



Primary assessment and accountability under the new curriculum

SCORE's response to the Department for Education consultation

11 October 2013

Introduction

1. SCORE is a partnership of organisations, which aims to improve science education in UK schools and colleges by supporting the development and implementation of effective education policy. The partnership is currently chaired by Professor Julia Buckingham and comprises the Association for Science Education, Institute of Physics, Royal Society, Royal Society of Chemistry and Society of Biology.
2. In summary:
 - SCORE supports the principle of giving teachers more freedom over what they do in the classroom, including assessment. However, they must be given sufficient support and training in order to do so effectively.
 - SCORE has some concerns about the appropriate use of commercially-produced material in schools. We are particularly concerned about the quality of the materials and whether school budgets will reflect these additional expectations.
 - SCORE agrees that data on performance, added value and context can be valuable tools in judging how well schools are doing in fulfilling their aims. However, there are risks associated with the publication of the data that can drive behaviour in an unintended way, for example narrowing the focus of the curriculum. SCORE would therefore recommend that data is only published as part of a report on a school.
 - SCORE supports the principle of value-added measures, but is concerned about whether it is possible to devise a robust baseline measure for very young children. There is also the danger that schools could be incentivised to depress their baseline scores in order to inflate their value-added scores at the end of key stage 2.
 - If it is decided to introduce baseline measures, it is vital that different approaches are properly piloted. It is also imperative that the same measure is used for all schools.
 - Further work is needed to define what is meant by secondary ready. If it is to be a useful measure of what students have achieved by the end of primary, it must be fully understood by students, teachers and parents, and encompass more than just attainment data for mathematics and English.
 - SCORE has concerns about the use of scaled scores; it is not clear in the consultation why they are needed or how judgements would be made to scale the tests, and thus how useful they will be as a measure of student performance.
 - SCORE is concerned about the use of the decile measure; it implies a level of accuracy in grading that is not present, and it is not clear whether it will be useful information for students, parents or schools.
 - SCORE is pleased that sample tests will continue in science, though there is a need to be mindful of the fact that removing the national curriculum tests has had unintended consequences on the provision of science in some primary schools.

General comments

3. In 2010, SCORE published a set of principles for the assessment of key stage 2 science¹. We are pleased that many of these principles, particularly around the introduction of sample tests, have been taken up. However, the need for clarity around the different accountability mechanisms used in schools remains a significant concern. Any changes to be implemented, following piloting, will need to be carefully communicated both to schools and parents so that all parties fully understand the new arrangements.

Freedom over the curriculum and assessment

4. Many of the proposals in the consultation are based on the principle of giving teachers more freedom over what they do in the classroom. While SCORE is supportive of this, it is important teachers are given sufficient training and support to be able to do so effectively. This is particularly true in science, since very few primary teachers have a science background². There is a danger that without support primary teachers will not have the confidence to plan effective learning in science (that will encourage students' curiosity), and will find it difficult to carry out effectively the ongoing assessment proposed in the consultation. This is particularly true of the scientific enquiry aspects of the curriculum, which are less straightforward to assess than scientific knowledge.
5. It is not clear what quality control will be in place to monitor schools' science. SCORE's understanding is that the focus of Ofsted inspection will primarily be student outcomes (in the statutory tests in English and mathematics) rather than the curriculum they are being taught, with inspections concentrated on those schools whose outcomes fail to meet expectations. While measuring outcomes is an appropriate means of monitoring delivery of the curriculum, we are concerned that this may not be sufficient to ensure that schools teach an engaging science curriculum.
6. Both these issues – teacher confidence in science and quality control – are also influenced by the use of commercially produced teacher resources, including schemes of learning and assessment packages.
 - While there are many high quality resources available to teachers to support their teaching and assessment, selecting the most appropriate for their students requires thorough knowledge of the subject and how it is taught. It is important that this is provided as part of training and ongoing professional development.
 - There is a danger that commercial providers will compete on assessment outcomes (so that schools can choose the assessment materials that are likely to provide them with optimum results, rather than encouraging good learning), which will not be beneficial for schools or students. This is likely to be more of an issue in mathematics and English than science, however, where there continue to be national curriculum tests for every student.

¹ http://www.score-education.org/downloads/assessment/principles_KS.pdf

² *Science and mathematics education, 5-14: A 'state of the nation' report*, The Royal Society (2010), p.36

- There could also be an issue of comparability between different commercial products, which will be a particular problem if results are to be used for accountability purposes. If commercial products are to be used, consistency and comparability between them needs to be assured.
- It is not clear whether school budgets will be increased to take account of the additional expenditure needed for external resources. Science has particular resource requirements to support practical science, and it is essential to safeguard expenditure in this area. SCORE would rather see adequate provision of equipment and consumables in science, with knowledge of how to use them effectively, than the short-term solution to a lack of support – the purchase of published materials.³
- There could be issues with quality control if schools are making individual decisions about which providers to use. Although the Department for Education states in its consultation document that they ‘will work with teaching schools, professional associations, subject experts, education publishers and external test developers to signpost schools to a range of potential approaches to identify and share examples of good practice for schools to draw upon’⁴, it is not clear how this will work in practice. A model based on schools making individual decisions is not necessarily the most efficient or effective.

Accountability

7. Data on performance, added value and context are extremely useful for judging how well a school is achieving its aims. They are also useful for researchers and policy makers. However, if they are published as raw scores (that are sequenced in league tables), then they are open to abuse and can lead to perverse behaviours – some, but not all, of which can be predicted. It is much fairer and more powerful (as a way of encouraging positive behaviours) for the data to be filtered through an expert body – such as Ofsted – and published as part of a digest or report on a school.
8. While SCORE supports the principle of value-added measures of accountability, we have some concerns about how this will work in practice, or whether it would add useful information in terms of judging a school’s performance. Producing a baseline for key stage 2 by testing children when they are very young will not give reliable data, and it will be very difficult to ensure that this will be robust enough to become the baseline from which students are measured for the rest of their time in primary school. There is also the danger that there will be perverse incentives on schools to adjust their assessment scores in order to improve their VA score. It may be more appropriate to provide data on the school context, which will allow schools to be compared like-for-like.
9. Whichever measure is chosen, it is imperative that it is the same for all schools so that meaningful comparisons can be made. We are supportive of the statement, made in

³ See SCORE’s research on resourcing practical science for some of the difficulties in this area. The reports can be downloaded here: <http://www.score-education.org/publications/publications-research-policy>

⁴ paragraph 2.1

paragraph 1.10 of the consultation document, that ‘pupils in maintained schools, academies and free schools are all subject to the same statutory assessment system and accountability regime’, and the same should be true of any baseline measure that is used.

10. Given the difficulties we identify above, we would suggest piloting the different options for establishing a baseline, so that their different impacts can be evaluated.
11. Further work is needed to define what is meant by ‘secondary ready’. It implies a measurement both of the knowledge and skills in a particular subject as well as less tangible qualities such as maturity or attitude. Careful communication will be needed to ensure that all those with an interest in students’ achievement – including teachers (both primary and secondary), parents and students themselves – fully understand what is meant by the new standard. While SCORE supports the importance, implied in the term ‘secondary ready’, of ensuring a seamless transition for students between primary and secondary education, this will be achieved through a well-planned and coherent curriculum and good communication between the different phases of education.
12. It is also unclear how the scaled score will be used. The implication in the consultation document is that ‘secondary ready’ will remain as a fixed standard for students to achieve; in other words the tests will be criterion-referenced. Given that there is an expectation that student performance will improve over time, it is hard to see how reliable judgements will be made about where to set the scaled score, since both the tests and the performance levels are expected to change. There will also need to be some explanation of what a score means in relation to the curriculum.
13. SCORE is also concerned about the effect on both schools and students of the proposed decile ranking. While the achievement of a criterion-referenced standard such as ‘secondary ready’ is reasonable, we do not believe the proposed system of deciles will be helpful to students or their parents. We have concerns about whether it is a statistically valid measure, since it does not take account of potential error in externally marked written assessments.
14. It is not clear in the consultation document what use will be made of teacher assessment for accountability; it would appear that teacher assessment won’t form part of the floor standards, but it will continue to be published. While SCORE welcomes the principle of using a broader range of criteria to judge schools than national curriculum tests alone, we are concerned about the prospect of teacher assessed data being used for purposes for which it was not intended. For this reason, we would recommend that these data are shared only with students and parents. In addition, a clear distinction should be made between assessment for learning purposes, and assessment for accountability purposes.
15. SCORE supports the findings of a joint project between the Association for Science Education and the Nuffield Foundation into the assessment of primary school science, which proposed a framework for assessment which differentiates between individual, school and national performance⁵.

⁵ *Developing policy, principles and practice in primary school science assessment*: report from a working group led by Professor Wynne Harlen (2012)

Sample tests

16. SCORE is pleased that science will continue to be assessed through sample tests, which are an appropriate way of monitoring national performance over time. However, it is not clear in the document what sort of information the tests will provide for schools (for example on students' conceptual development and ability to work scientifically), how the results will be used to improve standards, and who will be responsible for ensuring that schools take appropriate notice of the results.
17. It is worth noting, however, that there is firm evidence of unintended consequences from the decision to remove national curriculum tests in science. A survey of schools carried out by the Wellcome Trust in 2011⁶ found that the removal of the tests had led to a change in the status of science in schools, with less time and resources being devoted to it. These findings were borne out by SCORE's research into the resourcing of practical work.
18. SCORE is not advocating a return to whole cohort testing in science at key stage 2, but it is important to ensure that appropriate science retains its place in the curriculum. This could be achieved by strengthening Ofsted's ability to inspect the time and resources allocated to science, and ensuring that schools make appropriate use of the assessment they themselves are doing to measure the effectiveness of their own science curricula. It is also important to make better use of the results from the national sample tests, and use them when planning professional development.
19. SCORE has some concerns about whether practical work should be included as part of the sample tests. It is unlikely that a simple test can adequately cover the breadth of knowledge across different areas and the practical and thinking skills which currently make up the science curriculum. The danger in trying to develop replicable and reliable practical tests that will provide comparable data is that any tasks will be so reductive and formulaic that will not give any real indication of children's scientific thinking. If practical work is part of the national sample tests, and we would welcome this, it should be more a measure of the school's provision of experiences for students (linking to SCORE's Practical Resourcing benchmarks⁷), rather than the individual students' abilities. We would recommend that this provision is also part of Ofsted's inspection role for primary schools.

⁶ *Primary Science Survey Report*, December 2011. The report can be downloaded here: http://www.wellcome.ac.uk/stellent/groups/corporatesite/@msh_peda/documents/web_document/wtvm053596.pdf

⁷ The SCORE resourcing benchmarks can be found on the SCORE website at <http://www.score-education.org/publications/publications-resourcing-benchmarks>