

VIEWPOINT

Trust in testers

John White*

Institute of Education, London, UK

It is time to replace the examination régime at 16 and 18 by something more appropriate. The coalition government has been solidifying its place by its Bacca-laureate reforms at both ages, but this is a move in quite the wrong direction.

Whatever the wider purposes that the examination system may serve – competition for higher studies, comparisons between schools, or whatever – its core aim is to find out how well students are faring in their learning. Most people take it for granted that public examinations are the most appropriate vehicle for this. They do not always ask whether there are better alternatives.

The examination régime has many faults, among them financial, psychological, sociological and ethical. It is a multi-million pounds drain on school budgets; it causes anxiety in many examinees; it disproportionately benefits students from families and schools who know how important examination success is in getting ahead in life and do what they can to guarantee it; it encourages undesirable attitudes like teaching to the test, an instrumental attitude to learning, and even, sometimes, cheating. In addition, its inclusion of norm-referencing as well as criterion-referencing features can obstruct the core aim just mentioned. We saw an example of that in 2012 in the use of five-year-old data from SATs results at 11 to predict GCSE performance at 16.

The examination regime also has epistemological deficiencies. These are more serious in that they strike at the heart of what examining is supposed to be. The more public examinations diverge from testing facts or routines, the more problematic their results can be. Following a philosophical line of argument that Davis (1995, 2008, chap. 1) has opened up, if we are looking for evidence of how far students have a deeper grasp of a subject – the English Civil War, say – this means taking into account the way in which a wider background of ideas bears on the connections they make between one consideration and another. What links do they make between notions – both past and present – like economic development, social class, religion, kingship, democracy, power? These connections, and this background, differ from individual to individual. Unlike the recording of an isolated fact, they are not present on an examination script. They belong to its hinterland. This raises a conceptual difficulty for examiners looking at the work of multiple candidates all of whom are strangers to them. How well can they know what lies off-the-page? If they make a judgment about what this is, how far will this judgment depend on their *own* background of ideas, including their value judgments, and so is likely to differ from that of other examiners marking the same question?

*Email: John.White@ioe.ac.uk

To what extent all this is a problem for examiners in modern languages, mathematics and the exact sciences I do not know. But it certainly applies to many humanities subjects. In the case of the literature most notably, these also often involve moral and aesthetic judgments on the part of the examinee. These will again differ from person to person. Added to the obstacle discussed in the last paragraph, they must make the job of a public examiner all but impossible – not least if one also throws into the pot his or her own moral or aesthetic hinterland, which may colour their judgments and, again, vary from one examiner to another.

The difficulty that Andrew Davis has pinpointed casts doubt on the validity of public examinations – on how far they can test what they purport to test – where it is a matter of deeper levels of understanding. It is also a challenge to their reliability insofar as it points to likely divergence between examiner judgments. If you try to improve reliability by setting questions less dependent on the ‘off-the-page’ phenomena in question, you threaten an examination’s validity (Davis 1995, 13).

Are there better alternatives to public examinations? As we have seen, a fundamental problem with them is that candidates are only numbers. If the basic reason for assessing students is to find out how well they are doing, for all but more rule-bound achievements assessors have to know something about examinees’ wider intellectual and perhaps ethical or aesthetic horizons. And for that, assuming, rightly, that they lack extensive written evidence of this, they need personal contact with those being tested.

This all points to school-based, rather than nation-wide, assessments. If our focus is attainment in school subjects with a view to qualifications for more advanced education or for employment, we could explore Tim Brighouse’s suggestion of ‘an examination system based on university methods – internally marked and externally validated and moderated on criterion referencing by chartered teacher examiners’. <http://www.guardian.co.uk/teacher-network/teacher-blog/2012/oct/22/2015-school-reform-predictions>

But there is a case for going further than this. Records of student progress or attainment can supplement or replace internal examinations. The present examination régime fits hand and glove with a secondary curriculum built around familiar school subjects: curricular and assessment arrangements are mutually reinforcing. There are good grounds for detaching the curriculum from examination constraints and rethinking it as an aims-based enterprise (Reiss and White 2013). This would encourage us to see subjects as only one kind of vehicle for pursuing aims: others, like projects, out of class activities and whole school processes would become more salient. As it is a record of achievement system enables schools to celebrate progress on a far wider front than subject-based examinations, and the more aims-based the curriculum, the more this system can come into its own.

Until now, one of the main blockages to the reform of secondary education has been university entrance requirements. It is these that have fuelled the competitive scramble for good grades at A level that has done so much to narrow young people’s educational horizons. But perhaps it is from the universities that schools will now be looking for inspiration. *The Guardian* (3 October 2012) reported a nation-wide piloting across 90 universities and colleges of higher education achievement reports (Hears). ‘Students and former students, who will govern online access to their documents – up to six pages in hard copy – will be able to provide a more rounded picture of their college life, including sporting, volunteering, employment and student union successes. It will be easier for employers to verify jobseekers’ credentials.’

This initiative has been widely applauded and is likely to lead to a more extensive scheme. Sir Robert Burgess, vice-chancellor of the University of Leicester, who led the pilot study, has said ‘In time, the steering group hopes that the wider information contained in

the Hears will eclipse the single degree classification and, where appropriate, serve as a replacement for it.’

Just as secondary schools have been plagued by pressure to get good GCSE and A levels, universities have suffered from what the *Guardian* piece quotes as the ‘damaging obsession’ with first and upper second degrees. It is becoming clearer that both institutions have a common interest in replacing traditional assessment patterns by ones that give a fuller picture of student achievements. In time, perhaps this will lead both of them to jettison traditional examinations altogether.

Perhaps, then, as far as secondary schools go, in the early twenty-first century we are standing on the watershed between two systems – an old one, in which the necessity of public examinations has been taken as read, and a new one, based largely on records of achievement. Whether or not this proves to be true, we can at least look back from the vantage point of the present and ask ourselves how it is that the old system has become as powerful as it is despite the serious flaws that more and more people are coming to see in it.

We have to see public examinations in the wider context of attitudes to authority. As is well known, a century and more ago, Britain was a far more deferential country than it is now. Master–servant relationships were much more common; clergymen, teachers, bankers, politicians, lawyers, policemen, army and navy officers were all looked up to as reliable sources of authority in their various fields; husbands ruled the household. Today things are very different. The servant class has all but vanished. Relationships within the home are more on an equal footing. Other ‘pillars of the community’ just mentioned have lost, or are losing, their old unquestioned position; some, indeed, sometimes suffer public antagonism.

Matters were different in the more religious world Britain inhabited until the early twentieth century. Monarchs were no longer seen as God’s appointed rulers, but clergymen were still held to derive their authority from God; and the same was true, in many cases, of teachers and of heads of households. In a hierarchical society conceived as part of God’s creation, these and other social leaders had a religious backing for their superiority. This helps to explain how they came to have the taken-for-granted trust which many placed in them.

Until the 1970s, those working in the field of public tests and examinations also enjoyed such a taken-for-granted trust. The first major sign of a change came with the claims that Cyril Burt, the panjandrum of intelligence testing, had invented crucial data to support his hereditarian beliefs (Kamin 1977, 55–71). In the last few years, trust has begun to erode further, first through general accusations from the political right of dumbing-down in secondary examinations, and in 2012, more specifically through charges of changing GCSE goalposts within the academic year.

The profession of examiner has swollen massively since Victorian times with the rise of intelligence testing and secondary school examinations. Examiners were far thinner on the ground before that but they did exist. In the form familiar to us today, of people from outside the examinee’s institution checking on their progress in assimilating an academic subject, they seem to have originated in the academies set up for the sons of dissenters from 1660 onwards, after dissenters were barred from Oxford and Cambridge in the wake of the Restoration. A leading historian of the Dissenting Academies, Herbert McLachlan, has favourably compared their examining arrangements in the early eighteenth century with those of Oxbridge. He quotes an account of how Cambridge undergraduates performed at their BA examination:

... One of them was asked what was the English of Anno Domini, but the blockhead was unable to tell. Another was asked how long it was since our Saviour’s birth: he said, about a hundred years ... (McLachlan 1931, 41)

Meanwhile, Academy tutors regularly set internal tests within their subject-based curricula – the remote ancestors of our present National Curriculum (White 2011). In many Academies, external examiners from Presbyterian or Independent headquarters conducted annual, on-the-spot examinations, especially of scholars funded by their community. A two-day examination of this sort in 1817, for instance, covered the following subjects: Hebrew, classics, Syriac, French, divinity, logic, ecclesiastical history, biblical criticism, mathematics, chemistry and natural philosophy (op.cit. 171–2, note 11).

In the radical protestant tradition, examination had a religious significance. Comenius (1638/1907: XXIV, 6, 9), the seventeenth century educationalist closely allied with the puritan precursors of the dissenters, drew on Luther's belief that meditation, prayer and examination 'are essential to make a true Christian' in writing that

Examination is the continual testing of our progress in piety and may come from ourselves or from others. Under this head come human, devilish and divine temptations. For men should examine themselves to see if they are faithful, and do the will of God; and it is necessary that we should be tested by other men, by our friends and by our enemies.

The examination arrangements we find in the Dissenting Academies are to be seen against this background. They were partly tests in piety, increased knowledge about divine creation being seen as a virtue, and ignorance a sin.

In the nineteenth century, the dissenters came back into public life and indeed achieved political power through the rise of the Liberal Party. Their traditional way of conducting examinations influenced the educational reforms that marked that age, not least professional qualifications for the higher civil service, engineering, accountancy and other areas, as well as school matriculation examinations for entry into London University and affiliated colleges (White 2011, 68–9, 74).

This basic structure is still with us. Indeed, the new EBacc requirements – laid down by (the ex-Presbyterian) Michael Gove – are a whisker away from those for the London University matriculation examination of 1858. Both demand high achievement in English, maths, science, a foreign language, history and geography – except that the earlier examination made the last two compulsory, not alternative.

The religious origins of our examination system have been lost to view in the more secular society of the past century or two. We no longer think of examinations as testing deliverance from the sin of ignorance. Yet something of their old aura remains in the authoritativeness we still attribute to examiners. They have long since ceased, along with other social pillars, to derive this from God, but the soundness of their professional judgment is still, recent glitches notwithstanding, by and large taken as read.

It is time, as Andrew Davis and others are now doing, to put this to the test. Those involved in public examinations are not well placed to assess students' achievements in areas beyond rule-bound knowledge, where wider background understanding and ethical or aesthetic attitudes come into the picture. We trust our examiners because we have grown used to trusting them. If we subject their role to conceptual interrogation like Davis's, as well as empirical enquiry such as Kamin's (1977) critique of the IQ, its authoritativeness crumbles.

Perhaps the next few years will see such doubts spreading from academic circles into the public more widely. 2012's GCSE shenanigans may well feed into this. It will be interesting to see how far sceptics among examiners themselves add more fuel.

Certainly, there will be many of these among university teachers. So far in this piece, as in Tim Brighouse's analysis quoted above, the target has been public rather than institution-based

examinations, like those in higher education. But any internal university examiner will testify to the difficulties – sometimes agonising – of assessing masters’ scripts or – not least – doctoral theses. This is true even where he or she has good, first-hand acquaintance with the student. Often the problem arises because of the multiple criteria brought into play and the impossibility of cut-and-dried guidance about how to weigh one against another. This is most evident in assessing doctoral theses, perhaps because they are much longer than one-hour examination answers and the examiner has so much more time to mull over them. Is this dissertation just up to scratch because a lot of work has gone into it, is well organised and presented, and breaks new ground here and there? Or is it in the danger zone because it is pretty pedestrian on the whole and has one very weak chapter? Other, more ideological, value judgements also sometimes add to the problem. Should this thesis on homosexuality and education be bounced back for not having more on traditional Christian perspectives? Is that one imperilled by its overt attachment to Marxism?

Performance at the viva adds a further dimension of uncertainty, as Louise Morley and her colleagues have demonstrated (Morley, Leonard, and David, 2003). Overall, their verdict is that ‘there still seems to be considerable variation, and some mystification, in how doctoral assessment is conducted and experienced.’ This is no comfort to the PhD candidate and their equally quaking supervisor, for both of whom several years of solid work may be at risk of rejection by an external examiner who for no good reason takes against it.

Interestingly, the traumas of institution-based, e.g. doctoral, examinations reveal the even greater shakiness of public, e.g. secondary school, ones. Blemishes invisible through a telescope can show up in a microscope. The significance of incommensurable criteria, idiosyncratic value judgements and other sources of examiner variability is all too evident in PhD assessments; but it can vanish from sight in a system where multiple examiners of the same topic each spend only a few minutes rating a complete stranger’s answer.

But the invisibility is not complete. There must be many working in the school examination industry who have doubts about these and other challenges to the possibility of objective judgment. Perhaps more of them will now reveal their misgivings publicly. They should find an attentive audience. We live in a culture where, rightly, misplaced trust in people in powerful positions – media celebrities have now been added to the list – is increasingly seen for what it is. True, scepticism about whether school examiners should retain their long, taken-for-granted reputation for solidity of verdict may not generate headlines as big as those about Jimmy Savile or Lance Armstrong. But it is likely to have impact nonetheless. For whether traditional school examinations deserve a future is an issue that, unlike the Savile case, touches the fortunes of every family.

For some people, radically to question the present examination regime is to threaten a central feature of our British social structure. Agreed. It is. But that does not make it not worth doing. Preserving it suits those with a vested interest in maintaining that structure – that is, mainly an existing elite prepared to let down a ropey ladder of opportunity to those below whom they want to join them. As with other, more prominent, pillars of an old order which sceptics have given a good shaking over the last century and a half – non-state-intervention, the dominance of the church in public life, a limited franchise, the myth of British superiority, male domination ... – thoughts once beyond the pale can come to be mainstream. How long has the examination regime still got?

References

Comenius, J.A. 1638/1907. *The great didactic*. Trans. M.W. Keatinge. London: Adam and Charles Black.

- Davis, A. 1995. Criterion-referenced assessment and the development of knowledge and understanding. *Journal of Philosophy of Education* 29, no. 1: 3–21.
- Davis, A. 2008. *Educational assessment and accountability: A critique of current policy*. Impact Paper No. 16, Salisbury: Philosophy of Education Society of Great Britain.
- Kamin, L.J. 1977. *The science and politics of the IQ*. Harmondsworth: Penguin.
- McLachlan, H. 1931. *English education under the test acts: Being the history of nonconformist academies 1660–1820*. Manchester: Manchester University Press.
- Morley, L., D. Leonard, and Miriam David. 2003. Quality and equality in British PhD assessment. *Quality Assurance in Education* 11, no. 2: 64–72.
- Reiss, M.J. and White, J. 2013. *An aims-based curriculum*. London: Institute of Education Press.
- White, J. 2011. *The invention of the secondary curriculum*. New York: Palgrave Macmillan.