Standards setting: how we can ensure our assessment system is fair and supports all young people to achieve their goals?

During the pandemic, policy makers, school leaders and exam boards were suddenly faced with the challenge of how to award end of qualification grades to young people without the use of traditional exams. This raised issues of fairness, regional inequality, parity of standards and the impact of grade variations on next steps, both for pupils and post-16 and higher education institutions. At events held by the Education Policy Institute (EPI) at the political party conferences in 2022, it became apparent that many of these issues remain of real concern, both for policymakers and the education sector. Themes that ran throughout the discussion included whether our current assessment system is inclusive enough; its (lack of) resilience during the pandemic; the role of technology and whether its use in assessment systems will widen or narrow attainment gaps. Following these events, EPI and Pearson chose to bring together experts to consider the following key questions:

1. Is our current assessment delivering on its intended aims? If not, why not?
2. Does the assessment system need to be reformed? If so, how?

This paper is a summary of the discussion of the roundtable (conducted under Chatham House Rule), drawing together reflections on the current academic assessment system and in particular, its approach to standards setting and the impact of the accountability system. We are grateful to Pearson UK for their support of the event and to all participants for their contributions.

The current assessment system in England

In England, we currently have a system where most students sit their GCSEs at age 16 in a series of high stakes pen and paper assessments occurring over a short period in the summer of their year 11. These results are used not only to evaluate a young person's level of competency and how well they have learned the skills and knowledge in the curriculum, but also to demonstrate their performance to employers and further education institutions (depending on the route they choose) as well as to evaluate school performance. The discussion made clear that we are asking far more from our assessment system than simply determining a pupil's level of knowledge and this multifunctional nature causes problems.

During the discussion, some participants highlighted positive attributes linked to high stakes assessments that were generally centered on the role of accountability measures in schools. High stakes assessments provide an opportunity to collect data on students and teachers, allowing comparisons to be made in terms of performance. It was purported that, before current accountability measures were introduced, schools across the country were 'failing' students, however it was difficult to intervene without any sense of the challenges being faced. When 'failing' schools are now taken over, exam results are scrutinised and provide a key indicator of broader issues occurring in the school.

Despite the need for performance data, a negative aspect of high stakes assessments being used as a measure of accountability is that it creates intense competition between schools, and some participants claimed that teachers often feel frustrated as progress cannot be recognised without negative outcomes in other schools. Accountability measures are based on an understanding that

healthy competition can drive up performance, however the pressure to perform has been shown to encourage educational rationing. Research conducted by Gillborn and Youdell (2000) revealed that resources such as high-quality teachers are often channeled towards 'borderline' students who were working on the cusp of a passable grade to drive up a school's overall performance.[1]

Participants agreed that grade reforms missed an opportunity to value each grade by defining grade 4 as a standard pass, which produced another threshold and defined grade 3 as a failure. Rather than celebrating diverse talents, it was suggested that high stakes assessments have been more concerned with selection, and that this selective approach, evolving during the 1980s when GCSEs were introduced and education was rationed, is producing wash-back effects that have a negative effect on many students, particularly those who 'fail'. Despite this, a recent survey conducted by the Association of School and College Leaders (ASCL) about assessment preferences following the pandemic found over 90% of senior leaders wanted to return to exams and move away from teacher assessed grades, and that these views were linked to accountability as headteachers were concerned that some schools would 'game' the system. For this reason, most senior leaders who took part in the study regarded exam-based, high stakes assessment as the fairest approach to ensuring that all students are receiving a quality education.

Participants also discussed the idea of differential assessment, and discussed whether this method should be prioritised over a 'level playing field' approach. One attendee highlighted that university admissions recognise challenging circumstances when evaluating applications, so asked could we do the same for exams? There was considerable pushback in response, with one participant asserting they were strongly against exams being "dumbed down" for disadvantaged students and another highlighting that though this happens in university admissions, when degrees are awarded, standards are very much maintained; a university would not award a 2:1 to a student that had 2:2 grades due to their circumstances. The consensus was that it is important to focus on closing disadvantage gaps in the classroom to ensure that pupils, no matter their circumstances, are able to sit assessments equitably.

In England, standards are set based on the data produced from each year of exams, and the current system therefore requires all students to take their exams at the same time so that the results are comparable. To offer more flexibility in relation to assessment, our approach to standards setting would also need to be reformed. Attendees highlighted that students who are excelling academically can take assessments a year early, however this does not offer flexibility to account for the wide variety of needs represented in a cohort, including catering for students with special educational needs and disabilities (SEND) and those who have faced socioeconomic challenges.

This lack of flexibility was highlighted as an argument against the use of high stakes assessments in that it does not take account for the heterogeneity of a cohort. In addition to disadvantaged students and those with SEND, research shows that assessment results are significantly affected by arbitrary factors such as the time of year a student is born, yet the current system classifies students by age rather than stage and requires all students to take their exams at the same time[2,3]. Participants discussed the possibility of assessing children by stage and allowing assessments to be taken when students are ready, but again mentioned that we are limited by the need to collect data at the same time to set standards.

---

[1] Rationing Education - Google Books
[2] When you are born matters: the impact of date of birth on child cognitive outcomes in England | Institute for Fiscal Studies (ifs.org.uk)
[3] The impact of month of birth on child development - Nuffield Foundation

The possibility of assessing children by stage also raised concerns that this could be seen as a return to a modular system. Several participants recalled their experiences as teachers under the modular system where they found that parents and teachers would often 'game' the system, and that there were also more assessments as students were assessed continuously throughout the year. Despite this, other participants suggested that assessing children by stage would not necessarily mean a return to a modular system, and that we can avoid previous pitfalls by learning from past mistakes. Assessing children by stage is more concerned with addressing diversity and inclusion and highlighting that a single mould does not fit for all. To consider such an approach, it would be necessary to reflect on different ways to demonstrate skills and create a different way of thinking about the relationship between knowledge and skills, for example using outcomes-based education where feedback is offered, and children are given the opportunity to apply their knowledge multiple times.

Participants considered the role that teacher assessment could play in such an approach, however there were mixed reactions that mirrored recent research findings. In contrast to ASCL's survey findings that 90% of senior leaders wanted to return to exam-based assessments, a survey conducted by the NEU found that 68% of classroom teachers prefer a mixed economy of exams and teacher assessment. Despite workload pressures, 54% of teachers felt their mental health would improve, and that a mixed economy might improve retention and lessen teaching 'to the test'. This mixed reaction arose in the discussion as some questioned whether we have adequate safeguards to prevent against 'game' playing behaviours, which the exam-based system had originally sought to address, as well as implicit bias in student-teacher relationships. With mixed multi-modal assessment process, others suggested that there is more flexibility to compensate for challenging circumstances that can occur during adolescence and more opportunity for students to demonstrate their skills.

When asked about international comparisons, there was some debate. One participant felt that England is not an outlier as many countries, such as Estonia, Singapore, Hong Kong, Belgium and Ireland, also have high stakes assessment at age 16 and 18, yet it is important to note that this is often not to the same extent as the English system. For example, they highlighted that Estonia only has three exams at age 16 and England is unusual in the number of hours of GCSE assessment. They continued to say that many countries that have teacher assessment wish they had exams due to inconsistencies in teacher-assessed grades. In response, another attendee raised the fact that policy comparisons across countries are challenging as education systems are culturally embedded. They mentioned the example of Japan which has a ranking system throughout a child's time in education where they are ranked from 0 to 100, highlighting that, while there are aspects of many systems that we need to look at carefully to understand what we can learn, it is key to recognise that the systems are rooted in a wider culture and we should be careful about the conclusions we draw and where we want to be heading.

On this concern around bias in teacher assessment, raised several times during the discussion, EPI research has found that fears that the switch to teacher assessed grades for GCSEs in 2020 would penalise students from disadvantaged backgrounds were largely unfounded, and that there was no evidence poorer GCSE students lost out under this system[4]. While there is still the possibility that teacher assessments might be inconsistent and therefore less fair than exam-based assessments, participants agreed that the current system places too much pressure on children, which cannot be justified given that high stakes assessments are mostly concerned with assessing schools and teachers, rather than students. Despite the possible need for reform, participants agreed that changes will have

---

[4] Covid-19 and Disadvantage gaps in England 2020 - Education Policy Institute (epi.org.uk)

a significant impact on pedagogy and classroom management, and that policy makers should instead consider introducing slow, long-term reforms.

Essential skills and their place in curriculum and assessment

Participants considered whether we are truly teaching the right things, emphasising the need to look at the curriculum in any discussion on assessment. A significant challenge to the current system that was raised is that essential skills needed to thrive in the 21st century are often not being developed during early educational stages. This is becoming more apparent as students move through different pathways as, upon reaching university for example, many students struggle to demonstrate skills such as metacognition, creativity and critical thinking as these have not been cultivated at school. Similar concerns are arising in the job market and before the pandemic, 1 in 6 employers were conducting their own tests to assess their prospective employees. According to these employers, GCSEs, A Levels and university grades were not necessarily providing a meaningful assessment of the relevant skills needed to succeed.

To consider how we can include essential skills in the curriculum, it was suggested that we must also discuss how we are assessing these skills, as what happens in the classroom, including pedagogy and classroom management, is largely shaped by the assessment process. Rather than cultivating essential skills, passing exams has become the primary focus and students are consequently unprepared when they leave formal education. One speaker made the plea: "can we stop making people sit behind a desk and do what they are told to pass the exam? What is this actually teaching young people?" Rather than explicitly teaching essential skills, it was suggested that these should be implicitly worked into the curriculum and assessed through alternative mediums which allow for creativity in the process, for example creating and delivering a presentation, debating, analysing and problem-solving, but for accountability purposes, to only assess the thread of knowledge being delivered. Such an approach could have the potential to reduce the burden on students and enable richer learning.

Not assessing individual skills explicitly was deemed important by some participants who cautioned that too much assessment can be damaging. Previously, the education system differentiated between specific and general goods of schooling, where general goods refer to the ability to feel engaged and relate well to other people. Assessment can change behaviors and have other washback effects, and for this reason, policies have not aimed to assess these 'soft' skills. Under the current system, however, it has become clear that essential skills are not being cultivated, partly because the pressure of accountability has meant that teachers have been forced to prioritise methods that will enable students to excel in an exam context. There is currently an open question about how we should proceed in teaching and assessing such skills.

Participants discussed an approach that understands the notion of creativity as complex problem solving, so that the process is quantified rather than the end product, and the Extended Project Qualification (EPQ) was referred to as an example which allows for flexibility but which has also been proven to be robust. To minimise potential variance in quality, it was suggested that a lexicon should be created to produce a common frame of reference and clearly define multifaceted concepts such as creativity. This should allow for skills and competencies such as problem solving, communication and presenting to be measured by putting theory into actionable terms. In doing so, it may become possible to expand the curriculum while concurrently narrowing assessment and creating more fluidity in the process.

A challenge remains that assessment results must be comparable, but some participants suggested that there could be an opportunity to implement adaptive testing to provide a backbone of hard skills such as numeracy, literacy and IT, as well as soft skills such as creativity with a core with comparable indicators. In this way, it would be possible to produce the kind of data required by governing bodies such as the Department for Education and Ofsted, but also provide a greater variety of assessment models for students to nurture essential skills. Ultimately, the participants agreed it will be necessary to spend more time thinking about what we're assessing, why and when, who the data is useful for, and how it is preparing young people for the challenges they will face after formal education.

The use of technology in assessment

Speakers also considered the use of technology in assessment, and in turn teaching, and how it could be used as a tool to ensure fairness and appropriate support.

The use of Artificial Intelligence (AI) resources such as ChatGPT was seen by some attendees as a potential threat to some forms of assessment such as essays or written coursework, with AI having the capability to write essays on complex topics in a matter of minutes. Some saw it as a direct threat to coursework-based assessment, arguing that the only solution would be to continue with invigilated, in-person exams, in the way that GCSEs are currently predominantly carried out. Yet, others highlighted that AI has the potential to enrich the teaching and learning experience for young people in the classroom, through its capacity to generate content. An example was shared where ChatGPT was used to generate essays on alternative outcomes to key historical events, which could then be discussed in class to encourage critical thinking and knowledge of historical process.

School leaders and teachers round the table also raised the point that it is straightforward for teachers to identify work that has not been written by students themselves, given their knowledge of their pupils' writing styles, with easy to identify trends such as excessive use of punctuation or incorrect grammar or word order.

AI was considered to be a potential threat only if assessment methods are not modernised and reformed to recognise the omnipresence of technology in today's world. AI and technology was considered by some participants as an important part of assessment to encourage the development of skills that young people will need as they move into the world of work. A statistic was shared showing that 76% of those sitting assessments felt more comfortable when it included on-screen aspects, given that many students regularly use technology in their day-to-day life. Some participants felt there should be further efforts to incorporate on-screen assessment in all qualifications, with the necessary provisos that accessibility should be central to any reform. One participant highlighted that the majority of assessments for computing qualifications were carried out through pen and paper exams.

Attendees shared concerns over affordability and the challenges faced by students from low-income backgrounds to access devices, highlighting the risk of widening disadvantage gaps if learning and assessment become increasingly dependent on technology whilst access to technology is not improved. Beyond access for students, fears were also shared over affordability, access and technological skills development in schools themselves. Anecdotally, one school leader outlined a typical classroom of 32 students who share 12 devices and use a local authority network of patchy quality, highlighting "this is the reality in many schools." The need to provide continuing professional development for practitioners to keep up with the pace of technological evolution and the challenges

around typing speed for children both with and without SEND during assessment were also highlighted as key barriers to consider.

A final concern shared was the risk that technology poses in relation to social interaction between child peer groups and between pupils and teachers. One speaker felt "technology companies have an impoverished view of what it means to be human," highlighting that while digital skills are critical to flourish in the modern world, there are also concerns around how technology could negatively impact students ability to learn, interact socially with peers or communicate, all of which are key development skills needed to prepare young people for the future. At the same time, an example was shared where a smart board was used in a classroom with over twenty year 10 students. The students' work was visible to all on the smart board and activity was periodically paused for pupils to peer-review their work and discuss mistakes and opportunities to improve. This virtual environment was seen as a catalyst of collaboration, giving students the opportunity to share workings, methodology and resources, with a resulting increase in peer discussion and questions to the teacher. Moreover, this example was used to demonstrate the "blurring line" between learning and assessment, where pupils could use assessment as a learning tool and teachers could gather data about their understanding in a low-stakes environment, facilitated by technology.

Overall, it was felt that technology has great potential, but its implementation must be carefully considered in both assessment and the classroom to ensure it is supported by well-evidence pedagogical approaches.

Conclusion

In this complex discussion on assessment reform, one participant emphasised that when they look at exam scripts, those pupils who get higher grades clearly know more about the subject than those who receive lower grades, a reassuring point. As this paper outlines, there are clearly contested issues in our assessment system, not least of which is the accountability system's dependence on grades to evaluate school performance. Another participant highlighted that "assessment, from first principles, is just a series of questions and tasks that tell us about the mental life and behaviour of individuals and we can absolutely open up the way we ask these questions with all the provisos of fairness." They suggested expanding the number of assessments to ask questions in multiple ways and including more assessment of different kinds would lower the stakes and give a clearer understanding of where the gaps in comprehension lie. Another felt that we need to expand the curriculum but narrow down what we are assessing to ensure the children and young people do not lose their curiosity and love of learning during their time in formal education.

Ultimately, participants broadly agreed that there is a need for reform but what this should look like remains an open question, though loosening the ties between assessment and accountability and investigating where else in the system accountability could come is an important place to start. Any reforms or system change however must be carried out slowly in conjunction with teachers and the wider education sector to ensure that they are able to be implemented effectively in an already stretched system. Any reforms must also be clearly communicated with universities and employers to ensure that standards are maintained and qualifications retain their currency.