



Pointmaker

TICKING THE RIGHT BOXES

A RELIABLE, FASTER AND CHEAPER ALTERNATIVE TO SATs

TOM BURKARD

SUMMARY

- In the last six years, failure over exam marking has led to three separate inquiries and the resignation of one Education Secretary and two chiefs of the QCA. It has, more importantly, disrupted both primary and secondary schools and undermined the validity and reputation of external examinations.
- Last July, over half of all exams were marked late; and a quarter were estimated to be marked incorrectly. Yet, despite previous ministerial promises that it “would never happen again”, all warning signals were ignored. The QCA now warns of more delays in 2009.
- The difficulties with English SATs in particular are systemic. The underlying problem is the nature of the open-ended essay questions which are intended to test pupils’ ‘critical thinking skills’.
- Yet it is simply not realistic to expect examiners to be able to accurately mark essays written by 600,000 11 year olds.
- Pressure to end external examination at 11+ must be resisted. Parents, teachers and politicians need to have rigorous evidence on pupil performance.
- This can be achieved by replacing the essay format with multiple choice tests. This would:
 - provide a far more accurate and reliable picture of how well pupils and schools are performing;
 - be far cheaper and quicker to mark;
 - be an accurate test of knowledge and ability;
 - mean that it is impossible for teachers to ‘teach the test’; and
 - enable accurate year on year comparisons on school performance.

THE FIASCO

The 2008 fiasco over the marking of SATs tests was only the most recent in a series of unfortunate events.¹ The exam agency responsible for the crisis (ETS, an American testing service) had, after all, been employed because of the previous failures of the exam boards to mark tests accurately and to deliver results on time.

As far back as January 2002 *The Guardian* reported that “Officials are desperate to avoid a repeat of this term's debacle in the main summer examinations in England.”² The problems with A-level results later that year led to the resignation of the then Education Secretary Estelle Morris and the then Chairman of the Qualifications and Curriculum Authority (QCA) Sir William Stubbs.

Matters came to a head again in 2004, when Key Stage 3 English test results were delayed. Despite the following inquiry citing “poor leadership” at the QCA, there were no resignations. Rather, the then School Standards Minister David Miliband vowed that: “it's vital for schools and parents that a delay in delivering results does not occur in the future.”³

¹ The majority of SATs results were delivered late, with over a quarter being reported to be inaccurate. For example, Ron Naylor, head teacher of Forefield junior school in Crosby, Liverpool stated: “The standard of marking on the English papers that we have now got back is absolutely appalling.” (*The Times*, 20 July 2008). As SATs results are also used to band children in secondary school, this caused much disruption to both primary and secondary schools. Lord Sutherland, the former chief inspector of schools, who led an official Inquiry, commented: “Failures occurred at almost every stage of the test delivery process.”

² www.guardian.co.uk/education/2002/jan/22/aslevels.secondaryschools1

³ *The Guardian*, 17 November 2004.

For the next three years, Edexcel (the only exam board willing to take the SATs contract) continually struggled to meet deadlines. The QCA then awarded the exams to ETS, despite the firm's failure to manage exams in other countries. The problems with ETS were apparent as early as 1 May 2008, but Ken Boston (then Chief Executive of the QCA, who has since resigned) waited until 26 June to inform the Secretary of State, Ed Balls.⁴

The ETS contract has been cancelled, and Edexcel reappointed. It has already warned that further delays with the 2009 SATS papers are likely.⁵

There is clearly a systemic problem. But does it need to be like this?

WHAT ARE SATs?

The benefits of learning to read, write and calculate are not controversial: virtually all parents and politicians agree that these are essential skills.

Yet most also agree that they have been, and remain, badly taught in many state-supported schools. Hence the need for mandatory tests of these subjects.

SATs (Standard Assessment Tests) originated with the 1988 Education Act, to meet this need and to check the growth of ‘peace studies’ and other subjects that were replacing traditional academic disciplines.⁶

⁴ *The Independent*, 24 July 2008.

⁵ *The Daily Telegraph*, 31 December 2008.

⁶ The problems that have plagued SATs have also affected other exams, and for similar reasons. However, this paper is restricted to comments on the former.

Pupils are tested in English and Maths at 7+ years old, 11+ and 14+; and in Science at 11+ and 14+. As of the current academic year, the 14+ tests are not mandatory. In the last few years, teacher assessments have replaced external tests at 7+: the 11+ test is therefore the only remaining exam which is externally graded.

The tests have always been the centre of controversy: in 1996, Melanie Phillips chronicled the 'capture' of the National Curriculum by the very same progressive ideologues that it was intended to thwart.⁷ And SATs have consistently been opposed by teaching unions, especially the NUT. More recently, there has been a groundswell of opposition from educators, who maybe sense an opportunity to get rid of external assessment altogether.

Indeed, in educational circles, it is no secret that SATs may soon be abolished.⁸ Tests at 14+ have already gone, and external assessment of 7+ tests has been abandoned. This also explains the reluctance of exam boards to bid for the all-but-unmanageable job of marking the remaining 11+ tests. However, in the absence of a suitably rigorous alternative, abandonment of externally-graded tests is, for political reasons, unlikely to happen before the next general election.

⁷ M Phillips, *All Must Have Prizes*, Little Brown, 1996.

⁸ For example, the Government is already trialling a new system of single-level testing, in which children sit less formal papers twice a year. Children are entered only when their teachers assess they are capable of passing. Ed Balls has commented: "Nothing is being ruled out or in... With these (single-level) tests, it's more in the hands of the teachers. You could combine this with teacher assessment." See *The Times*, 16 August 2008.

The need for a reliable and creditable replacement for SATs is urgent; unfortunately, there is a lot of confusion. On one hand, we are led to believe that our children are relentlessly drilled in traditional academic subjects. On the other, calculating a simple percentage is beyond the great majority of our 16-year-olds while basic reading skills are still beyond a quarter of 11 year old children.

ROTE LEARNING OR RELEVANCE?

SATs are intended to test subject knowledge as well as 'critical thinking skills'. However, despite the denunciation of SATs as encouraging rote-learning, the emphasis is now on the latter: hence the need for open-ended essay questions. For example, consider this question from the 2007 English test for 11-year-olds:⁹

Each of the texts in this booklet looks at the subject of drumming but in different ways. Which text might inspire someone to take up drumming?

One thing stands out: the question has nothing to do with any academic subject, not even music. It has been selected for its presumed 'relevance' to pupils' interests.

Another example from the 2006 paper illustrates how it is empathy, not knowledge, that is being tested in pupils:¹⁰

Can I Stay Up?

In this scene, Joe is desperately trying to persuade his parents that he should be allowed to stay up late to watch TV.

⁹ *The Daily Telegraph*, 26 May 2007.

¹⁰ See <http://orderline.qca.org.uk/gempdf/1847213073/9999067999.pdf>

Your task is to continue this scene until a decision is reached.

Remember that Joe is trying to persuade his parents.

Your task is to continue the play script set out below.

Scene 1

Joe: (pleading) Dad, can I stay up to watch something special on the TV tonight?

Dad: I don't know, it depends on what it is ...

Mum: (coming into the room) ... And what time it finishes.

The following question from the 2007 SATs is preceded by the statement that "Your teacher will read through this section with you":¹¹

A mystery story starts with these words:

Ali stood silently, looking at the door. With a slow creaking sound, it opened. Taking a deep breath, Ali walked inside ...

Your task is to continue the beginning of the mystery story by describing what it was like through the door.

There is of course an intractable problem with these types of question: there are no right or wrong answers. Is it not unrealistic to expect an army of hastily-trained part-time exam markers to judge the relative worth of papers written by 600,000 pupils?

¹¹ <http://orderline.qca.org.uk/gempdf/1847214444/1847213715.pdf>

The attempted solution to this problem is to set down rigid criteria for grading essays, specifying 'appropriate' responses. Form counts above content. It is easier to judge whether the answer fits the format than whether it makes sense. Of course, this defeats the whole purpose of the exercise; every trace of originality is extirpated. Pupils therefore spend a large part of their last year in primary school writing practice exams, learning to use 'writing frames' that shoehorn their words into standard formats. The result is the worst of both worlds: rote learning of material with no discernible academic value.¹²

TEACHER ASSESSMENT

There has in recent years been increasing emphasis on teacher assessment (where the class teacher makes an assessment of the level attained by the pupil). These have now replaced the SATs test for 7+ year olds.

There is increasing pressure from the education establishment to end external assessment. The Gilbert Review in particular has called for a much greater reliance on teacher assessment, as well as self-assessment and peer-assessment.¹³ The latter two are defended on the grounds that they help children take control of their own learning, which is assumed to be a motivating factor. The hard evidence is lacking; without a strong teacher to guide them (as was the case in a seminal study¹⁴), the pupils may also take control of the

¹² In contrast, the results of Maths and Science SATs have answers that are clearly either right or wrong. While they have their own problems (particularly with dumbing down), Maths and Science SATs have not suffered the delays experienced by English SATs.

¹³ DFES, *2020 Vision, The Gilbert Review*, 2006.

¹⁴ P Black and D Wiliam, *Inside the black box: raising standards through classroom assessment*, NFER-Nelson, 1998.

classroom, the teacher and the school. The children's rights agenda, unsurprisingly, is not as popular with teachers as it is with romantic theorists in education colleges.

Of course, not all of the pressure for teacher assessment is ideological, nor does it all stem from ministers' desire to avoid another SATs fiasco. It is cheap. £700 million was spent on all exams in 2005, enough to fund 1,400 primary schools.¹⁵ By transferring the burden to teachers, much of this expenditure can be saved.

Unfortunately, in doing this, the Government is merely passing the bill to others. Firstly, teachers suffer. Formal assessment of any variety is time-consuming. Since these assessments are really tests of teacher performance, teachers are put in an invidious position: do they exaggerate a bit, knowing that it is unlikely that they will be caught out, or do they jeopardise their own careers with an honest appraisal of their pupils' progress?

Second, pupils suffer. Anything that distracts teachers and absorbs their mental energies detracts from the time and effort they can devote to teaching.

Third, there is evidence to show that pupils from the least advantaged homes suffer the most. One of the Government's main objectives is to reduce the gap between the achievement of the most and least able pupils (this is explicit in the Gilbert Review). Unfortunately, evidence from other countries suggests that teacher assessment can have the opposite effect:¹⁶

Most importantly, there is recent evidence from Sweden ... of substantial "grade inflation afterwards"... Over five years, performance, as judged by teachers, rose by 13%, as a result of the move from external examinations to teacher assessment, using a comparison with externally set tests. Furthermore, it appears that children in private schools, where the pressures for high grades are greatest, have seen the largest grade inflation, and there is also higher grade inflation for the higher achievers. In other words the relatively low achievers in state schools suffered most from the change.

THE HIDDEN VIRTUES OF MULTIPLE-CHOICE EXAMS

Replacing the essay format for the English SATs taken by 11 year olds with multiple-choice examinations would have many benefits. It would, for a start, eliminate the recurring crises that have plagued SATs. The technology for delivering them is reliable and cheap. And they would provide a far more accurate picture of how well pupils are performing.¹⁷

The essay exam is a traditional feature of British pedagogy.¹⁸ It does of course have its advantages. In particular, it provides a much more intimate picture of the student's mind. This may well be important for *formative assessment* – when teachers have to decide where their pupils are going astray. But this is not the purpose of SATs. They exist mostly

¹⁵ *The Guardian*, 11 January 2005.

¹⁶ *The Guardian*, 23 February 2005.

¹⁷ SATs do currently contain some multiple-choice items, simply because it is the only way they can test a broad range of objectives.

¹⁸ Essay exams would of course continue at other exam levels such as A level and GCSEs.

to tell parents and policy-makers how well schools are doing. They are, in other words, high-stakes tests of teachers and ministers.

And for this purpose, multiple-choice tests have huge advantages. The psychologist Jeffrey Jones explains that:¹⁹

... in the 1960s, Godshalk and others decided to...determine if objective, standardized questions could be prepared that would predict how well a student would score on essay tests... They discovered multiple-choice test items which yielded test scores that were quite comparable to scores obtained by having students complete an essay examination requiring several hours. In fact, they ended up concluding that a multiple-choice test that took 20 minutes to complete gave the same information as an essay examination which required two full and three half classroom periods (240 minutes of testing time). The multiple-choice test could be machine scored in less than a second whereas the collection of essays written by each student took over two hours for the raters to score.

When compared to essays tests that are adequately marked, multiple choice tests are 12 times more efficient in terms of the time taken to sit a test, and over 7,000 times more efficient in terms of marking time. And even more to the point, they are completely fair.

Using multiple-choice exams will also end the annual debate about rising standards and dumbing down. Individual tests of

equivalent weight can be created with tests created by random selection from a large bank of graded questions. Providing that the bank is big enough, it becomes impossible to 'teach the test' without teaching the relevant skill, knowledge and concepts. This also eliminates the possibility of cheating, as no two tests are identical.

Modern tests are far more sophisticated than most people imagine. In the US, a recent study examined:²⁰

...the relationship of Graduate Record Examinations (GRE) General Test scores to selected personality traits, including conscientiousness, rationality, ingenuity, quickness, creativity, and depth. Analyses revealed statistically significant, *positive* correlations of GRE verbal, quantitative, and analytical scores with both creativity and "quickness." (Quickness was defined here by, for instance, the ability to handle a lot of information and the ability to understand things.) In other words, the "deep-thinkers" did better on multiple choice questions, just what we would hope for.

It is worth noting that in commerce and industry, the use of multiple-choice tests to assess applicants and to evaluate performance on training courses is all but universal. In the real world, where a lack of knowledge can have disastrous and costly consequences, the essay exam is increasingly irrelevant.

This is not to say that the ability to write well is not useful – indeed, it has become so rare that it is now a highly-valued skill. The point

¹⁹ <http://my.execpc.com/~presswis/assdbt.html>

²⁰ www.illinoisloop.org/test.html

is different: if we want to measure how much 600,000 11 year old pupils have learnt, essays are grossly inefficient. Also, they unfairly penalise bright pupils who, for whatever reason, find writing difficult.

One might argue that essay exams provide an incentive to teach children to write well. However, it clearly has not worked this way. Children are now being forced to write before they have learnt to spell, and before they have learnt the basic rules of grammar and punctuation. Although our pupils are writing more than ever before, all too often they are just re-enforcing bad habits. The widespread use of writing frames has done nothing to correct these problems. A recent survey by the Institute of Directors disclosed that the 71% of employers believe that the writing ability of young recruits has declined over the last ten years. Only 5% think they have improved.²¹

Even more damning were the following comments made in *The Guardian* by an exam marker:²²

...after marking GCSE exam scripts for a major UK examining board for the past two weeks, I can honestly say that not only are standards dropping, but also they are unbelievably low... In relation to the GCSE candidates' general standard of writing, as a part-time lecturer at a university, I had already become aware that many undergraduate students had abysmal reading and writing skills. However, even that did not prepare me for the written skills of your average GCSE candidate. The handwriting, most of the time,

resembled that of a five-year-old toddler or a drunk (grotesquely simple or an illegible scrawl). A lack of basic punctuation, such as full stops, commas, capital letters etc, was commonplace. There were countless inarticulate, immature sentences, which did not make any sense to the reader.

If this is true of GCSEs (usually taken by pupils aged between 14 and 16), how much worse must the situation be for 11 year old pupils taking SAT tests?

Replacing essay exams with multiple choice tests will not immediately help our children become better writers. But it will at least eliminate a lot of counter-productive activity in and out of the classroom, and make it possible to introduce teaching practices that will improve children's basic literacy skills.

That would of course be the greatest benefit of all.

²¹ M Harris, *Education Briefing Book*, IoD , 2008.

²² *The Guardian*, 25 August 2005.



THE AUTHOR

Tom Burkard is the Director of The Promethean Trust, a Norwich-based charity for dyslexic children. He is the author of *Inside the Secret Garden: the progressive decay of liberal education* (University of Buckingham Press, 2007). His main academic interest is the interface between reading theory and classroom practice, and he has written numerous articles for academic journals and the press. He is the co-author of the Sound Foundations reading and spelling programmes, which are rapidly gaining recognition as the most cost-effective means of preventing reading failure. He is the author of (with Martin Turner) *Reading Fever: Why phonics must come first* (Centre for Policy Studies, 1996), *The End of Illiteracy? The Holy Grail of Clackmannanshire* (CPS, 1999), *After the Literacy Hour: may be the best plan win* (CPS, 2004), *A World First for West Dunbartonshire? The elimination of reading failure* (CPS, 2006) and *Troops to Teachers: a successful programme from America for our inner city schools* (CPS, 2008). He is a member of the NAS/UWT.

ISBN 978-1-905389-92-6

The aim of the Centre for Policy Studies is to develop and promote policies that provide freedom and encouragement for individuals to pursue the aspirations they have for themselves and their families, within the security and obligations of a stable and law-abiding nation. The views expressed in our publications are, however, the sole responsibility of the authors. Contributions are chosen for their value in informing public debate and should not be taken as representing a corporate view of the CPS or of its Directors. The CPS values its independence and does not carry on activities with the intention of affecting public support for any registered political party or for candidates at election, or to influence voters in a referendum.

© Centre for Policy Studies, January 2009