

The English assessment regime: how consistency and standards stifle innovation and improved validity for the assessment of mathematics

Sue Pope

Liverpool Hope University

This paper describes the national assessment regime for mathematics in England for 5 to 16 year olds which is the basis of school accountability. Most of these assessments comprise timed written tests or exams that are designed to assess the statutory national curriculum programmes of study. For pre-16 learners the assessments are developed nationally and teacher assessment is reported alongside test outcomes. There is considerable evidence that teachers are over-reliant on the tests and adjust their assessment to match that of test outcomes. At age 16 independent commercial organisations (awarding organisations) develop public examinations (GCSEs) in a regulated market place. There is fierce competition between awarding organisations to gain and maintain market share. The regulatory system for the development of tests and exams and maintenance of standards is rigorous but restricts innovation and improvements in validity.

Assessment; accountability; regulation; measures of performance

Introduction

The extent of national testing increased considerably in England following the introduction of the National Curriculum (NC) in 1988 for 5 to 16 year olds. More recently it has begun to decline but there is still concern from the Office for Standards in Education (Ofsted 2008) that 'teaching to the test' dominates much classroom practice and teachers are over reliant on testing as opposed to other forms of assessment.

In this paper I describe the current system of assessment in England for 5 to 16 year olds and how it came about. I explain the processes in place to maintain consistency of standards over time and discuss the implications of these for innovation that might help to improve the validity of these assessments.

Background

The original NC included a ten level scale of attainment which was intended for use across the entire age range. Children's levels of attainment in the core subjects of English, mathematics and science were expected to be reported at the age of 7, 11 and 14 (known as the end of key stages 1, 2 and 3 (KS1, KS2, KS3) respectively). The level of attainment was to be determined by teachers using their knowledge of children's achievements on the basis of 'best fit' to the level descriptions in the NC. The levels were designed in such a way that the 'average child' might be expected to make one level of progress in two years. Assessment of KS4 was through GCSEs which are graded A (level 9) to G (level 4) with grade C (level 7) acting as the gatekeeper for progression to many vocational, professional and academic courses.

Subsequent revisions of the NC in the early 1990s replaced levels 9 and 10 with exceptional performance (EP), and restricted the use of levels up to age 14.

Key Stage	Age	Expected level of attainment for the 'average' child at the end of the key stage
1	5-7	2
2	7-11	4
3	11-14	5-6
4	14-16	7 (GCSE grade C/D)

Table 1: English NC expected levels of attainment.

National exemplification of standards materials were provided by the Schools Examination and Assessment Council (SEAC) to support teachers in making these end of key stage assessments (SEAC 1992). Teachers were encouraged to work within and across schools and phases to moderate their judgements and develop a robust understanding of what attainment at different levels looked like.

No funding was provided for teacher moderation of level judgements and concerns about consistency and rigour led to the introduction of national testing. In the early 1990s, national tests were initially introduced for KS3 mathematics, then KS3 English and science and KS2 mathematics, science and English. The first tests were criterion referenced and marked by teachers. Test results were expected to be reported alongside teacher assessment. However, no additional funding was made available to teachers to undertake test marking and teacher union action resulted in the external marking of KS2 and KS3 tests. The only concession won by one teacher union was that scripts would be returned to schools.

However, the time taken for external marking of tests meant that the potential of the tests to inform teaching and learning was lost, scripts were often returned after schools had finished for the summer and in the following September children would either transfer to a new school (for the start of KS3) or to a new teacher (for the start of KS4). Teachers tended to wait until test results were received before reporting their assessments, as if the teacher assessment didn't match the test outcomes they were criticised for over or underestimating children's level of attainment (Dainton et al. 1999).

KS1 English (reading and writing) and mathematics tests and tasks (for level 1) were marked by teachers and monitored by local education authorities. Since 2005 the use of reading, writing and mathematics tests and tasks remains statutory but the outcomes now inform teacher assessment which is reported for English and mathematics. Local authority monitoring has switched from the accuracy of test marking to the effectiveness of teacher assessment arrangements, and the system is nationally monitored by the Qualifications and Curriculum Authority (QCA, which became the Qualifications and Curriculum Development Agency (QCDA) in 2009).

National testing came at a time of increasing accountability for schools and led to the high stakes performative culture that now dominates English education (Ball 2003). This includes the use of national, local, school and individual targets and publication of results in 'league tables'. The school inspection regime focuses on school results relative to national norms and expectations. Schools that fail to meet government targets, regardless of the nature of their cohort, are named and shamed and given 'notice to improve' or put in 'special measures': failure to improve ultimately leads to closure.

In 2008 KS3 tests were abandoned, as were national targets and 'league tables' for 14 year olds. Teacher assessment is still collected nationally and a system of national sampling is under consideration. In 2009, following a review of national

assessment arrangements (DCSF 2009), KS2 science tests were scrapped in favour of national sampling. The national sampling process for science is still in development but in 2010 a sample of KS2 children completed science tests, resulting in a drop of attainment by seven percentage points (81% at level four and above from 88% in 2009, DfE 2010). The mathematics community, led by the Advisory Committee for Mathematics Education (ACME), argues strongly against national tests for KS2 mathematics (ACME 2008). Many primary schools abandon teaching any new mathematics and begin test preparation as early as January because of the high stakes nature of the KS2 tests that are currently taken in May. However, a recent review of KS2 assessment and accountability stated:

We have not received any evidence to suggest that there are significant issues with an externally-marked mathematics test. We recognise that it is relatively straightforward to create a valid and reliable test of mathematics, and we feel that the current mathematics tests achieve this. We believe that it is legitimate to use a test to establish how well a pupil can perform a range of mathematical operations within a finite period of time. We recommend that mathematics should continue to be subject to externally-marked testing. (Bew 2011, 63)

Assessment at Age 16 - GCSE

In 1988 the GCSE was introduced as a single qualification for the vast majority of 16 year olds. It replaced GCE O levels² and CSEs³. GCSE was expected to maintain standards by ensuring that grades C and above corresponded to O level passes and grades D to G corresponded to CSE passes (grades 2 to 5). There is evidence to suggest that grades were inflated in the new GCSEs (Hodgen et al. 2010). When first introduced GCSE could comprise up to 100% coursework (teacher assessment) and consortia of schools could work with an exam board to develop their own syllabi.

The examinations available to 16 year olds were originally through exam boards, many of which had started within universities that had devised entrance examinations for undergraduate courses. Over the years the number of exam boards has decreased as smaller organisations have merged, and there are now just three main awarding organisations in England who provide GCSEs within a 'market place'. These high stakes qualifications are regulated by the Office of Qualifications and Examinations Regulation (Ofqual). The system for developing curriculum, assessment and qualifications in England and its increasing centralisation and narrowness is described by Isaacs (2010). A GCSE is intended to be a two year course and, for NC subjects, should assess the entire KS4 programme of study (Ofqual 2004).

Ofqual is responsible for the maintenance of standards of live qualifications. A code of practice describes the processes awarding organisations must have in place to ensure consistency of standards in both exam paper development and awarding the qualification. The regulator monitors awarding organisation processes and undertakes 'standards over time' scrutinies looking at exams and candidate scripts to see how different awarding organisations compare with one another.

GCSEs tend to be written by a small number of examiners approximately a year in advance and their draft papers are subjected to various quality assurance processes including a Question Paper Evaluation Committee (QPEC) meeting where each paper is subject to expert scrutiny (*shredding*) to ensure that it will be accessible and is consistent with the demand of previous papers. This means that novel items

² GCE O levels were introduced in 1951 as subject-based assessments for the top 20% of 16 year olds.

³ CSEs were introduced in 1965 to enable a broader population to gain a qualification at age 16, the highest grade being notionally equivalent to an O level pass.

(e.g. those which expect candidates to make decisions about what mathematics, information and strategy to use) are altered to include more scaffolding and require less thinking.

National curriculum tests

Ofqual is also responsible for quality assuring national assessments. NC tests have always been developed by a single contractor and for a number of years they were developed 'in house' at QCA. There is a regulatory framework (Ofqual 2009a) which describes precisely the nature of the tests including how many marks for each level and each aspect of the curriculum. There is a general expectation that the threshold mark for a particular level on a particular suite of papers does not change by more than one or two marks year on year. The development process for NC tests is significantly different to that for GCSEs and is much more expensive. There is a lead time of at least two years. Initially 200% of material is generated and, after scrutiny by subject experts and practicing teachers, it is trialled with a representative sample of students (known as pre-test one). This provides a facility (the proportion of students that are successful) and discrimination value (how students working at different levels perform) for each item. The outcomes of pre-test one form the basis for developing a suite of papers for a further larger scale pre-test. All changes to items and mark schemes after pre-test one are minimal so the statistics can be used. Novel items which expect learners to make decisions about what mathematics, information and strategy to use are trialled with students preparing for the current assessments. As responses are not as statistically robust as more traditional items, few survive pre-test one. Where they do survive items are likely to have become shorter and more structured.

The implications of regulation

The purpose of regulation is to ensure that tests and qualifications are valid, reliable, manageable, comparable over time and across awarding organisations, and minimise bias. The regulators are responsible for ensuring standards are maintained and there is public confidence in regulated assessments. However, the rigorous processes in place to ensure consistency of standards actually conspire to stifle innovation (Wolf 2009). The 'market place' for GCSEs means that awarding organisations will not want to do anything that might jeopardise market share. For mathematics this means that questions tend to be highly structured and require little initiative from candidates. Where examiners attempt more innovative questions these will inevitably be remodelled through the QPEC process. Similarly, as national curriculum test questions are trialled with children who are being taught for the current tests the outcomes on novel items tend to be less predictable and consequently less statistically robust, which means that novel items rarely survive the pre-test process. Given the requirement for minimum variation in threshold marks for each level the tests have become increasingly predictable over time.

The assessment of mathematics process skills

The NC for mathematics is principally constructed around content: number and algebra; geometry and measures and statistics and probability. Mathematical process skills are articulated in the curriculum as 'using and applying mathematics' (UAM) which includes problem-solving, reasoning and mathematical communication. At

GCSE UAM was assessed through coursework until 2009, when following a national consultation coursework was removed and assessment of UAM was to be incorporated into exams. Few teachers noticed any changes in the exams for summer 2009 awards, although there was a notable decline in girls' performance.

In NC tests UAM was not included until 2003 when a small number of marks (approximately 10%) were allocated to UAM. Items that assess UAM are designed to assess mathematical content first and foremost but include an element of problem-solving or require an explanation.

The high stakes nature of the assessments (national curriculum tests and GCSEs) means that they tend to comprise items for which the outcomes are relatively easy to mark – there is one right answer and the method is relatively obvious. Teachers tend to believe that the assessments are the curriculum and certainly all that needs to be taught (Dainton et al. 1999, Wiliam 2001). Despite incorporating UAM into GCSE and NC tests there is little evidence that this has actually impacted on classroom practice. Many teachers think that UAM is an add-on to the curriculum rather than something to be integrated (Ofsted 2008).

In the latest revision of the NC for KS3 and KS4 UAM has been strengthened as Key Processes (QCA 2007). Key Processes draw on the mathematical process skills set out in the 2003 PISA framework for mathematics (OECD 2003) and include representing, analysing, interpreting, communicating and evaluating.

KS3 tests and GCSEs needed to change to reflect these curriculum changes. However KS3 tests were abandoned in 2008 so no changes were realised, although new optional tests include some items which require students to think for themselves and make choices about what mathematics to use and how to tackle a problem (QCDA 2010). Changes to GCSE were piloted but the independent evaluation of the pilot found that whilst there was some evidence of change it was limited and many items continued to lack authenticity, complexity and genuine opportunities for problem solving and reasoning (Noyes et al. 2008).

The new GCSE introduced from September 2010 (first awards in 2012) has a much greater emphasis on mathematical thinking and problem solving. Compared to the current GCSE which has just 20% of the marks allocated for UAM, 50% of the marks are allocated for selecting and applying mathematics in context and devising strategies to solve problems (Ofqual 2009b). Whilst the accredited materials look different, it remains to be seen to what extent the live assessments are different to the current GCSE. Awarding organisations will be expected to maintain year on year standards whilst working to the new subject criteria, which represent a significant and sudden change.

Conclusion

The rigorous systems in place for developing and regulating national assessments for 5 to 16 year olds in England whilst ensuring comparability and reliability are unable to accommodate innovation. Adjusting assessments to improve validity, even when the curriculum undergoes substantial change or inspection evidence identifies a significant weakness in the system, as is the case for mathematics, is highly problematic.

Despite substantial pilots of changes to GCSE the potential for change was barely realised and awarding organisations, constrained by market forces and the need for year on year comparability, are likely to make as little change as they can get away with as the new GCSEs are rolled out. Whilst national tests could undergo a

gradual process of change, the development process makes it difficult for novel items to make it through to the tests.

Consequently, the assessment of mathematics which drives teaching and learning is unable to change significantly. Narrow, predictable assessments enable 'teaching to the test' in England's high stakes performative education culture.

References

- ACME. 2008. *ACME's Views on an Alternative Regime for Key Stage 2 Assessment in Mathematics*. London: ACME. www.acme-k.org/downloaddoc.asp?id=182. May 2011.
- Ball, S. 2003. The teacher's soul and the terrors of performativity. *Journal of Education Policy* 18 (2), 215-28.
- Bew, P. (2011) *Independent review of KS2 testing, assessment and accountability*. London: Department for Education.
- Dainton, S., G. Evans, S. Ford, J. Moss and S. Pope. 1999. *An evaluation of the 1999 KS3 tests in English, Mathematics and Science*. London: ATL, 27-49.
- DfE. 2010. *School Performance Tables*. www.education.gov.uk/performance/tables/ June 2011.
- DCSF. 2009. *Report of 'Expert Group in Assessment'*. London: HMSO.
- Hodgen, J., M. Brown, D. Küchemann, and R. Coe. 2010. Mathematical attainment of English secondary school students: a 30-year comparison. *British Educational Research Association (BERA) Annual Conference*, University of Warwick.
- Noyes, A, G. Wake, P. Drake, and R. Murphy. 2008. *Independent Evaluation of the Mathematics Pathways Pilots, Stage 3 Report*. available on request from QCDA.
- Isaacs, T. 2010. Educational Assessment in England. *Educational Assessment: principles, policy and practice* 17 (3): 315-34.
- OECD. 2003. *The PISA 2003 Assessment Framework: Mathematics, Reading, Science and Problem Solving Knowledge and Skills*. Paris: OECD.
- Ofqual. 2004. *The statutory regulation of external qualifications in England, Wales and Northern Ireland*. [www.ofqual.gov.uk/files/6944_regulatory_criteria_04\(1\).pdf](http://www.ofqual.gov.uk/files/6944_regulatory_criteria_04(1).pdf). April 2011
- Ofqual. 2009a. *National Curriculum Assessments Regulatory Framework*. Ofqual/09/4123. Coventry: Ofqual. www.ofqual.gov.uk/files/2009-02-nc-assessments-fr.pdf. May 2011.
- Ofqual. 2009b. *GCSE subject criteria for mathematics*. Coventry: Ofqual. www.ofqual.gov.uk/files/2009-03-gcse-maths-subject-criteria.pdf. April 2011
- Ofsted. 2008. *Mathematics: Understanding the Score*. London: HMSO.
- QCA. 2007. *The National Curriculum Key Stages 3 & 4*. curriculum.qcda.gov.uk/key-stages-3-and-4/ June 2011.
- QCDA. 2010. *Optional tests for mathematics year 7, 8 and 9: New materials for 2011* http://orderline.qcda.gov.uk/gempdf/1849625859/QCDA105038p_OptionalMaths_years789_2011.pdf. April 2011.
- SEAC. 1992. *Pupils' work assessed: Mathematics*. London: HMSO.
- William, D. 2001. *Level Best? Levels of attainment in National Curriculum Assessment*. London: Association of Teachers and Lecturers.
- Wolf, A. 2009. *The role of the state in educational assessment*. Speech delivered at the Cambridge Assessment Annual Conference, October 19, Cambridge, UK.