeper approach, it was argued, helps to bring pupils and teachers back to the centre of assessment by using self-evaluation, rather than producing data that is exclusively for external accountability purposes.

This approach could yield a much richer data set to inform standards setting and also support the students themselves to track their own progress, which can then also be sent home to parents. As highlighted above, one of the concerns with the current assessment model that was discussed by attendees is that some parents report being unaware that their child was

[4] The More Than A Score report run through Teacher Tapp and You Gov found that 70% of school leaders and 47% of parents believe a 'broad curriculum' is the best measure of high standards; 57% of heads and 35% of parents said 'love of learning'; 39% of heads and 47% of parents said 'mental health'; 40% of heads and 36% of parents said 'high quality teaching', while 4% of heads and 12% of parents said 'SATs results' were the best measure of standards.

working below the expected standard until receiving their SATs results. E-portfolios could create an opportunity to open up communication with parents, and the process of creating a portfolio can help teachers and pupils to know themselves as learners. E-portfolios also provide a large amount of evidence for curriculum leaders to look back on to support curriculum development. Despite this potential, the downside to using e-portfolios is that they can put greater pressure on teacher workloads. Participants felt there also needs to be a careful balance so that teachers are not constantly assessing rather than focusing on teaching, and in the early years, it has been observed that continuous paperwork created by assessment can strain this balance.

Finally, participants discussed the extent to which current standards and assessment policies should consider changes to the world of work over the next thirty years, including in relation to emerging technologies such as Artificial Intelligence. Some participants stated that the assessment system should be designed with this in mind, rather than informing decisions based on the world of work as it exists today. However, others felt strongly that there are fundamental skills (such as numeracy and literacy) that have remained relevant through the age of the internet and are likely to remain relevant through the age of AI. Instead of embarking on a likely impossible task by trying to predict what the future will look like, several participants felt that perhaps our focus should once again return to the needs of children including developing basic skills and allowing assessment processes to be driven by this purpose. However, participants agreed that not only literacy and numeracy should be seen as basic skills; rather this category should be broadened to include 'soft skills' such as creative problem solving, critical thinking and collaboration. One participant highlighted that, given the rapid pace of change, we need to "liberate children to react". It was agreed that developing these skills would help to facilitate this, encouraging adaptability and flexibility in the face of change.

Conclusion

The discussions across the two events covered a large amount of ground, touching on our standards system, the efficacy and impacts of primary assessment and links with Ofsted and the accountability system. When considering how standards should be defined and measured, the panel ultimately returned to fundamental questions of what we want children and young people to learn and how the levers of assessment and the accountability system can support that.

As the paper outlines, there was considerable debate, but consensus was generally reached on

- the need for clarity on the purpose of assessment, particularly at the end of primary
- the fact that the current system has resulted in unintended, negative consequences
- the need to consider, alongside whether SATs need reform, what role Ofsted inspections should play.

While opinions on the nature of reforms differed, participants felt that any reforms must take into account the need to prioritise closing the disadvantage gap, recognising that any assessment system that relies on support outside the school day is inherently unfair and that when we judge schools, we are also judging society as poverty, housing, mental health support and other contextual factors all have an impact on performance. Therefore, changes to the assessment system need to work alongside and take into account these underlying problems. Secondly, any reforms should consider the already significant burdens placed on teachers and school leaders and be implemented in partnership with the sector, rather than imposed, to ensure that the system is able to cope. Finally, that our current interpretation of success, defined largely by SATs, is considered to be a narrow one which jeopardises the development of a broad range of knowledge and skills.