

How Will the Government's Proposed Reforms to Qualifications in the UK Impact on Standards?

15 July 2013



Purpose

An audience of colleagues from numerous interested governmental, private, public and charitable organisations was invited to consider and respond to some of the challenges which the Government's proposed changes to qualifications raise for the setting and maintaining standards in general qualifications. The afternoon included presentations from leading academics on the key findings of their recent research into this issue, followed by the opportunity to discuss implications for assessment, standards and policy.

Rod Bristow, President of Pearson Education UK opened the proceedings, welcoming the opportunity this event has provided to discuss some key issues surrounding proposed GCSE reform and encouraging governmental bodies, regulators and awarding bodies to engage and collaborate in debate in this crucial reform period.

The Chair of this event, Roger Murphy, Professor of Education at Nottingham University, then put forward a series of provocations, asking if enough attention is being paid to research outcomes on assessment; whether we have the right balance between curriculum and assessment and if we are learning sufficiently from high performing jurisdictions. These provocations provided the focus for the afternoon's series of presentations.

Standards and Policy



Jo-Anne Baird, Director of the Oxford University Centre for Educational Assessment, is the Pearson Professor of Educational Assessment. She took as her theme [The Return of the GCSE](#). Jo-Anne described the current situation and government imperatives and then presented an analysis of the standards debate and research findings behind some of the myths, assumptions and interpretations of the current system. One of her observations was centred around the greater transparency in the examination system in England over recent years, which is generally considered as a positive development, and has had an impact on performance in assessments. The development of assessments themselves in terms of their structures and reliability is greater than ever, however she also noted that higher order thinking skills were currently insufficiently rewarded and that she was yet to find question papers and mark schemes having a positive backwash on the curriculum. This could be seen as an unrealistic expectation of assessment.

Jo-Anne ended her presentation with an analysis of one of the proposed developments in the GCSE reforms; the use of a reference test. Audience feedback and discussion from this presentation raised concerns about the measurement error associated with such tests and the risk that the debate and currency of a reference test may simply replace and potentially distort the current debate on GCSE standards. Questions were also raised whether in general, reliability of assessments had improved at the expense of their validity. Discussion also centred on the present system of awarding and accountability measures, delegates wondering if they did indeed encourage a 'race to the bottom'. One participant suggested that one national test in each subject area, administered by awarding bodies, might offer a solution, however Jo-Anne voiced the general concern about a 'one size fits all' model of assessment. There was general consensus across the audience that change was needed in the accountability measures to reduce the gaming and distorting effects of the present high stakes examination system.

Standards and Design

The second presentation was given by Dr Rose Clesham. Rose is Head of Assessment Research and Design, Pearson UK. In her talk, [Good Assessment by Design](#), Rose shared findings from her team's recently completed global comparative study of externally assessed examination systems for 16 and 18 year olds. This research study took the form of in-depth analyses of the structures and demands of assessments on a number of key criteria using a methodology developed by Rose for the redevelopment of PISA frameworks. The methodology uses a three-dimensional analytical approach considering the expectations of content standards, cognitive operations (more specific versions of assessment objectives) and cognitive demands of assessment instruments.

Rose shared some of the research findings, which included a global comparison of the assessment of higher and lower order cognitive operations and demands. The evidence strongly indicated the dominance of lower order knowledge and skills in many assessment systems, including those in England.



International research evidence suggests that the knowledge and skills required in a global economy will increasingly require higher order thinking skills and yet there seems to be a widening gap between these desired cognitive skills and those that dominate teaching and assessment systems. At a time of reform we need to ensure our educational curriculum and assessment reforms reflect national and international requirements. We need to ensure as a country that we have a clear view on the type and proportion of knowledge and skills we want our learners to be taught and assessed on and be pro-active in the principles of 'good assessment by design'. This should not be assumed- however we have the opportunity to embed these principles into the current reforms.

In her conclusions, Rose observed that all countries need to take a wider view of their curriculum and assessment systems and that England has much to learn by taking more of an international view.

Audience discussion and feedback following Rose's presentation included exploration of the research methodology employed in her international comparative study and a general consensus view that the reliability of examinations has dominated the assessment landscape for some time and now would be a good time to reconsider what our aspirations should be in terms of curriculum and assessment.

Standards setting

Robert Coe was the next presenter. Robert is Professor of Education, Director of the Centre for Education and Monitoring (CEM) at Durham University. He took for his theme, [Standard Setting and Maintenance for reformed GCSEs](#). Robert began by considering the different mechanisms used around the world to set standards and the differing interpretations of what may be meant by standards and grades, noting that high-performing jurisdictions in terms of PISA do not necessarily produce high-performing schools.

He then went on to review the pros and cons of standards setting and maintenance using judgements against criteria, norm referencing and reference tests. On the basis that complex grading decisions should be made using as much of the available data and information as possible, Robert's recommendations were that all three mechanisms should be used where appropriate, and that we should also consider reporting both grades and fine scores. He was pragmatic that performance culture was here to stay but asserted his view that the issue of standards is not directly about education.



Ideally, Robert would like to see reference tests for each subject, although he acknowledged that this was an unlikely scenario. Key to the success of the reference test is a common understanding of what construct a reference test would be based on and what it would look like. Discussions ended on the potential use of PISA itself as a reference test. It was felt that using it in this way didn't

necessarily solve the problem of how to anchor GCSE grades, however there might be some way of linking PISA to GCSE grading.

Standards and International Lessons



The final presentation was given by Dr. Peter Hill. Peter was able to share some of his valuable insights from his leadership roles in Hong Kong and Australia educational system reform programmes. In his presentation, entitled [The International Experience](#), he described in particular how Hong Kong managed system reform. He was quick to point out however, that success had come through twelve years of implementation with no threat of a change in government.

Peter then went on to highlight some of his recommendations for the GCSE reforms. He felt that international quantitative benchmarking of content standards to uncover best practice design principles was the way to go. He believed that standards could be maintained through the use of professional judgement, psychometrics and script sampling around cut scores and that there was a false dichotomy between extreme positions taken on criterion and norm referencing systems. Peter also described the Hong Kong monitoring test, showing that in terms of manageability, only a very small sample of schools was needed in order to gain evidence and confidence levels.

Emerging themes

The four presentations paved the way for an interesting and lively panel session in which three main themes came to the fore.

The first of these was the use of a national reference test. Though not everyone agreed on the efficacy of its introduction, it was felt that the content of any reference test should be closely aligned to the construct and content standards of GCSE. Some delegates expressed concern that a reference test might become a form of IQ test by proxy, whereas other delegates thought a general ability test would be an effective anchoring device. To remain useful, it was felt that a test would need to be regularly refreshed and more importantly there should be a shared understanding and agreement on the purposes to which the data could be used.

The second discussion theme centred on assessment design. It was felt generally that exams in England have developed into aptitude measures rather than cognitive achievement tests and that students were too easily rote-prepared. It was voiced that the system should encourage the teaching and assessment of future proof skills within content domains (such as non-routine problem-solving, analytical and evaluative skills). It was also acknowledged that this was a challenging proposition for us all: examiners, teachers and students alike. One delegate thought that a reference test might be helpful in trialling these higher order thinking skills question types and produce a positive wash-back by encouraging them to be taught and assessed in schools. It was felt generally that there is currently little incentive to teach higher order thinking skills because they are not particularly well represented in assessments.



This remark heralded in the final theme of system change. The accountability system in particular was heavily criticised for pushing the wrong incentives and encouraging 'poor' behaviours. The risk of focussing on a single qualification was highlighted with a call to aim for systematic change across all qualification types.

Looking forward, the development of online assessment was seen as an opportunity to encompass system change for the good, particularly in the light of many international assessments now moving towards e-assessment.

Discussion then turned to how system change might be brought about. There was a general consensus that it could only be done through a concerted effort and open debate but also that this reform period has created the necessary conditions and climate for change.

Conclusion

From the very engaging and interesting debate about critical issues associated with the current qualification reform, three key areas warrant further debate and interrogation.

There is clearly a desire for further investigation and debate about the purpose, quality and demand of the GCSE curricula and its associated assessment. What do we want our learners to learn and be assessed on in terms of breadth and depth? We know that external high stakes assessments have a significant washback in what and how content is taught. We need to ensure that any washback effects will be positive, and assessments reflect the intended aims and purposes of future GCSE specifications. International comparisons can inform this debate.

The purposes and structure of a reference test merits further investigation and debate with experts and stakeholders. Reference or anchor tests would potentially create a means of providing a better understanding of national attainment and greater confidence in standards. However there are significant issues to be debated, including how many reference tests would be needed; what would their constructs be; when would students take them and what would their purpose or purposes be? These are difficult educational and technical questions and require informed and considered discussion.

The longer term vision for standards and assessments should encompass large scale system change in particular the introduction of on-line assessment offering the capability to assess in new and adaptive ways that are fit for purpose for 21st century learners. We know there is increasingly a gap between the way young people learn using sophisticated technologies and the pen and paper methodologies used in high stakes assessments. We must embrace this paradigm shift in our qualification assessment structures.

Pearson is very grateful for the stimulating contributions to this event made by the speakers and the audience. We recognise that collaboration to further understanding of standards and assessment is critical to the education reform agenda and we look forward to continuing our engagement and discussions.