

# A brief history of the assessment of practical work in science

Martin Hollins November 2009

## Overview

This article presents a brief account of the major themes and developments in the assessment of 'practical science' over the period from the introduction of GCSE and the National Curriculum, to the recent and proposed changes in both. Aspects covered include research findings, policy development and the tools and agencies for its implementation, and evidence of the effects of this on classroom practice.

## The introduction of GCSE

After several years of trials and pilots of courses, for example '16+', the GCSE examination was introduced in 1988. This combination of GCE 'O' level and CSE posed several challenges:

- This single examination for almost the complete ability range led to differentiated content and exam papers – 'tiering'
- CSE used coursework assessment substantially (not just for practical work) whereas 'O' level used only terminal examinations.
- The inclusion of elements of teacher assessment for high status examinations raised questions about reliability, and methods of moderation of teacher judgment.

At this time also there was an increased attention to processes and skills in science, in contrast to the more traditional emphasis on content knowledge and understanding.

The developers of the new examinations had available to them a range of sources of assessment tools:

- Assessment of Performance Unit (APU) had developed a range of materials in school trials through the 1980's. Researchers followed a cohort of 11-14 year-olds, and developed criteria with which to measure attainment and report on progress across all of science. In practical science there were two types of 'process-based tasks': *practical tasks* involving measuring, observing and describing relationships, and *written tasks* involving handling data, graphing and planning investigations. Reports of this work were published by SEAC<sup>1</sup>.
- The Graded Assessment in Science Project (GASP) further developed the work of APU on criteria for progression.
- The Technical & Vocational Education Initiative (TVEI) supported a wide range of curriculum and assessment innovations and in particular led to the development of portfolio assessment of experiences. This was later used in GNVQ assessment.

## The National Curriculum (NC) and Task Group on Assessment and Testing (TGAT)

At the end of the 1980's there was a substantial development of both the school curriculum and its assessment framework. Professor Paul Black, one of the directors of

APU, led TGAT, which proposed a radical new model of assessment that was based on a comprehensive range of broad criteria for each subject, set in a framework of levels to record progression. The assessment was primarily formative, by teachers, but externally moderated. It recommended the use of 'A mixture of standardised assessment instruments including tests, practical tasks and observations should be used in the national assessment system in order to minimise curriculum distortion' (TGAT<sup>2</sup>).

This national assessment system was designed for the new National Curriculum. The NC for science originally had 22 (then 17) attainment targets (ATs) with 10 levels of attainment at four key stages (KS). GCSE was the main assessment at KS4, but the exam boards rejected the NC scheme as unmanageable in its extent and complexity. National Curriculum testing at KS1,2 & 3 by teachers also ran into criticism, though the authors of the TGAT report argued that the criticism was politically inspired, rather than based on classroom evidence:

'The stereotyping of all expert opinion and evidence is very common in political argument. ... I fear that critics in government do not really understand the deep difference between those who want to break away from traditional tests in order to improve assessment and testing because they care about it, and those who want to abandon it altogether. The result of such indiscriminate arguments will be a return to tests of poor validity, dangerous unreliability and a heritage of damaging effects on pupils' learning.' (Black<sup>3</sup>).

Not surprisingly, there was considerable turmoil at all levels of the education system as a result of the major changes being introduced at that time. The political drive was to simplify the system to make sure it was manageable, so NC science was reduced to 4 ATs and National Curriculum testing at KS1,2 & 3 became a combination of written tests (SATs) and teacher assessment (for 'practical work')

### 'Practical work' in the National Curriculum

If the moves to specify the whole curriculum and its assessment revealed deep political divisions, the attempt to define the practical aspects of science caused comparable turbulence in the science education community. This was detailed in a substantial report from an education research team (Donnelly<sup>4</sup>). A brief chronology of the changes will have to suffice here:

- 1989 NC order: 'AT1' as one of two profile components (the other had 16 ATs). Called 'The Exploration of Science' it had a strong investigative emphasis
- 1991 Revised NC: 'Sc1', which incorporated AT17 'the nature of science', as the ATs were reduced to 5 and then 4. Called 'Scientific Investigation', the model was strong on variable handling, and was criticised for not being applicable across the sciences.
- 1995 Dearing Review resulted in a simplification of NC and 'Sc1' became 'Experimental and Investigative Science'
- 2000 Revised NC incorporated additional material called 'ideas and evidence' into 'Sc1' which became 'Scientific Enquiry'
- 2005 KS4 revision with 'AT1' becoming 'How Science Works'
- 2008 KS3 revision to match the programme of study at KS4

## Assessment developments in the 1990's

Inevitably, the result of these changes in the NC was that assessment in schools of Sc1 at KS3 & KS4 in the early 90's was very variable. Reports noted that at KS3, Sc1 practice was weaker than in other aspects and at GCSE there was a narrow emphasis on coursework requirements (Ofsted<sup>5</sup>). There were further changes to the assessment systems at that time, to try to simplify and clarify what was expected of teachers.

It was decided that Sc1 at KS3 to be assessed only by teachers, as the proposed auditing system was abandoned. There was official support for teacher judgment of pupils' performance at KS3, in the publication of annotated pupils' work, exemplifying levels of attainment and showing how 'best fit' for a NC level could be arrived at (SEAC<sup>6</sup>).

The value of teacher assessment (TA) was promoted by the work of the Assessment Reform Group with publications such as 'Inside the Black Box'. Their research highlighted the importance of assessment to pedagogy, recommending strategies such as feedback to pupils and peer and self-assessment. (Black and Wiliam<sup>7</sup>)

The AKSIS project reviewed practice in schools of Sc1 at KS2 & KS3 found a limited range of approaches, with 'fair testing' dominant, reflecting the earlier formulation of Sc1 as variables-driven. AKSIS reported that there was a need for better moderation of practical assessment judgements at KS2 & KS3. They also recommended a better match between assessment procedures at KS3 and KS4, as the GCSE system of criteria, leading to marks and grades, was by now very different from the NC level descriptions and 'best fit' by teacher judgement.

The initial coursework moderation arrangements for GCSE had been judged inadequate, partly due to different interpretations of 'investigating' and of criterion-referencing, so exam boards agreed a common framework and criteria for investigative work: (Plan, Observe, Analyse, Evaluate). These used 'performance descriptors' which were 'criterion-based' and required teacher judgment, with moderation. Despite these changes however, schools continued to find the GCSE requirements complex (Donnelly<sup>4</sup>)

## Preparing for the Twenty-first Century

The moratorium on NC changes between 1995 and 2000, which was agreed by the government after the Dearing review, allowed time for development work. Some of the key influences on the curriculum were:

- Schemes of Work for KS1,2&3 published. These provided substantial coverage of Sc1 but with little detailed advice on assessment practicalities (QCA<sup>8</sup>).
- The influential 'Beyond 2000' review made radical proposals for extending the nature of assessment in science, to include reports of contemporary science and its practice, but did not directly address assessing 'practical work'.

- The ‘Twenty-first Century Science’ project was developed from recommendations of this report and led to a revised KS4 NC with broadened Sc1 into ‘How Science Works’

In response to reports that classroom practice of Sc1 was often formulaic, intended for assessment purposes and rarely done as a normal part of the lesson (Donnelly<sup>9</sup>), the planning of the GCSEs based on the changes to the KS4 NC attempted to free up the assessment of AT1. The new specifications drew on a range of research including:

- The assessment of ‘Ideas and Evidence’ by objective and structured questions (Osborne & Ratcliffe<sup>10</sup>)
- The development of written tools to assess skills and procedural understanding (Gott & Roberts<sup>11</sup>)

At KS3, a substantial teacher assessment resource was published. Assessing Progress in Science (APS) contains exemplars of pupils work with levelled assessment comments, advice on pedagogy and additions to the level descriptions as assessment criteria (QCA<sup>12</sup>). This project was then further developed into Assessing Pupil Progress (APP) which has a criteria framework to complement the current KS3 level descriptors, assessment advice and exemplars (DCSF<sup>13</sup>). It now replaces SATs, and is the subject of a training programme by Secondary National Strategy, Science (alongside APP for English and Mathematics).

## Recent and forthcoming GCSE changes

In 2004 there was a significant change to the KS4 requirements: it was no longer allowed to disapply students from being entered for a science examination because of having a work-related curriculum, as all KS4 courses were required to include work related learning. This meant in effect that exams had to be accessible to a wider range of candidates than formerly. Perhaps surprisingly, at about the same time the GNVQ examination was phased out, after many alterations to its assessment system. The range of alternatives with an applied or vocational orientation was however extended and includes BTEC, OCR Nationals and Applied Science GCSE.

In 2006 the new ‘Twenty-first Century’ GCSEs were introduced with a wide range of assessments of ‘How Science Works’ to replace and extend the Sc1 assessments. A small proportion of the marks was given for unmoderated teacher assessment of practical tasks; most practical work was assessed in a combination of practical tasks and written answers, usually done under exam conditions. This type of assessment is now called Controlled Assessment and is a requirement at A level and for other GCSE subjects (QCA<sup>14</sup>). However reviews of the first set of examinations in 2008 revealed concerns about consistency of standards in GCSE science (SCORE<sup>15</sup>) which has led to calls for changes to exams in 2008/9 and criteria for 2011 (Ofqual<sup>16</sup>). In response to the recent consultation on new GCSE science criteria (QCDA<sup>17</sup>), ASE argued for more guidance to be provided ‘on rigorous and different approaches by which practical work can be assessed, without reverting to the unimaginative and constraining Sc1 assessments of the past’. An uncanny echo of Paul Black’s fears of nearly 20 years ago about assessment limiting the

curriculum! The new criteria from which the awarding bodies will write their specifications, are to be published by the end of 2009. specifications need to be accredited by the Summer of 2010 for first teaching from September 2011. So now is the time when the types of GCSE assessments for the next few years are being defined.

## Some other aspects

This brief review has not covered other areas of interest which may be helpful in considering the assessment of practical work:

- The assessment of vocational and work- related courses such as BTEC, Applied GCSE, and Diplomas where the context ( not purely the content), of the experiences in the course are significant.
- The language of measurement and its use in the terminology in exams (ASE/Nuffield<sup>18</sup>)
- Experience beyond England eg Scotland's sampling assessment of practical activities by field officers and the freedom of choice of the type of assessment of 'practical work' in IGCSE science examinations (CIE<sup>19</sup>).

## Conclusions

The following issues concerning good assessment practice, can be seen to have been important in this brief history and appear to the author to be relevant to current decisions:

- How can the distortion of the curriculum by assessment methods be minimised?
- Since products are easier to assess reliably than processes, what is the minimum of process assessment, to ensure practical assessment is valid?
- How much choice should teachers and students have in their assessment instruments and assessed work?
- Will the current developments of APP and Controlled Assessments improve both the reliability and validity of the assessment of Practical Science?

This account has not had the scope to detail technical issues; it has perhaps emphasised the political over the pedagogical and the science has been social rather than natural – such, it seems to the author, are the priorities which are revealed by this account.

## References

1. SEAC: *Assessment Matters eg No.5: Profiles and Progression in Science Exploration* SEAC/APU1991
2. TGAT: *Task Group on Assessment and Testing Report, paragraph 59* DES 1988
3. Black: *Presidential address to the British Association for the Advancement of Science* Paul Black1992

4. Donnelly: *Investigations by Order* Jim Donnelly et al Leeds 1996
5. Ofsted: *Science Review of inspection findings 1993-4, and Annual Report; Standards and Quality 1993-4* Ofsted 1995
6. SCAA: *Exemplification of Standards* SCAA 1995
7. Black and Wiliam: *Inside the Black Box* Kings College, London 1998
8. QCA: *Science: a scheme of work for Key Stage 3* QCA 2000
9. Donnelly *Secondary science teaching under the national curriculum* School Science Review, ASE March 2000
10. Osborn and Ratcliffe: *Developing effective methods of assessing Ideas and evidence* School Science Review, ASE June 2002
11. Gott and Roberts: *A written test for procedural understanding: a case study into pupils' performance*, European Science Education Research Association, Holland 2003
12. QCA: *Assessing Progress in Science eg KS3 Unit 1 Scientific Enquiry* QCA 2003
13. DCSF: *Assessing pupils' progress in science at key stage 3 eg Standards file* DCSF/QCA 2009
14. QCA: *Controlled Assessment* QCA 2008
15. SCORE: *GCSE Examinations Project* [www.score-education.org/2projects/gcse\\_science.htm](http://www.score-education.org/2projects/gcse_science.htm)
16. Ofqual: *GCSE science monitoring report 2007 – 2008* [www.ofqual.gov.uk/2220.aspx](http://www.ofqual.gov.uk/2220.aspx)
17. QCDA: *Consultation on GCSE science criteria for first teaching in 2011* [www.qcda.gov.uk/21855.aspx](http://www.qcda.gov.uk/21855.aspx)
18. ASE/Nuffield: *The language of measurement: terminology used in school science investigations* ASE/Nuffield 2010
19. CIE: *IGCSE Science – Combined* [www.cie.org.uk/qualifications/academic/middlesec/igcse/subject?assdef\\_id=884](http://www.cie.org.uk/qualifications/academic/middlesec/igcse/subject?assdef_id=884)