

What happens when high-stakes testing stops? Teachers' perceptions of the impact of compulsory national testing in science of 11 year olds in England and its abolition in Wales

Sue Collins^{a*}, Michael Reiss^b and Gordon Stobart^b

^aSchool of Sport and Education, Brunel University. ^bInstitute of Education, University of London. ^bInstitute of Education, University of London.

The research aims to articulate the experiences and opinions of Year 6 (Y6) teachers, primary headteachers and primary science coordinators as a contribution to the debate about the effect of statutory national testing in England and its abolition in Wales. A multi-faceted approach of quantitative and qualitative data collection and analysis was used in England and Wales including a telephone survey of 600 respondents and eight focus group interviews involving a total of 74 participants. The findings of this research provide insights into the strategies adopted by Y6 teachers in England in ensuring pupils gain the highest possible scores in high-stakes national tests, used as the basis of published *Schools and Colleges Assessment and Attainment Tables* at Key Stage 2. Test preparation is perceived by teachers to narrow the curriculum and for science in particular, as the inclusion of only those aspects of science likely to be included in paper and pencil tests is said to have reduced many aspects of investigatory science in Y6. Y6 teachers in Wales reported working to develop effective teacher assessment strategies, supported by the continued use of optional test materials, to inform summative judgements about pupils' levels of attainment in science at KS2. Practical science activities, including investigations, were reported as becoming an important feature of science lessons in Y6 in Wales. Although it is too soon to make predictions about the longer-term effects of the abolition of national testing on Y6 teachers and the teaching of science in Wales, early indications suggest increased opportunities for investigative science in Y6.

Introduction

Proponents of high-stakes testing for accountability purposes claim that it leads to higher standards in teaching and learning (Linn 2000; House of Commons 2008). If such testing is discontinued, what then happens to teaching and learning? One of the consequences of devolution in Great Britain is that Wales, which formerly was part of the same national testing regime as England, has withdrawn from national tests of 11 year olds (Key Stage 2) and 14 year olds (Key Stage 3). The tests at Key Stage 2 continue to be used in England and remain a key accountability instrument, whereas the government discontinued Key Stage 3 testing in England from 2009 following major problems with test marking in 2008.

The decision in 2004 of the Welsh National Assembly to abandon the national testing regime first introduced in 1991 can be viewed as a social experiment in which the impact of ending national tests can be monitored alongside their continuation in England. This decision was reached after the Daugherty review of assessment policy in Wales which found that, on balance, their continuation could not be educationally justified (Daugherty 2004; Daugherty 2008).

* Corresponding author. Email: sue.collins@brunel.ac.uk

This article presents the findings of an empirical study conducted using quantitative and qualitative methods to explore the effects of statutory national testing in science at Key Stage 2 (KS2) on Y6 teachers (i.e. teachers of children aged 10-11 in their 6th year of full-time schooling) and the teaching of science in England and the perceived effects of the abolition of national testing at KS2 in Wales.

Since end-of-Key Stage statutory national testing was introduced in England and Wales in 1991 there has been debate about the validity and impact of such tests and whether they can serve the many accountability purposes for which they are used (Brown et al. 1996; Wiliam 2001a; Stobart 2001; Tymms 2004; Black and Wiliam 2006; House of Commons 2008). Internal purposes typically include the monitoring of individual progress over time using National Curriculum levels of attainment to inform pupils and their parents. External purposes include monitoring the performance of schools and local authorities. It is these external, high stakes purposes which impact most on what is taught and how (Pollard et al. 2000; ARG 2001). One of the key validity concerns about the tests is their limited sampling of the National Curriculum. In science the curriculum strand ('attainment target') involving scientific investigation is only indirectly addressed in the written tests. In English the 'speaking and listening' attainment target is not represented at all in the tests. In high stakes contexts there is intense pressure on teachers to maximise pupils' attainment in areas of the curriculum to be tested to the detriment of those areas left untested. As a result, standards of pupils' attainment might falsely be seen to rise. Thus the validity of national tests as an adequate measure of attainment has been called into question (Wiliam 2001b; Tymms 2004; Wiliam and Black 2006; Stobart 2008).

The reliability of national tests has been challenged on the grounds of the inherent difficulty in assigning accurate levels of attainment to individual pupil achievement. Of particular concern at Key Stage 2 is sampling error, as the choice of test items from a broad curriculum will inevitably be a source of unreliability. The selection of other equally valid test items may well result in the attainment of different scores. This is compounded by other sources of unreliability: pupils' performance on the day and the accuracy of marking and the setting of levels of attainment. Estimates of the reliability of the levels awarded in National Curriculum tests suggest as many as one third of Key Stage 2 pupils may be awarded incorrect levels (Wiliam 2001a; Nichols and Smith 1998; Black and Wiliam 2006).

One consequence of high stakes testing at the end of Key Stage 2 is the tension for teachers, to balance teaching and assessment to meet the needs of the individual pupil, while trying to satisfy the requirements of a policy agenda that principally measures excellence in schools by test performance and associated performance tables. (Boyle et al. 2004; Boyle and Bragg 2005).

There is considerable empirical evidence of a culture of end-of-Key Stage test preparation in primary schools in England. This is not necessarily restricted to Year 6 – the year group in which statutory national tests for English, mathematics and science are administered – but may permeate the whole of Key Stage 2 (Years 3 to 6). Y6 pupils in some schools are unlikely to engage in activities outside the classroom such as field trips, drama productions or any activity that might divert attention from the main goal of learning and revision of the core subjects of the KS2 National Curriculum (Hopkins 2003).

One way of minimising the negative consequences of summative testing in England and Wales is through the use of teachers' assessments. Opinion is divided on the extent to which teachers' judgements are reliable in determining pupils' end of Key Stage levels of attainment. Arguments in favour of the use of teachers'

assessments for summative purposes focus on the unique position of teachers in constructing an accurate picture of individual pupil attainment across the spectrum of learning (Harlen 2005; ARG 2006). The main argument against is that if the results are to be used for accountability purposes, such as evaluating schools and their teachers, there will be pressure on teachers to generate the best possible results. This would be a potential source of bias. Teachers in England do make the final assessment at Key Stage 1 (Year 2) – but significantly these results are not published or used for national accountability. It was in this climate that the National Assembly of Wales, which had already stopped the publishing of performance tables, took the decision in 2001 to abolish national testing in English, mathematics and science for pupils aged 7 years (Key Stage 1) in favour of summative teacher assessment. Pressure from ministers and teachers' organisations in Wales to abandon national testing for pupils at KS2 and KS3 led to a review of the curriculum and assessment arrangements in primary and secondary schools by Awdurdod Cymwysterau Cwricwlwm as Aesu Cymru (ACCAC 2004), the Welsh equivalent of the Qualifications and Curriculum Authority (QCA) in England. Findings showed existing assessment arrangements to be inadequate for the purposes for which they were intended. In their place schools were urged to utilise summative teacher assessment, supported by optional test materials at KS1, 2 and 3 (Daugherty 2004).

The research reported here sought to explore the effects of this change on Y6 teachers and the teaching of science in Wales. At the same time the effects of the continuing use of statutory national testing on Y6 teachers and the teaching of science in England were investigated and contrasts and comparisons drawn where possible to inform representation by the Wellcome Trust to the House of Commons Select Committee on Education enquiry into Testing and Assessment in England in the spring of 2008.

Assessment of science at KS2

Science education in primary schools in England and Wales has a chequered history. Prior to the implementation of the National Curriculum in England and Wales in 1989 there existed considerable variation in the content, consistency and quality of science teaching, learning and assessment in primary schools (Swain 1989). Proposals by TGAT for an essentially formative system of assessment with teacher assessment central to the process were subsequently revised in favour national tests in science to support teachers in the moderation and calibration of their assessments (Black 1995).

Consistent attempts have been made over the past twenty years to promote the fundamental role of science process skills in the development of knowledge and understanding in science and to justify their inclusion in assessment procedures and practices for science. This aspect of pupils' learning in science does not feature strongly in national tests for science at KS2 in England (e.g. Duggan and Scott 1995; Harlen and Qualter 2004; Roden et al. 2007; Ward et al. 2008).

Aims and methodology

The aim of the research reported here was to explore the perceived impact of compulsory national testing in science on Y6 teachers and the teaching of science in England as well as the perceived impact of the abolition of such testing in science on Y6 teachers in Wales.

The methodology for this study involved both quantitative and qualitative data collection and analysis. A multi-faceted approach was used to explore the views of Y6

teachers about the challenges they faced in the teaching of science and of those who support them in primary schools in England and Wales.

In the autumn of 2007 and spring of 2008 data were gathered from two main sources using a sampling strategy designed to permit an interconnection between quantitative and qualitative data in the analysis. One source of data was a random sample of 600 Y6 teachers, primary science coordinators and primary headteachers – 300 in Wales and 300 in England – who responded to a telephone survey. The instrument was designed to explore the relationship between assessment procedures and practice in science at Key Stage 2 and strategies for teaching science utilised by Y6 teachers or recommended by science coordinators and headteachers.

The telephone survey instrument consisted of thirteen questions, eleven of which sought to elicit factual information about assessment procedures and practices in science at KS2 and views/perceptions of the effects of statutory national testing on the teaching of science in Y6. These questions required frequency ratings from *often* to *never* and Likert 5-point scale ratings ranging from *strongly agree* to *strongly disagree*. Two further questions invited open responses. In the first of these, respondents in both England and Wales were encouraged to identify the main challenges faced by Y6 teachers in the teaching of science. In the second, respondents in England were asked to consider the ways in which the teaching and learning of science in Y6 might alter if statutory national testing was abolished while respondents in Wales were asked to evaluate the effect, if any, of the abolition of testing on the teaching and learning of science in Y6.

The second source of data was focus group discussions with discrete groups of Y6 teachers and primary science coordinators, headteachers and secondary science teachers. The aim of this was to collect data on participants' experiences, views and opinions about the issue. Focus group discussions are appropriate for this as they enable exploration not only of what participants think but also why (Krueger and Casey 2009). The schedules were designed to pursue responses to telephone survey questions, probing underlying reasons for particular standpoints adopted or views expressed.

Over 100 schools in England and Wales were canvassed to take part in the qualitative strand of the research; as a result, eight focus groups were convened – 4 in London and 4 in Cardiff – involving a total of 74 participants (Table 1). Secondary science teachers were invited to take part in focus group discussions to provide insight into the nature and perceived value of information received from feeder primary schools about pupils' attainment in science at KS2.

Table 1. Sample of participants involved in focus groups in England and Wales.

England		Wales	
Number of participants		Number of participant	
Y6 teachers and science coordinators (2 groups)	25	Y6 teachers and science coordinators (2 groups)	18
Primary headteachers	8	Primary headteachers	11
Secondary science teachers	8	Secondary science teachers	4

Analysis of quantitative data involved the exploration of similarities and differences between responses in England and Wales, facilitating comparisons between sub-groups in each country where relevant. Statistically significant differences were identified by means of Chi-squared tests, with significance levels being set at five per cent.

Focus group discussions were recorded, replayed several times, coded and categorised for emergent themes, agreed by the research team.

Results and findings

Within this paper, the data reported are drawn predominantly from five themes that emerged from the research, as it is these data that provide greatest insight into teachers' contents and discontents with current assessment procedures, processes and practices in science in England and Wales. These five themes are:

1. Main challenges facing Y6 teachers in the teaching of science.
2. Assessment strategies for science used in Y6 classes.
3. Test preparation and the use of optional test materials for science in Y6.
4. Perceived and actual effects of the abolition of statutory national testing in science at KS2.
5. Perceived links between assessment and standards in science.

Main challenges facing Y6 teachers in the teaching of science

The main reported challenges for Y6 teachers in England and Wales can be summarised as *lack of time for science* and *lack of resources for science*.

A lack of time for science in Y6 was a common cause for concern among participants in the telephone survey and focus groups in both countries, though the implications for teachers in England were greater. The majority of Y6 teachers/science coordinator participants in England complained of the need for a selective review of those science topics most likely to be included in national tests, leaving them little option but to adjust their teaching strategies accordingly, illustrated by the following comment:

I can hear myself just going on and on not giving them a chance to talk about anything except to answer the occasional question – it's not what I call teaching.
(Y6 teacher: England)

There was a perception among focus group participants in England that the spiral curriculum, while supporting progression in pupils' learning in English and mathematics at KS2, was less effective in science where discrete topics failed to build on pupils' previous knowledge and understanding. Over a quarter (27%) of the telephone respondents in England thought it necessary to revise the entire KS2 science curriculum during the autumn and spring terms. For teachers in Wales this reduced to 19%. As one Y6 teacher explained:

They won't have done anything on teeth since Y3, so by Y6 they have completely forgotten the important bits and they only have the occasional words, not the proper vocabulary for it. (Y6 teacher: England)

A quarter of Y6 teachers/science coordinators interviewed in England emphasised that revision of the KS2 science curriculum did not simply entail repetition of work from previous years; it was necessary to further develop pupils' knowledge and understanding of key concepts within each of the Programmes of Study for science, presenting very real challenges for Y6 teachers.

There was consensus among participants in Wales that changes in assessment procedures at KS2 had not yet removed the need for comprehensive revision of the KS2 science curriculum in Y6. The need for Welsh teachers to formulate summative judgements about pupils' levels of attainment in science at KS2, supported by optional test materials, was said to encourage a revision schedule that closely resembled that reported by Y6 teachers in England. There were, however, subtle differences of approach described by Y6 teachers in Wales as indicated by the following comment:

We still revise all the topics for KS2, but we can take them a bit further now that we don't have to cram them for SATs¹, there is more time for hands-on science and the children like that. (Y6 teacher: Wales)

Within the more general category of *lack of resources*, participants in all focus groups containing primary teachers cited an absence of adequate class-based adult support for science as a major challenge. This view was shared by 26% of respondents to the telephone survey in England and 41% in Wales. Inadequate support was cited by three quarters of Y6 teachers/science coordinators in focus group discussions in England as a contributory factor in decisions to exclude almost entirely any investigative or other practical science activities in the spring term of Y6. The majority of Y6 teachers interviewed in England and Wales found the management of practical science activities in the classroom with 30+ pupils problematic. Two thirds of Y6 teachers interviewed in Wales highlighted as problematic the need for summative teacher assessment of pupils' attainment in all strands of the National Curriculum Programmes of Study for science, including the use of scientific process skills to complete hands-on investigations.

Assessment strategies for science used in Y6 classes

Despite the abolition of statutory testing in Wales there remained an emphasis on written answers to test questions in forming judgements about pupils' learning and attainment in science throughout Y6 (Figure 1).

Insert Figure 1 about here

Review of pupils' assessed work was used with equal frequency by teachers in England and Wales to provide information about pupils' progress at intervals throughout Y6 (Figure 1) and, while this strategy was reportedly deployed for summative purposes, it might be assumed to be potentially formative in informing teachers' subsequent planning and teaching in science.

In England the focus, understandably more than in Wales, was on use of *past SATs papers* ($\chi^2 = 18.6$, $df = 1$, $p < 0.001$) and *practice papers from published schemes* of work for science ($\chi^2 = 7.05$, $df = 1$, $p < 0.01$) to prepare pupils for statutory national science tests in the summer term (Figure 1).

There were no notable differences reported between the two countries in the formative assessment strategies used to support pupils' learning in science during Y6 (Figure 2). Over 90% of telephone survey respondents in England and Wales cited *discussion of learning intentions* as crucial in providing information for pupils about what they should know, understand or be able to do by the end of the lesson(s),

¹ SATs: Standard Assessment Tasks – commonly used term to describe statutory national tests in English, mathematics and science.

encouraging pupils to take some responsibility for their own learning. Linked to this strategy, 90% of respondents from both countries rated highly the use of *discussion of success criteria* with pupils where teachers typically summarise the key elements and actions required by pupils to fulfil the learning intention. Over 90% of respondents in England and Wales acknowledged the essential role of *feedback without marks with advice to pupils on their progress* to support pupils' learning in science in Y6.

Insert Figure 2 about here

Despite a reliance on past national test papers and practice papers to form judgements about pupils' levels of attainment in science, respondents to the telephone survey in England and Wales reported a commitment to the provision of feedback to pupils, based on the learning intention of the lesson, designed to encourage pupils to discuss their learning with the teacher and recognise for themselves what they needed to do to improve subsequent learning.

Formative strategies involving pupil *self-assessment*, reported as utilised frequently by 90% of respondents in England and 83% in Wales, again signalled a commitment among those surveyed to the development of independent learners in science. Although *peer assessment* was less frequently reported as a strategy commonly used among respondents, 72% of teachers in England and 61% in Wales perceived benefits to pupils' learning in science in providing opportunities to clarify the learning intentions and success criteria of the lesson(s) by scrutinising and/or assessing the work of another pupil.

Test preparation and the use of optional test materials for science in Y6

The schools in England sampled in this study all reported committing considerable time to science test preparation in Y6, particularly during the spring term. This confirms findings of other studies that focused test preparation is routinely undertaken at KS2 in response to the demands of high stakes testing (Black 1995b; Brown et al. 1996; Sturman 2003; Mansell 2007).

Teachers reported that test preparation typically begins in January each year – the beginning of the spring term in England – when the schools represented followed the common pattern of timetable reduction:

In the spring term it's as if only English, maths, science and PE exist, there's no more history and geography and certainly no more art or anything remotely creative like that. (Y6 teacher: England)

The inevitable consequence of a reduction in the scope of the curriculum in Y6 was to increase the time available for science in the spring term. However, as this additional time was reported by Y6 teachers to be utilised solely for test preparation purposes, the effect was a narrowing of the science curriculum to include only those aspects of science testable in paper and pencil tests.

Test preparation techniques varied, though all reflected the nature of national tests in focussing on careful reading of test questions, understanding of question content and accurate written responses to questions that tested science knowledge. Resources used for test preparation were said to include published revision schemes of work, based on National Curriculum Programmes of Study for science. The majority of Y6 teachers viewed such materials favourably as they provided a useful

structure for revision and a range of reinforcement materials including vocabulary lists to support test preparation.

While all participants in focus groups in England used or recommended the use of practice test papers in Y6, it was widely acknowledged that the process contributed little to pupils' knowledge and understanding in science, for example:

... they are retaining loads of information in their short term memory, but they're not really understanding it. (Science Coordinator: England)

Although statutory national testing was abolished in 2004 in Wales in favour of teacher assessment, primary schools were urged to make use of optional test materials for English, mathematics and science in Y6 to support summative teacher assessments in determining pupils' levels of attainment at the end of KS2 (Daugherty 2004). Optional test materials differ from statutory national tests in that teachers are encouraged to select appropriate tests from a bank of materials designed to assess pupils' knowledge and understanding of each of the Programmes of Study (PoS) of the science National Curriculum from level 2 to 6. Pupils' individual scores from each PoS are then aggregated using a weighting formula to provide a single overall level of attainment in science at KS2.

There was general agreement among focus group participants in Wales that optional test materials for science were valuable in informing or supporting professional judgements and in providing additional evidence of pupils' attainment in Y6. While consensus was reached concerning the potential of teachers' summative assessments to reflect more accurately pupils' levels of attainment in science at KS2, some participants agreed with the following statement, questioning assumptions that teacher assessment invariably provided accurate information:

I would challenge anyone to say that they are able to accurately assess a child's understanding through a couple of lessons on a topic when you have thirty children in the class all working away. Tests mean you can formalise it a bit more, gather evidence that they are saying this and saying that, but then answering questions correctly as well. Then you have hard evidence that they understand. (Headteacher: Wales)

The use of optional testing had been formalised in some primary-secondary clusters of schools. In two clusters represented, secondary schools required all feeder primary schools to administer the same optional tests for English, mathematics and science in the summer term of Y6 to maintain consistency of assessment at KS2. However, this was by no means the norm; the majority of schools represented reported the selection of optional test materials based on judgements about pupils' learning in science to support summative assessments of individual pupil attainment.

Teachers in Wales: views of the abolition of statutory national testing in science at KS2

The majority of Welsh respondents fully supported the changes at KS2, identifying positive implications for science teaching and learning within the cognitive and affective domains. Essentially, change had reportedly led to enhanced flexibility for teachers to determine lesson content, within the framework of the National Curriculum, and to adopt teaching strategies designed to respond to the learning needs of pupils. Practical, investigative science lessons, a part of the curriculum not directly assessed in the KS2 tests were an increasingly common feature of science education in Y6:

There has been a huge move in science education away from content onto skills. The children's attention is being drawn to Sc1 all the time, more, "What skills are you using here?" than "Can you tell me what's a conductor, what's an insulator?" (Y6 teacher: Wales)

With assessment based on teacher assessment, the majority of Y6 teachers said they had found it necessary to make adjustments to existing teaching strategies for science. These included consistent use of observation combined with focused questioning and discussion with individuals and small groups of pupils to determine levels of knowledge, understanding and process skills development. The implication of this change in emphasis was that whole class, teacher-led science lessons were no longer perceived as practicable. Teachers explained the need to utilise elicitation strategies such as concept mapping and pupil-designed tables to record predications, hypotheses and results of investigations:

Often before the science was lost really because it became just a handwriting lesson when they wrote up their experiments. They'd write, "We asked this question, and this is what we thought, and this is what we did, and this is what we found out". And that's fine if that's what you're focussing on, but you had to make sure they were writing down the right answers ... " (Y6 teacher: Wales)

However, this is not to imply that focus group participants considered pupils' recording in science to be of little consequence; on the contrary, many Y6 teachers viewed pupils' written work as a key indicator of their comprehension of questions and understanding of scientific concepts involved in investigations. They took issue with former practice that emphasised the importance of pupils' recording correct factual information regardless of their own findings, or the stage of the investigation they had reached before the end of the lesson.

Changes in assessment procedures and practices in Wales were not met with universal approval by focus group participants. Reservations, expressed by a quarter of teachers and headteachers, included a lack of confidence in teacher assessment to provide accurate summative judgements about individual levels of attainment in science. The majority of Y6 teachers felt that their assessments were considered unreliable by colleagues in their own school and by receiving secondary schools, prompting the need to provide additional evidence of attainment for each pupil – most commonly based on optional test scores. The need for stringent internal and external moderation of teachers' assessments at KS2 was highlighted by headteachers and science coordinators to establish consistency in levelling pupils' work in science within and across primary schools.

Headteachers participating in one focus group in Wales were circumspect in their support for the prominence of *Scientific Enquiry* (Sc1) in revised assessment procedures as they believed this presented challenges for some Y6 teachers who were unfamiliar with teaching strategies required to support investigatory science activities. While it was acknowledged that a programme of continuing professional development for Y6 teachers would go some way towards alleviating concerns, headteachers in this group were reluctant to compel teachers to attend training courses pending implementation of a new curriculum for science in September 2008. Uncertainty about curriculum development was an irritant for headteachers:

We are in a state of flux here with the new curriculum coming in 2008. The draft documentation we have had has Sc1 and that's all. It makes it difficult for headteachers to know quite where to start introducing things in our schools;

there's just not enough information filtering through ... to know how to proceed effectively for September 2008. (Headteacher: Wales)

Teachers in England: views on abolishing statutory national testing in science at KS2

Perceptions of teachers in focus groups in England about the likely effects on Y6 teachers and the teaching of science if England followed Wales in abolishing national testing at KS2, typified by the following comment, were strikingly similar to the views expressed by teachers in Wales as to what had actually happened:

I would love it if we had no more tests. No teaching to the test all the time, we could go round and question the children and gauge their level of understanding for ourselves and not be so prescriptive about what they are doing all the time. We'd be able to look at the child as a whole, gather information over a longer period of time. (Y6 teacher: England)

Another perceived advantage in the cognitive domain included opportunities to introduce a more balanced science curriculum at KS2 by giving more prominence to investigations and other practical science activities. Perceived positive affective outcomes were identified as 'more interesting science lessons' and 'more fun science'.

Perceived links between assessment and standards in science

There was overwhelming agreement among focus group participants in England and Wales that national test scores for science at KS2 failed to provide accurate information about pupils' achievements and attainment. Support for teacher assessment was unequivocal, but with the caveat that summative teacher assessment was subject to rigorous internal and external moderation informed by clearly articulated criteria. An additional stipulation was the availability of optional test materials to support teachers' in making accurate summative assessments and to provide additional evidence for interested parties to substantiate their judgements of pupils' level of attainment in science at the end of KS2.

Participants were consistently critical of the *School and College Assessment and Attainment Tables* (SCAATs), based on statutory national test results, as the dominant means of public accountability for individual schools. Since the abolition of SCAATs in Wales responsibility for the formulation of reporting strategies has fallen to individual schools. Headteachers involved in this study welcomed opportunities to design systems that included individualised pupil reports and parent-teacher interviews where individual progress and summative attainment at KS2 was discussed and targets set for development at KS3:

We don't use computer-generated reports anymore, it is all personalised now so that parents are left in no doubt about their child's progress. This is very important for us because we have a lot of non-English speaking parents in Cardiff and they need to be able to understand how their child is doing and they won't do that through trying to read stock phrases taken from the computer. (Headteacher: Wales)

Agreement was reached among headteachers in focus groups in England and Wales that outcomes expressed only in terms of percentages of pupils attaining Level 4 in English, mathematics and science at KS2 failed to reflect the real achievements of many pupils over six years of primary education. The majority believed that parents recognised this, preferring to base decisions about the quality of education provision of a school on OFSTED and ESTYN inspection reports.

Discussion

Our research question asked ‘what happens when high stakes testing stops?’ – the current situation at Key Stage 2 in Wales. The study reported on teachers’ perceptions of Key Stage 2 science teaching, both of those in England, where test results are published and are therefore high stakes, and in Wales where tests have been abolished and replaced with teacher assessment.

What the telephone survey and focus group data reveal is that ending high stakes testing in science has not immediately led to radical changes in teachers’ practice. Where there have been changes, they appear to be focused on giving more emphasis to practical science enquiry, a strand in the National Curriculum that is not directly addressed by the tests and so relatively neglected by Y6 teachers in England. As scientific enquiry is generally regarded as central to effective teaching and learning in science, this can be seen as a positive development which is likely to improve the validity of science assessment.

What is also apparent is that teachers in Wales continue to use optional tests to validate their judgements. This is not surprising after a decade of public reports of pupil achievement based on national test results, with teacher assessment relatively ignored. This continuance of testing may reflect both teachers’ own lack of confidence in their assessments, having been relatively de-skilled over years of high stakes testing, and the perceived demands of parents and local authorities for ‘objective’ evidence.

Alongside this, teachers in both England and Wales reported a lack of confidence in the accuracy of national test results in assessing pupils’ attainments in science. There was a general recognition that suitably moderated teacher assessment would provide a more dependable estimate. While Year 6 local moderation clusters are now being developed in Wales (which include secondary school teachers from the schools to which pupils will transfer) there was little evidence of teacher participation in these at the time of this study.

The reported changes to teaching and learning indicate a fuller coverage of the science curriculum, with scientific enquiry (Sc1) being given a significant role. This has not meant a radical change in teaching methods, in part because of the difficulties of conducting practical enquiries in classrooms with limited resources and support. There were significantly more teachers in Wales than in England who commented on the lack of classroom resources. This is not surprising given the lack of practical science in lessons in England as teachers selectively prepare for tests which do not include a practical element.

Teachers in England in this study indicated they would welcome an end to Key Stage 2 national tests. They anticipated that this would free them to include more investigative science in their teaching and to allow more detailed reporting to parents. They also indicated that they would continue with optional tests to support their judgements. These projections align with the changes in practice reported in Wales.

From this snapshot four years after the removal of national tests in Wales, teachers are reporting incremental yet significant changes to their teaching practices in science. Any generalisation to other subjects would have to be cautious – however for science the message seems clear: the sky does not fall in when high stakes testing stops and teaching and learning may be enriched.

Acknowledgements

We are grateful to the Wellcome Trust for commissioning this research.

Note

A report of the full findings is available at: <http://www.wellcome.ac.uk/About-us/Publications/Books/Education/index.htm>.

References

- ARG. 2001. *Summative assessment by teachers: Evidence from research and its implications for policy and practice*. Working paper 2: Assessment Systems for the Future. London: Institute of Education, University of London.
- _____. 2006. *The role of teachers in the assessment of learning*. Report for the Assessment Reform Group: www.assessment-reform-group.org.
- Black, P. 1995a. 1987 to 1995 – the struggle to formulate a National Curriculum for science in England and Wales. *Studies in Science Education* 26: 159-188.
- Black, P. 1995b. Curriculum and assessment in science education: The policy interface. *International Journal of Science Education* 17, no. 4: 453-469.
- Black, P. and Wiliam, D. 2006. The reliability of assessment. In *Assessment and Learning*, ed. J. Gardner, 9-26. London: Sage.
- Boyle, B., While, D. and Boyle, T. 2004. A longitudinal study of teacher change: What makes professional development effective? *The Curriculum Journal* 15, no. 1: 45-68.
- Boyle, B. and Bragg, J. 2005. No science today – the demise of primary science. *Cambridge Journal* 16, no. 4: 423-437.
- Brown, M., Taggart, B., McCallum, B. and Gipps, C. 1996. The impact of Key Stage 2 tests. *Education 3-13* 24, no. 3: 3-7.
- Daugherty, R. 2004. *Daugherty Assessment Review Group: Final Report May 2004*. Welsh Assembly, Cardiff: Wales.
- _____. 2008. Reviewing national curriculum assessment in Wales: how can evidence inform the development of policy? *Cambridge Journal of Education* 38, no. 1: 73-87.
- Duggan, S. and Gott, R. 1995. The place of investigation in practical work in the UK National Curriculum for science. *International Journal of Science Education* 17, no. 2: 137-147.
- Gipps, C., McCallum, B. and Brown, M. 1996. Models of teacher assessment among primary school teachers in England. *The Curriculum Journal* 7, no. 2: 167-183.
- _____. 2005. Teachers summative practices and assessment for learning: Tensions and synergies. *The Curriculum Journal* 16, no. 2: 207-223.
- _____. 2007. *The Primary Review: The Quality of Learning: Assessment alternatives for Primary Education*. Research Survey 3/4. Cambridge: University of Cambridge.
- Harlen, H. and Qualter, A. 2004. *The Teaching of Science in Primary Schools*. London: David Fulton
- Hopkins, D. 2003. Instructional leadership and school improvement. In *Effective Leadership for School Improvement*, eds. A. Harris, C. Day, D. Hopkins, M. Hadfield, A. Hargreaves and C. Chapman, 112- 135. London: Routledge Falmer.
- House of Commons. 2008. *Testing and Assessment Third Report of Session 2007–08. Volume I. Children, Schools and Families Committee*. <http://www.publications.parliament.uk/pa/cm200708/cmselect/cmchilch/169/169.pdf>
- Krueger, R. A. and Casey, m. A. 2009. *Focus Groups: A Practical Guide for Applied Research (4th Edn.)*. London: Sage Publications.
- Linn, R.L. 2000. Assessment and Accountability, *Educational Researcher* 29, no. 2: 14-16.
- Mansell, W. 2007. *Education by Numbers*. London: Politicos.
- Nicholls, P. D. and Smith, P. L. 1998. Contextualising the interpretation of reliability data. *Educational Measurement: Issues and Practice* 17, no. 3: 24-36.
- Pollard, A., Triggs, P., Broadfoot, P., McNess, E. and Osborn, M. 2000. *What pupils say: Changing*

- policy and practice in primary education*. London: Continuum.
- Roden, J., Ward, H. and Ritchie, H. 2007. *Primary Science: Extending Knowledge in Practice*. London: Learning Matters.
- Schagen, I. 1994. Multilevel analysis of Key Stage 1 National Curriculum data in 1991 and 1992. *Oxford Review of Education* 20, no. 2: 163-171.
- Stobart, G. 2001. The validity of national curriculum assessment, *British Journal of Educational Studies* 49, no. 1: 26-39.
- _____. 2008. *Testing times: The uses and abuses of assessment*. London: Routledge.
- Stohr-Hunt, P. M. 1996. An analysis of frequency of hands-on experience and science achievement. *Journal of Research in Science Teaching* 33, no.1: 101-109.
- Swain, J. R. L. 1989. The development of a framework for the assessment of process skills in a graded assessments in science project. *International Journal of Science Education* 11, no. 3: 251-259.
- Sturman, L. 2003. Teaching to the test: Science or intuition? *Educational Research* 45, no. 3: 261-271.
- Tymms, P. 2004. Are standards rising in primary schools? *British Educational Research Journal* 30, no.4: 477-494.
- Wiliam, D. 2001a. What is wrong with our educational assessments and what can be done about it. *Education Review* 15, no. 1: 54-73.
- _____. 2001b. Reliability, validity and all that jazz. *Education 3-13* 2, no. 1: 17-21.
- Ward, H., Roden, J., Hewlett, C. and Foreman, J. 2008. *Teaching Science in the Primary Classroom (2nd Edn.)*. London: Sage Publications.
- Wood, R. 1991. *Assessment and testing: A survey of research*. Cambridge: Cambridge University Press.

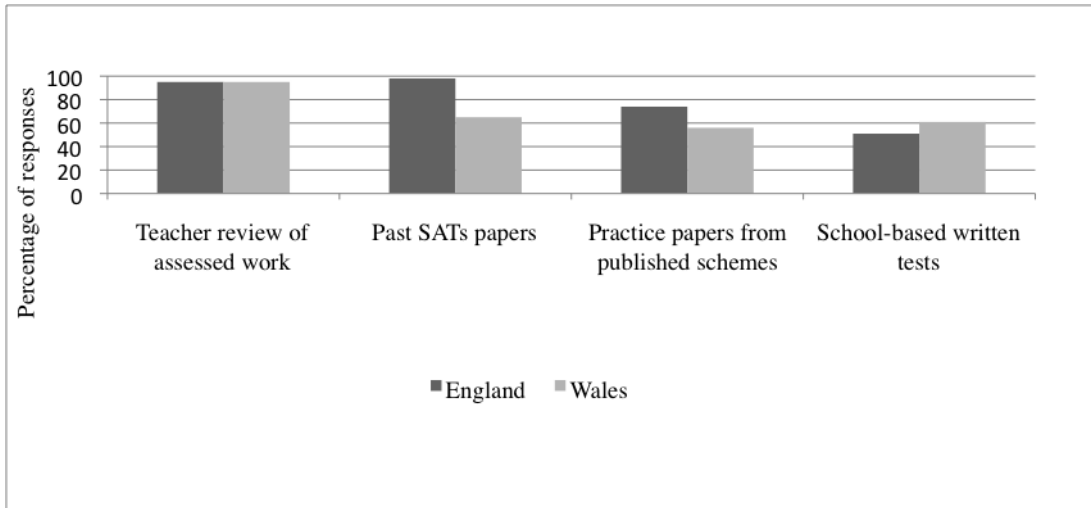


Figure 1. Summative assessment used in Y6 classes in England and Wales

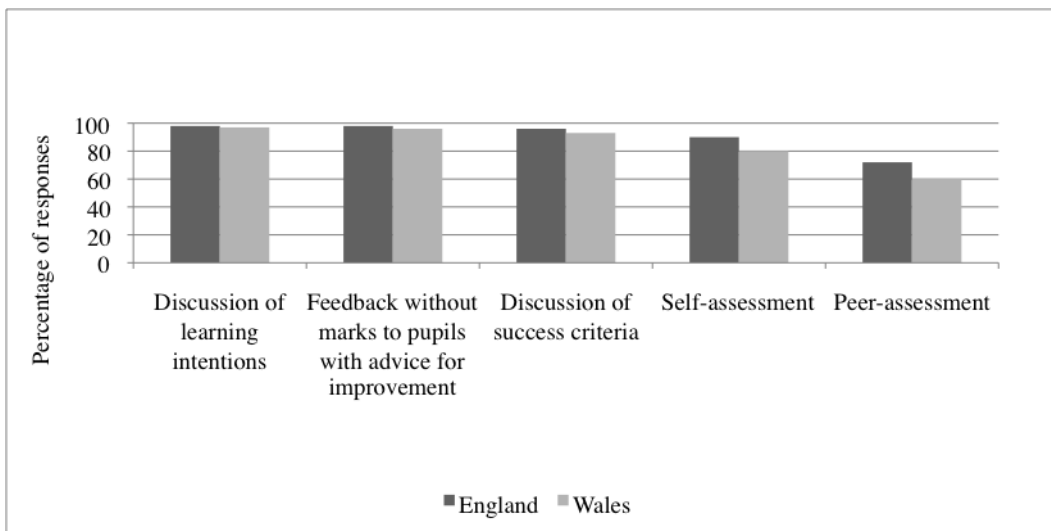


Figure 2. Formative assessment strategies used by Y6 teachers in England and Wales