

# Can better be worse?

## The conundrum of examination results

The examining season is behind us. Many thousands of pupils were processed by schools and examinations bodies (in May, June, July and August) through many tens of thousands of examinations at GCSE and Advanced levels. The data will all now be a matter of record – to resurface at unspecified points in the future as performance measures of departments, schools, LEAs and ultimately of the nation.

And there is a depressing but inevitable tendency amongst commentators to make capital out of the data in any direction that suits them. If GCSE A-C grade percentages are an improvement on last year, this is a good thing – or a bad thing, depending on who you talk to. The pupils and teachers will see it as a just reward for endless hours of hard graft, and most parents will be rightly proud of their children's achievements. Some, however, will see it as yet further evidence of the relentless decline in national "standards". If 40% of pupils in my school got A-C last year and 50% got A-C this year, then either teachers are doing a better job or standards are slipping.

Even before the exam results were officially released, this problem of interpretation was made quite explicit in the headlines of even the 'serious' press. **"Exam board denies easy grades at GCSE: ... The board says that the mark for grade C in the most demanding maths exam was lowered this year because candidates had found the exam so difficult... a spokesman for the SCAA said 'we shall be scrutinising this year's results...'"**. And then the following day **"A level pass rate up by 2 percent: ...exam officials believe that the improvement (believed to be 1.8%) is low enough to enable them to ward off criticism that A level standards are falling."**

What are we to make of this conundrum? Is it better to be getting better results – or should we be getting worse results to be sure that we are getting better ?

There are at least two questions wrapped up in this problem. **First**, is the examination system seeking to register the achievement (or not) of particular competences – or is it seeking to identify the best (and second

best; and third best) performers in a particular field. In short is the system criterion referenced ('she can do X') – or is it norm referenced ('she is better than him'). **Second**, does the examination system seek merely to judge – year by year – the excellence of the nation's pupils, or does it also seek to measure and *maintain standards over time* – so that this year's results equate to last year's and those of the year before?

### Norms and criteria

In the last 15 years we have supposedly moved progressively from norm referenced assessment to criterion referenced assessment. From the norm referencing of GCE and CSE days we gradually evolved 'grade related criteria' in the mid 1980s and then moved to 'grade descriptions' in the early days of GCSE. We then went the whole hog with the National Curriculum – with its innumerable and detailed criteria in the 'statements of attainment'. But these descriptors (criteria) could not provide absolute measures of excellence as a thermometer measures temperature, because they seek to describe something that only exists in the judgement and experience of the teacher. One of the original (1990) National Curriculum statements of attainment was *"does the pupil use specialist modelling techniques to develop a design proposal."* Teachers had great difficulty with this and most other SoA criteria because their level of difficulty was impossible to fix. It is completely adjustable according to how one chose to interpret 'specialist modelling techniques' and 'developing a design proposal'. It could be interpreted as level 2 or level 8 or as postgraduate design student.

Educational assessment is about **judgement**. The judgements that teachers are required to make are aided by *criteria* that identify the competence that is being targeted; and these judgements are also aided by teachers' awareness of their pupils relative strengths and weaknesses (i.e. *norms*). In a theoretical assessment debate it might be possible to separate 'criteria' and 'norms', but in practice teachers are constantly cross referencing between the two – using one as a benchmark to help clarify and exemplify the other. Long may it

**Prof Richard Kimbell**

*Technology Education Research Unit, Goldsmiths University of London*

last; and as Eisner perceptively observed "...when you scratch a criterion – underneath it you will find a norm."

**Standards over time**

What then of the other conundrum? Is GCSE as hard now as it was in 1987? Are A level standards going to the dogs? Two facts need to be remembered in this debate. (i) The school examination system was not set up to monitor standards over time but to provide measures year by year that enabled selection for employment and for higher forms of education. (ii) Whilst a purely norm referenced system might be able to maintain absolute consistency year by year simply by calling the top 10% A and so on down the scale to the bottom 10% being J (or whatever increments one wants to use), the use of criteria for assessment muddies the water somewhat. If more than 10% are judged to have achieved the criteria of A-ness then presumably they ought to be given As. We could of course achieve the illusion of consistency by arbitrarily drawing a line at 10% each year, but this would not be maintaining *standards*; it would be maintaining *levels* regardless of standards.

Our present examination system is quite unable to measure standards over time – except through the expertise of our examiners. Grades are awarded through the judgement of examiners, and only they can say whether this years As are exactly the same as last year's. It is a matter of judgement – not of statistics.

Interestingly, in the USA, politicians are absolutely confident that their relentless (and often mindless) testing of pupils provides absolute measures of standards over time. The tests are multiple-choice and machine-scored and hence provide very 'reliable' data. But the assessment experts in the USA know very well that the stability of their standards over time is an illusion that arises because of their test development process. Test writers produce 'test items' (i.e. questions) and each of these is pre-tested on a sample of pupils. From this pre-test emerges the 'item facility index' i.e. the % of pupils able to answer the question. A test item will not be put into the final test unless its facility index falls within acceptable tolerances. If a question is too

hard (a low facility index) or too easy (a high facility index) it is not useful in discriminating between pupils. So the whole test paper is 'normed' to the target population from the outset and there is nothing to ensure that standards stay the same over time.

But then we have always known that there are lies, damn lies and ...

- a) right answers
- b) SCAA statistics
- c) acceptable compromises
- d) regrettable errors

(candidates should tick the most appropriate response).