

Grades of Uncertainty? Why examination grades will never give us anything more than very approximate answer to questions about changes in standards over time.

In the UK, as in so many other developed countries, the interest in improving the quality of both educational provision and educational achievement is very high. In such a context examination results, because they appear to offer an enticingly simple view of how the education system is performing, can easily provide a highly tempting focus for public and media attention. Unfortunately examination results rarely provide straightforward answers to the questions to which people want to ask. As Tim Oates has outlined in his paper, the media frenzy which often accompanies the annual publication of GCSE and A level examination results, rarely contributes anything worthwhile to answering important questions such as:

‘Are educational standards in the UK improving?’

This is an issue which has interested me for a number of years, and I want to respond to Tim Oates’s provocative paper by agreeing with parts of it, and disagreeing quite profoundly with some other parts. In doing this I want to draw on some of my own research, and from that try to set out a clear position on the issue of the contribution which examination grades can make to debates about changes in educational standards over time.

The first plank of my argument is to suggest that we need to acknowledge and understand very clearly both the strengths and the limitations of UK public examinations. Among their strengths are the fact that both GCSEs and A levels perform a very useful role as a focus for much of what goes on in secondary education, and as a way of producing graded results, which give some indication of the attainment of individual students. The particular strength of these examinations is therefore that they relate directly to the curriculum, which is being taught in most UK schools, and they can be adapted as that curriculum is updated, improved and change over time. Thus as a focus for student learning they perform a successful role, and because they are set, marked and graded by well resourced and very professional assessment organisations, their results have a credibility and a currency both within the UK and in many other parts of the world.

But then alongside these strengths we must consider the limitations of GCSE and A level examinations and the grades which they produce. As I have argued extensively elsewhere (Murphy, 2004) the grades produced by these examinations are not highly sophisticated measures, which can be used to provide a fair level analysis of individual students, teachers, schools and government reforms. They are a general indication of the achievements of individual students studying particular subjects, through particular syllabuses, given the context of their school experience, the teaching they received, and the particular examinations which they sat. Thus in measurement terms they are 'approximate estimates of educational achievement', which need a good deal of interpretation, rather than precise measurements on a highly sophisticated scale.

Here then is my major point of departure from Tim Oates. I do not agree that administering and evaluating public examinations in the UK is a 'highly technical area' (p1), or that such assessment should be seen 'as an exact science' (p6). This I believe is exactly the kind of stance that gets us into such difficulty year after year. Excellent though our public examination system is, it will never allow us to directly answer questions such as 'Does a Grade A in Chemistry equate exactly to a Grade A in French?', or 'does an increase the number of students getting Grades A-C in English over a ten year period from 2000 mean that the standard of teaching in English has got better?' Educational assessment in my view isn't really a massively technical area. Anyone who has tried assessing anything from ice skating to vegetables to boxing fights will know that creating good and fair grading systems isn't easy and in any situation depending on the judgement of human assessors there needs to be an acceptance of the approximate nature of grades which are awarded. Much as Awarding Bodies would like to, they have not managed to come up with failsafe procedures to issue grades to candidates which avoid all the pitfalls of other assessment systems. Educational assessment is not an exact science, and those who try to turn it into such a thing generally start to distort the systems of assessment away from the complex curriculum areas which GCSE and A-level examinations seek to assess.

So my argument is that UK public examinations have a purpose and that is to give feedback, in broad terms, about the achievements of individual students who have followed specific examination syllabuses, relating to different curriculum areas (as defined by published examination syllabuses). Some sixty years after the introduction of our current examination system (GCE examinations were first introduced in 1950), it has clearly met with a certain amount of public acceptance. However among the things that it is not good at is making comparisons between the standards achieved by students in different areas of the curriculum, or making comparisons across time even in the same subject area.

There is much in life that we might like to compare 'scientifically' and can't. It is in my view generally not possible to come up with precise scientific answers to questions such as:

Would that person have been a better headteacher of school X than the person who was appointed?

Was Fred Perry a better tennis player relatively speaking in the 1930s than Andy Murray is now?

Is an Upper Second Class degree from University X exactly the same standard as an Upper Second Class degree from University Y?

And in exactly the same way it doesn't make a lot of sense to look at trends in national public examination statistics over periods of time to try to find an easy answer to questions about whether educational standards are going up or going down. Different students in different years are different people with different backgrounds. They almost certainly will follow a curriculum which is changing year on year, and they will certainly be examined through the use of different examination papers, which will be marked and graded by different examiners. After all of that and the fact that the examinations which they sit only produce approximate outcomes, there is little chance that minor fluctuations in national statistics will reveal great insights for those avidly wanting to know if standards are going up or down.

There is I think a general reluctance to go public with the news that examination grades are approximations and should only be treated as such. Life would be much simpler perhaps if we could rely on them as exact unquestionable statements of fact. This is I think in some ways, surprising. As long ago as 1980 the Schools Council offered the following advice:

'research has suggested results on a six or seven point grading scale are accurate to about one grade either side of that awarded'.

Very early in its life Ofqual, the new qualification regulator has once again drawn attention to the approximate nature of examination results, but such a message

hasn't been received all that readily by those who like to inhabit a more straightforward world.

So I would like to conclude my initial contribution to this debate by asserting that:

- 1 We have an excellent system of public examinations in the UK.
- 2 Even good GCSE/A level examinations only produce grades, which give approximate estimates of educational achievement.
- 3 Public examination grades depend upon a range of contextual factors, and carry the greatest level of meaning, when interpreted within that context.
- 4 Because the curriculum to which examinations relate is changing, comparisons of examination grades across years can be very misleading.
- 5 Because grades can only be defined in relation to specific subject-based syllabuses, they do not have much accuracy when compared between examinations in very different subjects.
- 6 Those who specifically want to answer questions about change in educational standards over time certainly won't find dependable answers from analyses of UK public examinations results.
- 7 The way forward is not to try to turn these examinations into an 'exact science' in order that their results will become exactly correct, true and unchallengeable. That is to go searching for 'a pot of gold at the end of the rainbow', which simply isn't going to be there. We need instead to value public examinations for what they are good at, and look elsewhere for other kinds of research data to try to answer the big political questions about how the UK education system is evolving in the context of major political initiatives to improve it.

Professor Roger Murphy

Centre for Research in Schools and Communities, School of Education, University of Nottingham.

Murphy, R (2004) Grades of Uncertainty: Reviewing the uses and misuses of examination results. ATL. London.