
Happy Fiasco! The National Curriculum Tests of 2008, and After

PATRICK YARKER

ABSTRACT This article, which draws heavily on the Sutherland Inquiry report into the delivery of National Curriculum testing in 2008, outlines important aspects of the failure that year to report test-scores on time, considers the extent to which ministers might have been held more accountable and reviews the state of the long struggle to replace the current form of NC testing with less-damaging alternative forms of assessment.

What can go wrong, will go wrong.

Competitive Dialogue: Final Bid Presentation, ETS, 18 July 2006

Perhaps it was Dr Ken Boston himself, Chief Executive of the Qualifications and Curriculum Authority (QCA) and a partisan for the technological modernisation of England's school-testing system, who submitted in June 2007 an unsigned 'further memorandum' to the Education and Skills Select Committee, chaired by Barry Sheerman, MP. This memo lauded the possibilities afforded by online testing and marking. It argued that: 'the wider use of electronic technology has the potential to benefit each element of the test system ranging from how we construct tests through to the delivery of results' (QCA, 2007). Online testing and marking would deliver more detailed information about student-performance, and deliver it more quickly and reliably to schools. It would enable test-markers to be trained and supported better. It would host mixed-media and so make National Curriculum test-questions yet more engaging. Schoolchildren, already familiar with the new technologies online testing could exploit, would be guaranteed 'a more authentic experience' of being tested (QCA, 2007).

Less than a year later, with the cohort of 2008 facing their Key Stage 2 tests, Sheerman's committee published its lengthy report on assessment in schools. The 'authenticity' of the testing-experience, together with the benefits of online testing, it chose to pass over in silence. But it had much to say about how the system of national testing actively damaged the educational opportunities of some children, and prevented teachers from doing their proper job:

[W]e find that the use of national test results for the purpose of school accountability has resulted in some schools emphasising the maximisation of test results at the expense of a more rounded education for their pupils. A variety of classroom practices aimed at improving test results has distorted the education of some children, which may leave them unprepared for higher education and employment. We find that 'teaching to the test' and narrowing of the taught curriculum are widespread phenomena in schools, resulting in a disproportionate focus on the 'core' subjects of English, mathematics and science and, in particular, on those aspects of these subjects which are likely to be tested in an examination. Tests, however, can only test a limited range of the skills and activities which are properly part of a rounded education, so that a focus on improving test results compromises teachers' creativity in the classroom and children's access to a balanced curriculum. (House of Commons Children, Schools and Families Committee Testing and Assessment Third Report of Session 2007-08, Volume I p. 3)

When the official written response of the DCSF to the Committee's findings eventually arrived it took the form, so the *Guardian* reported, of a blanket denial:

[DCSF] simply dismisses all the evidence. It claims that [the NC tests] should not be stressful, and therefore they aren't. It says testing should not distort education, and therefore it doesn't. (Russell, 2008)

Assuming the ostrich-position was perhaps only to be expected from ministers, given what had been going on behind the scenes as the testing-process took its course across the first half of 2008. In May the Education Committee's report had sparked a day's bad headlines. In June and July the unthinkable happened: New Labour's school testing-system imploded.

Scrapped

The educational and political repercussions of this spectacular and, for the government, humiliating debacle have yet to play out fully. The initial consequences were ground-shaking. The publication of all test-results for 2008 had to be delayed. Some schools did not receive their scores for months. A

small number of students have never received them. The company appointed to run NC testing was stripped of its five-year contract, and most (though not all) of the 156 million pounds of public money awarded it taken back. Dr. Boston tried to resign, as did his counterpart at the National Assessment Agency David Gee, but the duo found themselves suspended for months before their resignations were finally accepted in April this year. QCA was replaced by a new organisation, the Qualifications and Curriculum Development Agency (QCDA). Appeals by schools against their test-results increased some forty per cent beyond the figures for 2007, to over seventy thousand (Curtis, 2008). And having set up an inquiry, the Secretary of State for Children, Schools and Families found it necessary to announce in October the immediate scrapping of KS3 NC tests, with nothing similar ready to replace them. Suddenly the external testing-burden was lifted from Secondary schools and students.

But only from them. KS1 and KS2 tests would proceed in 2009, and baseline assessment would continue. In response to the retention of these tests the NUT voted at its annual conference to ballot members for a test-boycott in 2010. Sister-union the NAHT indicated it would do likewise, thereby uniting the two unions of most weight in the primary sector against residual NC testing. NASUWT leader Chris Keates, swayed no doubt by her union's policy of 'social partnership' with the government, criticised this action. Her union's conference threatened a strike if changes to the testing-system increased members' workload, laying the union open to portrayal as a self-interested defender of the discredited testing-regime.

If successfully called, the NUT/NAHT boycott may coincide with New Labour's attempt to go to the country in search of a fourth successive term of office. DCSF officials have already threatened teachers with the law.

Wholly Unacceptable

In the report of his inquiry into the delivery of NC testing, published in December last year, Lord Sutherland depicts the chaos which engulfed QCA and its subsidiary division the National Assessment Agency (NAA) as the 2007/2008 round of testing proceeded. He anatomises the many failings and complacencies which culminated in systemic breakdown, and apports blame. The last leg on the journey towards disaster began with the selection of ETS Europe (a division of the giant US-based educational-testing corporation ETS Global) to deliver NC testing for five years starting from 2008. But the drawn-out and still unconcluded demise of NC testing, at least in the form of SATs, might be said to have originated some years earlier. In 2003, on the tenth anniversary of the huge resistance mounted by NUT and NASUWT to the imposition of testing in secondary schools, the NUT attempted to boycott the tests a second time. The vote to boycott was won, but turn-out was too low to trigger the action and tests went ahead that year as usual. The government quietly altered arrangements for the testing of seven year olds, but seemed to

have faced down widespread public opposition to NC tests. It appeared, as ministers were so fond of repeating, that testing was here to stay.

But the structures set up by government to ensure testing took place on time and to budget were increasingly over-stressed. In 2004 KS3 NC test-results in English were substantially delayed as the system established to ensure that some nine million test-scripts were marked and reported-on in less than ten weeks failed to cope. Further breakdowns were averted in each of the three succeeding years only by retarding the marking-deadline for KS3 English. In 2007 QCA signed up ETS Europe to replace the privatised exam-board Edexcel, and a period of transition began, adding an extra layer of complexity and risk to the system.

ETS did not take the opportunity to submit evidence to the Sutherland inquiry, and in return the report is scathing about the company. ETS failed adequately to assess what was involved in running the testing-system in England, to identify the risks, or to ensure adequate resources. As things went awry it did not disclose the true extent of problems to other agencies. Its treatment of markers, schools and ultimately of students was 'wholly unacceptable and lacked professionalism' (Sutherland, 2008, p. 8)

But ETS does not shoulder the blame alone. QCA and NAA are also criticised for a range of failings, as is the newly-created regulator Ofqual, albeit to a lesser extent. In particular, the scrutiny of the track-record and reputation of ETS was inadequate. Had the company been better investigated, it might never have been appointed. Shadow Education Secretary Michael Gove submitted to the Sutherland Inquiry a catalogue of examples of ETS failures. Yet the 'due diligence' element of the procurement-process undertaken by PricewaterhouseCoopers, in which DCSF observers were fully involved, did not identify these. Sutherland argues that better checks should have been carried out (Sutherland, para 2.64). Ironically, the procurement-process was recommended as 'a case-study of best practice' by the Chair of QCA Board (Sutherland, 2008, para 2.44). Furthermore, the contract drawn up for ETS was ambiguous in important ways, and interpreted differently by different agencies. Oversight of the stages of its fulfilment was lacking in many respects, as were procedures for identifying and neutralising risks. NAA staff found themselves working evenings and weekend on tasks which were properly the contractor's in an ultimately futile attempt to prevent the breakdown. Had the narrow avoidance of catastrophe in each of the previous three years engendered a sense inside NAA and QCA that all would contrive to come good in the nick of time?

Under the Radar

Establishing such a scenario would appear to suit ministers. Once his resignation was finally accepted and he felt free to speak out, Ken Boston challenged its validity in strong terms. He wrote to Barry Sheerman's committee, before whom he had endured a torrid morning's questioning in the week after the test-results had failed to be produced in July 2008, and was

invited to appear before them again. He did so on April 22 2009, accusing ministers of having given a false account to Lord Sutherland of his own meetings with them, and calling the account offered in three particular paragraphs of Lord Sutherland's report (each of which was based on ministerial evidence) 'fiction' and 'without foundation'. Contrary to the impression ministers conveyed, Boston declared that he had not been 'pressed' by DCSF emails and phone-calls about the problems overtaking NC testing, nor had he given ministers 'strong reassurances' that the test-results would be delivered on time. He resented being cast as 'complacent' and 'disengaged' (Boston, 2009).

Dr Boston was certainly right to claim that particular evidence given to Sutherland's inquiry by Jim Knight, Schools Minister, was false. In paragraphs 4.92 and 4.137 of the report Sutherland refers to a meeting Knight stated took place on 17 June between himself, Boston and David Gee. Knight claimed this meeting, held a mere three weeks before the date set for publication of the test-results, was the first time he had talked in detail to Gee about what was going wrong with the marking-processes. He states that Dr Boston referred everything to David Gee. The Sutherland report declares that DCSF notes supported the minister's recollection of the meeting. However, Knight had made a mistake. Ken Boston was not at this meeting. He had not even been invited to it. Knight has made it known that he wrote to Lord Sutherland after the publication of his report to set the record straight on this matter, and that Sutherland had responded that the corrected information made no difference to the overall thrust of his findings. (Subsequent editions of the Inquiry report have been amended in line with Knight's correction.)

Ken Boston remains unsatisfied. But his own position is undermined by his declaration to Lord Sutherland about how he performed his role overseeing the fulfilment of the testing-contract:

My own involvement was essentially through David (Gee) and getting David (Gee) to do things directly with [ETS]... I was not hands on in the sense of doing David Gee's job for him. I was wanting to know what he was doing, what the result was, what information he was getting and what response. But yes, it was clearly an issue that was of high concern, clearly an issue as National Curriculum tests every year are of high concern. But I honestly cannot say at any point up until mid-June, that I was more alarmed or more worried than I had been in the two previous years. *It did not get that high on my radar.* (Sutherland, 2008, para 3.69. My emphasis.)

The frankness of the final admission is admirable, but it is bound to prompt the riposte: why not? Especially since Dr Boston had been here before. He told Barry Sheerman's committee that he came close to resigning in 2006 because of the government's unwillingness to modernise the way NC tests were marked and reported. 'Every year this has been a high-wire act', he said, calling the task 'almost impossible' (Boston, 2009).

Outsourcing Blame

The Department for Children, Schools and Families does not escape Lord Sutherland entirely unscathed, though it is let off much more lightly than the other agencies. Ken Boston claims the inquiry was set up in ways which would ensure this (Boston 2009). It is possible, however, to garner from the report evidence of complacencies and failings on the part of DCSF observers. They were after all embedded with the agencies as the breakdown unfolded. 'Fully involved' (Sutherland 2008, para 2.44) in the procurement-process which selected ETS, they attended the major meetings as arrangements for testing in 2008 proceeded, and exercised speaking (though not voting) rights. They seem not to have challenged ETS sufficiently when problems became obvious, nor to have pressed home such challenges as were made. Sutherland writes:

[DCSF officials] did not challenge NAA on the shortcomings of its project-documentation...

DCSF officials did not use all the sources of information available to interrogate the reassurances they were receiving [from QCA/NAA]...

[F]eedback from DCSF to QCA... could have been more challenging in the critical period leading up to the test-delivery in the summer.

(Sutherland, 2008, paras 4.42, 4.143 and 4.146)

Consequently, although DCSF observer-briefings began to indicate problems as early as December 2007, ministers appear not to have been helped to grasp the full scale of the failure or its implications until June 2008, by which time it was too late for the crisis to be averted. It seems legitimate to ask how ministers themselves could have allowed this to happen. A number of facts should have forewarned them that the 2008 testing-process needed particular scrutiny. To ensure the SATs ran on time was the biggest annual undertaking QCA was charged with by DCSF. Since the mini-meltdown of 2004, in the aftermath of which a previous head of NAA had been forced to resign, there had been further difficulties year on year ensuring the punctual delivery of marks. ETS arrived in 2007 not only as a new contractor but as one untried in the English NC testing environment, and seemingly ignorant of important aspects of it. Sutherland reports, for example, that ETS staff did not know what Ofsted was nor what it did (Sutherland, 2008, para 5.206), and criticises ETS for not taking due account of 'the UK context' when designing its systems and processes (Sutherland, 2008, para 5.204). For their part, ETS tried to instigate a number of changes to the system handed over by their predecessor which further increased the risk to the entire project.

Sutherland's report exposes the complex nature of the lines of responsibility across the structures and agencies involved, and quotes David Bell, DCSF Permanent Secretary, on the rationale for such a tangle:

[W]hat we do is delegate... a large number of our delivery responsibilities. And to some extent that reflects the theology of the last twenty years or so as central government departments have increasingly passed over responsibility to other government agencies... (Sutherland, 2008, para 3.5)

Were this theology, the devil could commend it. The policy of 'passing over responsibility' (which the Permanent Secretary is unashamed to acknowledge stems from the Thatcher era) has also been called 'the outsourcing of blame' (Mansell, 2009, p. 7). Central government's observers monitor the work of 'sub-contracted' agencies, but the mechanism of 'sub-contracting' serves to minimise the blowback towards ministers when things go badly. The buck stops some distance from the Secretary's desk. When asked, Ed Balls approved separating those with what he termed policy-making expertise from those with expertise in policy-delivery (Sutherland, 2008, para 3.4). But it could be argued that it is precisely this divide which works to prevent policy-making from being properly-informed by what happens during the process of policy-delivery. The effect and consequence of what the minister does or leaves undone at policy-making level is prevented from becoming part of her or his lived experience. It remains a kind of thought-experiment, under-resourced by the empirical. The determining of policy is harmed as a result. To embrace Bell's 'theology' engenders a dereliction of ministerial duty.

Ken Boston told Lord Sutherland that the maintenance of a gap between policy-makers and policy-deliverers, the so-called 'arm's length' principle, was in fact an illusion. In relation to the failures of 2008, DCSF officials had access to the same data and information as did NAA and QCA and were 'properly part of the process' of delivery (Sutherland, 2008, para 3.12), with the implication that DCSF was as blameworthy as everyone else. He repeated this view to Barry Sheerman's committee, arguing that it had major implications for Ofqual's claim to independence (Boston, 2009).

Premium Rate

So the slow collapse of the system took its course, exacerbated by what the QCA memo of June 2007 had called 'the wider use of technology'. Sutherland reports how problems with online aspects of the test-marking procedure had been foreseen but either went unremedied or were cured in ways which made matters decidedly worse. DCSF observers had questioned (but not robustly enough to alter it) the belief ETS appeared to hold that schools would act entirely in accord with procedures ETS set up, and in due time. In the event it transpired that some ten percent of schools failed fully to complete the initial on-line registration of their NC test-taking cohort. Consequently markers could not upload completed information on marked students to the ETS system. ETS had not readied any contingency to meet this situation. As the delay in processing marks extended, ETS decided to fix the problem by defaulting their

system so that it registered every student who was eligible for the tests as having been present for them. This in turn altered the parameters of the system so that it now expected markers to upload information on every eligible student, including students who had in fact been absent from the tests. Furthermore, and fundamentally, it meant that ETS could not now tell which students had actually taken which tests, nor where their scripts might be (Sutherland, 2008, pp. 97-98).

Among other consequences, these failures led to markers receiving more or fewer scripts than anticipated, at varying times and in varying conditions of security, and sometimes from unexpected sources. KS2 scripts went to KS3 markers, and vice-versa; Science scripts went to Maths markers, English scripts to markers expecting those for Science. At least ETS had established helplines. But, in keeping with that entrepreneurial ethos so admired by New Labour, these used premium-rate numbers, so markers and schools seeking advice paid over the odds '... despite the fact that the service was poor and customers had to call multiple times in many cases' (Sutherland, 2008, para 5.217).

Requesting help by email could prove cheaper but no less frustrating. Sutherland discovered ETS failed to answer ten thousand emails. He quotes one Headteacher's verdict on the company: 'In all my thirty-four years as a schoolteacher and twenty-five years as a Deputy Head/Headteacher I have never experienced such lack of concern and such an inability to help' (Sutherland, 2008, para 5.217).

On July 4 2008, four days before the official deadline for delivery of all marks, Ed Balls was forced to announce a delay in their publication and an extension of the marking-period. Most KS2 students received results a week late, and most KS3 students by the end of term, but a 'significant proportion' had to wait longer (Sutherland, 2008, para 13, 14). On October 14 Ed Balls returned to the House of Commons to announce the scrapping of KS3 NC tests.

In his statement Balls replaced KS3 tests with increased reporting by teachers to parents, enhanced classroom assessment by teachers and a yet-to-be-devised system of national sampling to ensure 'standards' are maintained. He set up an 'Expert Group' to advise him on these matters. With neither classroom teachers nor students represented, this group is due to report in May 2009. Their brief has been so tightly-drawn that they are unable to comment on the advisability or otherwise of maintaining the current testing-system. They may only recommend to schools ways to test-ready students which are 'proportionate, educationally appropriate and do not inhibit the delivery of a broad and balanced curriculum' (Stewart, 2009). But as Sheerman's committee made plain in great and well-sifted detail, the obstacle which prevents students being prepared for the tests in these ways is the existence of the NC testing-system in its current form. Ministers can have either a broad, balanced curriculum and that which is educationally appropriate, or they can have NC testing at Key Stage 2. They cannot have both.

Assessors, Samples, Seedbeds

To rival Labour's 'Expert Group' the Conservatives set up their own inquiry into testing. It is headed by Sir Richard Sykes, an unashamed educational elitist and fan of selection by 'ability' (Jha, 2007) who has criticised A-Levels for failing adequately to distinguish the 'brightest' from the rest. His inquiry is due to report shortly, and may influence the Conservatives' manifesto-pledges on education. The government continues to pilot its Single Level Testing (SLT) project in over four hundred school, allied to its 'Making Good Progress' policy and its quantitative-data-driven Assessing Pupil Progress scheme (APP). Reassuringly, none other than PricewaterhouseCoopers have been chosen to conduct the evaluation of SLT.

Alternative perspectives have been put forward. The Welsh replaced SATs with a more diagnostic system founded in teacher-assessment. Peter Tymms of Durham University's Centre for Evaluation and Monitoring has long argued for the testing of a small random sample of students each year using 'objective measures which don't have to be statutory tests' (Garner, 2009). Tim Brighouse, who accepted membership of the 'Expert Group' and hoped to be able to breach its narrow remit, warned that an externally-marked system of Single Level Tests would prove costly and bureaucratic. Better to accredit teachers as assessors, and for Ofsted to test and compare samples of students before renewing the teacher's accreditation-license (Brighouse, 2009). The NUT and NAHT issued a joint-statement outlining their ideas for national assessment. The statement is based on one made in June 2005 by all the main teacher-unions. It calls for teacher-assessment to determine NC levels at each Key Stage, for moderation to be conducted between schools, for a national bank of assessment-tasks which teachers would help compile and from which they could draw, and for a system of national sampling. The unions also call for a mechanism to recognise the breadth of a student's achievement, rather than a concentration solely on English, Maths and Science (NUT/NAHT, 2009).

The issue of testing and assessment will be discussed in the final report from the Cambridge Primary Review team, whose most recent publication confronted the distorting effects on the curriculum of current KS2 testing-arrangements. Such testing, it made clear, compromises the entitlement of students to a broad, rich and balanced curriculum, the seedbed for enhanced attainment by every student. Only by ensuring schools can provide it will all students have the chance to prosper. Scandalously, the suggestion is put forward that primary schools reserve one-third of their teaching-time for teaching not the National Curriculum but a Local Curriculum which schools themselves may determine. Such a policy could enable schools more closely to meet the needs of their students, and would return to teachers some measure of foundational decision-making about the educational offer they think it right to present.

One more review is awaited: that of Sir Jim Rose. Straitjacketed in the familiar way by a government remit which specifically forbade consideration of how to change the testing-system, it finds itself directly challenged on the terrain of the primary curriculum by the Cambridge Primary Review team. Rose

has already come out against teaching-to-the-test. Will he be able to make recommendations about curriculum without at least gesturing towards, if not stating directly, what must follow so far as assessment and testing is concerned? His Interim Report (Rose, 2008) spoke positively about the potential of Assessing Pupil Progress. It stated that in the future APP 'should reduce teachers' reliance on testing as the main source of evidence for achieving national standards' (Rose, 2008, para 1.26). But APP requires teachers to fit their students to pre-set criteria. It rivets teachers and students to the track of 'progression' already laid down by policy, and renders any other assessment-approach illegitimate.

Washed Up

Should Single Level tests and APP come to replace the current version of NC testing, it will be no real advance. A better way to think and talk about assessment, as well as to conduct it, remains urgently needed. An assessment-system drained of its poison would understand assessing a student as also an ethical and political act, one not discharged through judgement of a particular performance. Such an assessment-system would, in my view, need to be couched in the living and unanticipated language of considered observation and evaluation, not in the inert script of APP's tick-box robotics. Instead, an alternative assessment-system could mobilise a new set of questions and answers about how best to harness the teacher's own sense of accountability to herself with the need for a just account of student-progress and development to be rendered in the public sphere.

Lord Sutherland eschews the word, but the media were quick to label the events of 2008 'a fiasco'. In this they were apt. Its etymology is obscure and contested, but it seems 'fiasco', the Italian for 'flask' or 'bottle', washed up in English a century and a half ago specifically in relation to a theatrical or operatic flop. That is, to a failure of prepared public performance. Teachers condemn NC testing for the way it shrinks what matters in education to the outcome of a student's brief display, presuming to bound within a nutshell each singular mind, imposing a regimen of test-rehearsal on schools, and re-casting teachers as trainers. There is poetic justice in the enforced abandonment of KS3 SATs because of the failure of a private company to perform publicly in the way and on the day it had contracted to. The culpability of ministers for the happy fiasco which destroyed KS3 NC testing may be obscured between the lines of Lord Sutherland's report, and the DCSF troupe may still insist that what's left of the show must go on and that tests, of seven and eleven year olds at least, are here to stay. But, little by little, the curtain is coming down. The drama of SATs, part-tragedy, part-farce, looks almost done.

References

- Boston, K. (2009) Evidence before the House Select Committee for Children, Schools and Families, 22 April.
<http://www.parliamentlive.tv/Main/VideoPlayer.aspx?meetingId=3928&rel=ok>
- Brighouse, T. (2009) All I Want for Education... , *The Guardian*, 2 January.
<http://www.guardian.co.uk/education/2009/jan/02/tim-brighouse-wishlist>
- Curtis, P. (2008) Sharp Increase in SATs Score Complaints after Marking Fiasco, *The Guardian*, 20 September.
<http://www.guardian.co.uk/education/2008/sep/20/sats.schools>
- Garner, R. (2009) Testing of Pupils Has 'Serious Negative Effect', *The Independent*, 10 April. <http://www.independent.co.uk/news/education/education-news/testing-of-pupils-at-11-has-serious-negative-effect-1666843.html>
- Jha, A. (2007) Over the Rainbow, *The Guardian*, February 27.
- Mansell, W. (2009) Tested to Destruction, *Education Journal*, 114, 7.
- NUT/NAHT (2009) Statement on Assessment.
<http://www.teachers.org.uk/resources/pdf/NUT-NAHT-Statement-6256.pdf>
- QCA (2007) Further Memorandum Submitted by QCA to Select Committee for Education and Skills, Paper 3 Testing and Assessment-the Use of Electronic Media.
http://www.qca.org.uk/qca_13936.aspx
- Rose, J. (2008) The Independent Review of the Primary Curriculum-Interim Report.
- Russell, J. (2008) Balls' Test Answer? More of the Futile, Top-down Plans that Labour Loves, *The Guardian*, 28 July.
<http://www.guardian.co.uk/commentisfree/2008/jul/28/education.labour>
- Stewart, W. (2009) Expert Group Steers Clear of SATs Reform, *Times Educational Supplement*, 17 April, p. 1. <http://www.tes.co.uk/article.aspx?storycode=6011920>
- Sutherland, S. (2008) *An Independent Inquiry into the Delivery of National Curriculum Tests in 2008: a report to Ofqual and the Secretary of State for Children, Schools and Families*. London: The Stationery Office.

PATRICK YARKER is a doctoral student at the University of East Anglia. Correspondence: Patrick Yarker, Dols Hole Cottage, Hall Road, Beetley, Dereham NR20 4DE, United Kingdom (patyarker@aol.com).

