



# Looking for the best result

**Calls to replace external testing with more teacher assessment ignore a deeply corrosive feature of education in the UK: the volume of high-stakes tests facing students that are accredited, certificated and recorded**

Words **Alison Wolf**

In Britain, the discussion of assessment practice in schools is dominated by issues relating to formative assessment and Assessment for Learning. This speaks well of schools', teachers' and indeed researchers' priorities and concerns. However, this discussion is also often accompanied by a tendency to see formative assessment as "good" and summative assessment as "bad". Such a view obscures the importance and the value of summative assessments.

It blinds us to some key characteristics of British education systems. And it diverts attention away from key features of our current systems for testing schoolchildren, whose worst effects have little to do with the use of "summative" rather than "formative" techniques. In this article I will highlight the importance of formal and summative assessments, and then go on to highlight what I believe to be the most important problems with our current testing regime. Much of what is said will, inevitably, have greater relevance at secondary level: and as such, may well

confirm, for primary-based readers, that they made the right career choice!

**The summative advantage**  
Education is, of course, about learning. People enter teaching – at all levels – because they want to help students learn. But formal education is also, and intrinsically, about selection and certification. Students are very well aware of this. If you ask undergraduate or graduate students whether a university's reputation affects their choice, they look at you as though you were deranged – how could it not? If you then

ask them what they would choose if they had to pick either collecting the credential without learning anything additional, or learning for several years but without getting a formal degree, they find it very hard to decide. Rightly so. Your skills are crucial in determining your promotion and success in life – but it is the credential that gets you on to the shortlist and through the door. Everyone can, in principle, obtain a certificate. But selection does just that; it is about choosing some people at the expense of others. Societies may – hopefully will – offer multiple ladders to climb, second chances, re-entry points and all the rest. However, this does not mean that selection disappears. For teenagers in particular, school and education have, at their core, intense competition. Their contemporaries may be friends but they are also rivals. And as education becomes increasingly important in determining people's lives, this rivalry becomes increasingly consequential. How should selection decisions be made, if not on the basis of summative assessments? Do we really think it would be better to go back to a world where connections were the key determinant of life chances? Where you got into an Oxbridge college or a craft apprenticeship on the basis of family connections rather than performance and skill? We may still live in a world where family circumstances have an enormous effect on life chances, but without a system of national tests and certificates, do we really believe things would be better? Standardised summative assessments, marked in the same way for everyone, are necessary for any modern society, which aspires to fairness and justice for its citizens. For this reason, I get very uneasy about some of the claims made for teacher assessment. Of course, it is true that teachers know their pupils better than anyone else, and that many things are not well suited to formal examination. But we should not lose sight of an equally important point: allowing or requiring teachers to assess their students in high-stakes situations means giving them arbitrary powers, with good reason to doubt the fairness of the outcomes. University experience is relevant here. The most discussed assessment issues in most faculties have nothing whatsoever to do with formative assessment (or learning) and are, instead, plagiarism and the effectiveness of anti-plagiarism software. (This is also one key reason why centrally set examinations remain so important in most universities.) University students – or their representatives – also tend to be very keen on anonymised scripts, and I think they are right to be. It is only when I have finished marking coursework or exams, and agreed final marks with a second marker, that we

“

Students do feel under pressure to accumulate as many grades, credits, points and the rest as possible

”

find out the identity of candidates – and I am often surprised to find out who some of the high, or low, marks belong to. Can I be sure that, had I known the candidates' names when marking, I could have remained completely unaffected by my prior opinions and judgments? Maybe, but I doubt it – it is not the way the brain tends to operate. My first teaching experience was in an American university. I was let loose on a second-year class without any prior induction into what the standards for different grades were meant to be; no one checked the quality of my papers or the marks I gave. This remains common practice, and it still shocks me, as it did then. Arbitrary power is just that: "unrestrained, exerted at pleasure" as the dictionary explains. The fact that most of those wielding it are conscientious and well meaning may mitigate the effects, but does not change the situation. For the same reason, I have never shared the view that high-stakes final certificates based on teacher assessment must necessarily be more valid because markers who are teachers have more evidence at their command. What happens if the teacher doesn't like you? We also need to take note of research showing that it is hard for teachers to maintain common standards when they operate independently of each other, however hard they try. For example, in Sweden, where teacher assessments are used for high-school leaving certificates, there used to be centrally set tests that told each school where their pupils stood,



roughly, in relation to the national average. So they might discover that about a third of their pupils were in the lowest (or highest) quartile nationally: and would then use that information to anchor the assessments that they made, and the grades awarded. More recently, the country moved to a criterion-referenced system, in which teachers were expected to assess against written standards instead; and the result has been definite – and uneven – grade inflation (see references).

In Germany, teachers mark the examinations for high-school certificates, including the *Abitur* (the university entry qualification). Universities have to treat grades on the *Abitur* as standardised and equivalent across the country when admitting students. But work out by the Max Planck Institute in Berlin in the 1980s indicated clearly that this is not the case; and since then the results of the OECD Pisa studies, which enabled German researchers to look at *Länder* or state level, show big differences in secondary level attainment among the *Länder* that are not reflected in their *Abitur* mark distributions.

In this country, the relationship between pupils' grades on nationally set tests and teacher assessments at Key Stages 1, 2 and 3 is often very close. This is often advanced as evidence that we could simply do away with the national tests. But we also, and crucially, provide teachers with an annual update on what standards "ought" to be, and what content needs to be covered, through those tests. So this is a very different situation from one where teachers simply assess "to standards", without having tests (and results) available as benchmarks.

For all of these reasons, any large-scale move towards more teacher assessment



Summative assessments, marked in the same way for all, are necessary in a modern society



of a formal, high-stakes type is fraught with danger. Those who advocate it also, surely, ignore the impact it is likely to have within the classroom. In this country, teachers and students in schools can and do feel united in tackling the external quasi-enemy, in the form of the test-setter and examination marker. This sets up a very different classroom dynamic from the one in which the teacher directly controls marks that will affect the student's future. And if you are supposed to be carrying out summative assessment of a high-stakes sort – and trying to do it

as fairly and well as possible – will it really be feasible also to carry out good formative assessments that do not, in some way, feed into that summary grade? Or for students to believe that they do not, and take any risks?

**Volume matters**

The major problem for British – and especially English – schools is surely not the low status afforded to teacher assessment, or the importance of summative assessment. Rather, it is the sheer volume of summative and externally accredited assessment. Basing some of this accredited floodtide on teacher-awarded grades rather than test scores would make no real difference.

Somehow, in the past decades, this country parted from mainstream global practice. It is not the number of tests – or summative tests – that students take that marks us out. What is unusual about us, and increasingly corrosive, is the number of formally accredited, certificated and externally recorded test or examination grades that each student accumulates. When we note that we are the "most heavily tested" system in the world, it is in this more restricted respect that we are correct. And testing of this type has quite different consequences from within-class or within-school testing and examining (including within-school use of standardised, externally developed tests).

English pupils today accumulate a whole list of separate awards and "point scores" as they progress through secondary education, recorded on a pupil database. In most countries, there are one or perhaps two important, high-stakes awards in a pupil's career. Here, by contrast, it is not only that we have key stage tests, plus GCSEs, plus a variety of post-16 awards; but that we assign individual scores and grades to each of these. This process reaches its apogee at 18, when all sorts of things currently are awarded "points" by a process that it would be kind to describe as opaque. (Can anyone explain why a Diploma is worth as many points as three and a half A-levels? Why not four? Why not three? What does that statement mean?)

Students – and their parents – may not understand the details of levels, points, credits and the like. Like me, they may be utterly bemused by attempts to explain the assessment system for the new Diplomas. What they do feel, however, is that they had better accumulate as many points, grades, credits and the rest as possible: and the exam structure encourages them to do so, allowing modular entries and retakes in a way that is, again, internationally unique. Even though universities mostly take more notice of individual subject results than of any overall "points" score, there is still enormous pressure to obtain as many top grades as possible, at GCSE and A-level.

This has inevitable and harmful repercussions for the subjects people choose. Why risk taking a subject that you find difficult, when you can be far more confident of a B or an A elsewhere? A few years ago, the headteacher of one of London's most prestigious girls' schools proposed cutting drastically the number of GCSEs that pupils took, for the good of their broad education. The parents panicked: GCSEs stayed, the headteacher went. If current plans go through, this stacking up and adding up of discrete assessments will become even more marked – and distinctive – as England moves to a globally unprecedented and uniquely ambitious system of credit awards and accumulation.

**Yes, we should worry**

Given the defence I have just mounted of externally set examinations, does this, on balance, matter? Yes. First of all, the emphasis on accumulating points and passes is very expensive; second, it jeopardises assessment quality; third, it distorts the choices young people make; and fourth, it undermines learning.

The first point is obvious enough. Large numbers of externally developed and marked tests, along with their administrative requirements, require large numbers of individuals, and large amounts of time. The use of examination fees makes this transparent – but it would be true even if government paid for it all directly.

Second, this volume of testing pretty much guarantees that a lot of the tests will be fairly poor. Developing a good test is a costly and difficult business, there are only a limited number of people with the skills required, and ensuring marking consistency for anything that involves complex and stretching responses is itself a complex business.

To take the third point, although the official position has been and is that examinations at a given level are equivalent in difficulty, no one really believes this to be true. Nervousness over grades (and Ucas points) is a key reason for the low numbers taking maths, science and modern languages.

And fourth, the more tests and exams there are, the more time is spent teaching to them. Teaching to the test is not necessarily a bad thing, and in any case it is – like death and taxes – unavoidable. But the more discrete and separate tests you have, and the more you make all of them high-stakes (because they count towards a Diploma or carry points), the more you lose any balance in the curriculum. One can defend teaching to the test some of the time: doing nothing but teaching to multiple tests is utterly stultifying, for teachers, pupils, and for the intellectual health of a country.

Looking at current discussions of school-



level assessment from inside a university leaves one impressed in many ways. It is remarkable, in the sense of "worth remarking on", that so much time is spent thinking about how to use assessment in the classroom to promote learning. At university level, which accounts for an increasing part of the nation's education system, there is no such preoccupation. I wish, nonetheless, that the school assessment community would pay more discriminating and sympathetic attention to both summative assessment, and to our overall system of tests, examinations, qualifications and credit frameworks. They need the spotlight. ■

**About the author**



**Alison Wolf** is Sir Roy Griffiths Professor of public sector management at King's College London, and a member of the editorial board of *Assessment in Education* (Routledge)

**References**

**Further information**  
Wikstrom, Christina "Grade stability in criterion – reference grading system: the Swedish example" *Assessment in Education* 12.2 (2005) 125-144