

The Children of the Gap

Remember the great exchange of limericks on the Professional Assessment for Californian Teachers (PACT)?

There once was an assessment called PACT,
Which videotaped how teachers act.
They showed all their tricks
And were scored with rubrics;
And quality went up: that's a fact.

By Dr Linda Darling-Hammond

The sceptical student of PACT
Was not renowned for his tact.
He said, "I suppose
That nobody knows
If the measurements made are exact".

By Professor Gordon Kirk

You can guess that I favour the second limerick. David Laws, our Minister of State for Schools, probably favours the first one. He now wants to know what universities can do to help trainee teachers learn how to close the gap in performance between 'disadvantaged' children and the rest. It is as though the social and educational policies of government are not relevant: it is down to the teacher and those that trained the teacher.

Closing the gap has been an aspiration for a very long time. It keeps coming round every time we get new ministers with no relevant memory and very little relevant knowledge. In my view there *is* much that can be done at the stage of professional formation to help provide techniques and insights that could help to close a gap. To be effective, however, and to avoid simply showing trainee teachers the tricks of pushing a child temporarily up a grade or two, the work of HEIs needs to be part of a concerted, strategic and inclusive effort involving much change to how government perceives learning and the measurement of performance. Done in isolation during a period of policy turmoil it is less likely to be effective and may, therefore, be used in evidence against universities.

I am concentrating on the various public examination systems here because that is what I know most about.

To summarise the points I try to make below: it is no help at all if, when we try to close a performance gap, teachers have to become like tailors who are

trying to measure a fidgety customer with a tape measure that cannot make up its mind if it is using imperial or metric gradations. If David Laws is exercised about a gap in performance that he thinks can be addressed by Initial Teacher Education/Training then he should get the customer to stand still for a while and decide what and how he wishes the tailor to measure. Unfortunately, *customer* is a commonly used word these days.

He should also acknowledge the multiple changeable and fast growing numbers of variables at play, including the effects of very quickly coming off a relatively stable curriculum and assessment diet. School structures are changing very rapidly, as are conditions of service, examination systems, the curriculum and the imperatives of Ofsted. It is all unsettling. In a few years time with a variety of examination systems running simultaneously in different schools and varying parts of the country it will not be one gap we are looking at but several and as yet we are not able to predict their locations, nature or extent. What, however, we can be sure of is that parents, children, schools, schoolteachers and universities shall have caused none of this. It is government that places the fulcrum, owns the lever and has the force.

The changes to professional formation are also at very early stages. The relationships between universities and teaching schools have yet to be fully established and are hardly likely to be uniform, though we can hope for some commonality.

And the recent attempt by the right wing think tank Reform to show that increasing the level of resources invested has no effect upon attainment in schools should, I believe, be seen as a clear declaration of intent to lay the blame for low attainment at the door of professional educators (see Snoopy below).

Five Key Concepts (more later)

1. Differentiation by outcome.

When I worked with colleagues designing the new GCSE syllabuses (later called specifications), mark schemes, written examinations and protocols for coursework it was differentiation by outcome that we all had in mind. We felt that we had to fulfil the loudly announced and much admired philosophy of GCSE that all children should be provided with the opportunity to demonstrate what they knew, understood and could do. As a result questions became more open and rather than the old practice of looking for 'correct answers' examiners spent more time making critical sense of responses that might demonstrate unexpected but still valid knowledge, understanding and skills. To do this is to connect exams with education.

For me this is the best way of working. The question is, will Michael Gove's proposed changes to examinations, the curriculum and the way it will be taught prevent differentiation by outcome?

2. Differentiation by task.

There have always been people who believed that what we need are tasks matched to predicted levels of ability. Ofsted like this and it has become an orthodoxy. In my opinion, after he made his unexpected Café Royal speech in 1991 saying that he did not like coursework and preferred tiered papers John Major transformed GCSE from a means of supporting learning to a vehicle for grading. Writing tiered papers and mark schemes also felt like playing three-dimensional chess. I am sure that John Major had no idea how difficult it was to produce three differently graded but overlapping papers and mark schemes that each retained exactly the same balance across a range of both assessment objectives and syllabus content. You had to be absolutely equally fair to the candidates entered for each tier.

If Michael Gove wishes to get rid of tiered papers that will be good. If, however, he wishes exam papers to concentrate upon measuring the rate of regurgitation of his preferred facts then it will be bad because it will narrow and diminish the educational experience. As, of course, the drastic reduction (even abolition) of coursework already has by taking away and devaluing the development of the skills for sustained learning and research.

3. Accessibility.

This concept has to accompany differentiation by outcome. If we look at O and A-Level papers of more than thirty years ago you can often see why people say they were harder then. Sometimes they were linguistic obstacle courses. Let me give two examples from my own experience.

In order to prepare teachers for the GCSE in Government and Politics the Chief Examiner for A-Level and I, as the CSE Chief Examiner for the same subject, were asked to write a specimen paper and mark scheme. So I showed him one of my CSE papers. After reading it he remarked that I asked the same kinds of things that he asked but there was, he said, a difference. While I made clear to candidates the meanings of the questions his practice was to hide meaning to see if candidates were clever enough to decode the language of the questions (see Discrimination below).

Some time later after I became the Chief Examiner for GCSE someone invigilating one of my papers accused me of setting really easy questions. 'Anyone off the street could answer these questions', he said. I asked for an example. 'What influences the way people vote? Anyone could answer that'. I agreed but had to point out that they might not get many marks. 'What', I asked, 'do you know about the Michigan Voting Model?'

I am afraid that too many people see dumbing down where they should see that we are much better now at making both teaching and examining fair and accessible. The question is what do Gove and Laws want to see?

4. Discrimination.

This is what I fear we shall return to. In what I believe to have been the bad old days of examining when recall of facts was more important than analysis, interpretation and evaluation we tended to look for a rank order. All that was required was a means of discriminating somehow between candidates; hence linguistic obstacle courses.

Doing this is to remove examining from being part of a learning process and reduce it to a simple grading process. Is this what Government want?

5. Breadth and balance.

This phrase is usually reserved for the curriculum but I want to emphasise its use in looking at the range of knowledge, understanding and skills we should be examining. To remove, as is currently proposed, Speaking and Listening from the range of things to be examined in English GCSE is to damage the subject and diminish the learning experience of children. To remove it half way through a programme is a disgrace.

The question is, how broad and how balanced will be the range of knowledge, understanding and skills that Government wishes to see demonstrated and examined?

Additional Concepts to Confuse and to Vex

Now for the concepts of criterion and normative referenced assessment: the questions to bear in mind are a) shall we have un-researched norms (really, expectations) imposed by government or b) shall there be time to develop an understanding of criteria so that useful and reliable norms can be discovered?

It is easy to become confused about norms and criteria because they are each used in different ways and, most important of all, they must be constructed with great care: they offer a sense of certainty, robustness and reliability but sometimes all of this can be false. The consequences of badly constructed norms and criteria are most directly, but not exclusively, borne by children. So, the following is my attempt to create a little clarity about them, plus something on that vexed concept, 'standards'. Quite possibly I shall create confusion where previously none existed. I shall try not to.

A **criterion** is perhaps best seen as a statement of something that can form the basis for the construction of a judgment (an assessment). In the education field

we are probably thinking of statements (criteria) about items or kinds of knowledge, depth or nature of understanding and specific skills; even, at times, personal attributes. It becomes more complicated when we introduce grading because then we are ascribing differential value and making judgments in another way: introduce descriptions of grades and you have to get them all equal and connect them to what is known about natural rates of learning and progression. Gaps between grades that have been designed to be equal are often found not to be. Since we are also talking here about sticking labels upon children that will be there for life it is important to get this as right as possible from the outset but also to keep things under review.

Nevertheless, grade or level descriptors are merely words to help us break down bigger criteria into smaller criteria. Different systems often use different language such as *assessment objectives* and *attainment targets* for what are largely the same things. However we frame the definitions of criteria and however we deploy them, graded or not graded, they are starting points for making sense of learning. They are statements to which we can refer when constructing judgment.

Too often people from around the country who do not know each other have been put into a room for a few hours to write grade descriptors. Whenever I took part in such an event trying, for example, to nail down descriptions of what might constitute performance at G, C and A for GCSE you could place bets that the G descriptor would contain the phrase 'a basic knowledge of' while the C descriptor would say 'a more developed knowledge of' and the A descriptor would let rip with words such as 'complex', 'detailed', 'wider' and 'deeper'. The number of words would grow as we ascended the alphabet. That we really had very little notion of systematic relevant research relating to levels of performance did not matter because by being printed in a syllabus what we wrote was treated as gospel.

In education **Norms** are the shape and spread of performance results we can expect to see repeated if we continuously maintain the same kinds and conditions of learning and examining. They provide us with a recognisable pattern so that if anything unexpected and unusual happens we can spot it quickly. But what might be the significance of the unexpected and the unusual? A properly run examination system would have taken a lot of time and trouble to arrive at assessment criteria that are well constructed, well and widely understood and trustworthy. So when or if it encountered the unexpected and the unusual it would stop to think about, maybe to reconsider, how it described sets of knowledge, understanding and skill. A good examination system would be a learning system or, rather, a system that was willing to submit itself to testing and allowed itself to learn.

Supposing, however, the assessment criteria, especially the criteria for different grades, were put together in a hurry in order to meet the demands of a new secretary of state and dominating the entire process was a partisan belief that those that attained certain highly approved grades were worthy of being labelled 'successful' and their teachers celebrated as 'outstanding' while those that fell short or were simply different were to be condemned as failures. In such

circumstances it will be schools, teachers and, most prominently, children that would come under pressure to learn not to engage or interact with the system so that it can gain new knowledge and see things from different perspectives but, instead, to adopt its official values and fit in with its official vision of learning: to adopt and conform to given, rather than researched, norms: a case of the square peg submitting to the force of a crude but powerful hammer that will ensure that it goes through a round hole. In a system like this it does not pay to be different.

An aside

In 1974 I entered the Open Tenor class at the Liverpool Music Festival and won first prize. There were only two people entered and the other fellow never turned up. Some years later when I told my wife about this great achievement I said that it was a case of criterion-based assessment. She replied that, on the contrary, the assessment had been entirely normative because not only had I come first but I had also come last. Collapse of stout party!

Now for that favoured word of politicians and the media: **standards**. It is probably in the job description of every new secretary of state for education that they are required in every speech either to make reference to their intention to 'drive up standards' or to the 'disgraceful lowering of standards' during the tenure of their predecessor, especially if they belonged to a different party.

I sometimes picture Michael Gove as the Wolfie Smith of the Conservative Party, raising his right fist in salute and shouting 'Power to the Pedants'. He really does like to catch out educationists (or does he say *educationalists*?) slipping up on their syntax and spelling and choice of words. We might, therefore, gently chide him about standards for, you see, they are supposed neither to go up nor to go down. He wants to 'drive them up', thereby making clear that until he came along they had been falling but if they are any good, if they are based upon some decent research and if they are such that the profession can believe in them then they simply have to stay in one place long enough for useful comparisons to be made about performance.

Like the tailor's fidgety customer our educational standards refuse to stand still. Or, rather, successive secretaries of state refuse to allow them to. It is as though their only purposes are to provide rhetorical flourishes for speeches and, too often, sticks with which to beat professionals. I simply cannot see a useful educational or social purpose in standards that won't stay put.

I believe we need slow cooking when developing new examination systems but I fear that we shall get microwaving from frozen. That, quite simply, is how today's politicians operate.

And if the examination systems proliferate in future with some of them sticking to un-researched norms taken off a party political shelf while others opt to base their assessments upon barely understood criteria we shall have got ourselves into a fine stew.

Some background on the gap

Keith Joseph used to agonise about the 'bottom 40%' and ask what should be done. That was one reason why Records of Achievement were pushed hard in the 80s. The problem for Joseph was, however, that he and his predecessors had created a system that required the existence of that gap: they needed losers in order to have winners.

First, the 1944 Act gave us a school and examination system that was based upon a eugenic belief that only about 20% had the ability to go to grammar schools and take School Certificate (later O-Levels). Then we got CSE that was supposed to be for the next 40% down. Except that there was no way of measuring or discovering how this notional 40% of 14-16 year olds could be identified as a group. The fact that CSE was a far more educationally sound programme of study and assessment was brushed to one side because we were told to think of it as aimed at a clearly defined inferior group.

Second, as a further confusing factor we teachers were told to imagine that a Grade Four (roughly equivalent to an F in GCSE) CSE represented the performance of the average 16 year old. The big problem with this was that at the time about 80% of children left school at 15 without ever following a programme that could have led to an award. The average was, therefore, totally imaginary and, as research information, useless. In terms of labelling children it was, however, very powerful. If there had been any truth at all in this average then it would have made an unjust nonsense of the current requirement that all children should be performing at the same level as that 20% for whom the starred As to C of GCSE were, in the past, reserved. We have a system that separates while also requiring everyone to be, at the same time, on the top of a greasy pole.

I think that some people seem to believe that the imaginary average of CSE can be used to establish believable performance measures. It can't. Picture a PhD candidate telling their examiners that the statistics underpinning their findings were based on holding up a wet thumb in the wind: a variable wind.

Third, GCSE was designed to combine O-Level and CSE. In other words, it was designed for the so-called top 20% plus the next 40% leaving out the remaining 40%. I hesitate to use the language of Keith Joseph and refer to them as the 'bottom 40%' because all of these assumptions about groupings were not, and could not be, based upon any testing of complete cohorts. If, however, there was anything substantial in the notion of a bottom 40% we would have to explain how the GCSE duvet designed to reach our feet when we were ten years old was supposed to reach our feet when we grew to fourteen and older. It was designed for an imagined 60% but eventually deployed for 100%. I have to admit though that I loved the early GCSE as a support for learning but it had its downside (next paragraph).

When I was a chief examiner for CSE, 16+ and GCSE (later chair of a number of subjects) the distribution curve was everything. The key question was where did we draw the boundaries last year? We had to keep the same shape on the

curve or else. I would be asked for very good reasons if I changed a grade boundary by even 1%. This was how we demonstrated that we maintained quality. The norms to which we stuck so rigidly were those very same norms that told us A-C was for the 20% of children who would have got O-Level and the rest was for those lower 40% assumed to be where CSE had been. I can remember, at the very first GCSE Award of Grades meeting for my subject, in answer to the question of where we should draw the boundaries, saying that we had to remember that Grade Four CSE (F in GCSE) was what for years we had been told to think of as the average performance for someone of 16. That helped to give us our normative curve for GCSE.

Schools, however, were increasingly being told to enter complete cohorts so GCSE began to cater for all without any pause for thought about this. On the one hand we had a system designed to exclude a percentage (we might call them The Children of the Gap) and on the other hand increasing pressure to push them up the greasy normative pole.

Meanwhile, the Assessment of Performance Unit (APU) was abolished. In *The Politics of the Curriculum* (1980) Denis Lawton had some very good arguments against the APU's approach to assessment and saw it as an example of increasing central control but at least it was an attempt at discovering some sort of believable expectations of performance for all children. Since its abolition we have only been given the ever-changing un-researched norms and expectations of politicians. As a result, failure to achieve a C grade (which politicians value so highly) negatively labels children for life and does a lot of damage to the careers of their teachers.

My experience of using criteria in both GCSE (in the form of grade descriptors) and Key Stage One SATs (in the form of level descriptors) tells me that it can take an age to arrive at a common understanding of what they mean. Grade boundaries drawn by people uncertain of the meanings of criteria simply wobble around.

As for the levels achieved in National Curriculum assessment, they too are unreliable. Before it was decided how to assess Key Stage One, as an LEA advisor I was co-opted to a government funded research team led by the University of Leeds. Armed with research SATs and an awareness of the Attainment Targets and their level descriptors we worked with 7-year-old children in schools in Leeds and, in my case, Liverpool. At that time each Attainment Target had ten levels. One day I gave a piece of unseen writing to a girl to read aloud. The writing was full of speech marks, question marks, exclamation marks and multi cultural nicknames. She read it brilliantly and I searched in vain for a level descriptor at the top of the page that said, 'Not even Anna Ford could do better than this'.

Now it is possible that this girl was so exceptional at age 7 that she ought to have been excluded from consideration of results. But when something like her performance happens it reinforces the feeling that we ought not to be rushing too quickly to establish norms: that we ought to do a bit of research first. Back in the LEA I raised this problem with the senior inspector for mathematics. He

explained that I should see the level descriptors as criteria but that they must be used over a period in order for us to discover the norms that they would generate; simple stuff, but for me in those days an insight because we had been brought up as professional educators to believe what we were told about the meanings of grades and levels.

By this time Kenneth Clarke was Secretary of State. I naively expected him to take all of this seriously. He did not: the usual impatient politician. He simply told us that Key Stage One would be limited to the bottom three levels and that he expected to see the average in the middle. When the results came out he complained that too many children were below average. In English he said that those below average were illiterate, despite the fact that the level descriptor for Level One Reading said, 'Can read with some assistance'. (That applies to most of us, even secretaries of state). I think that when his primary teacher taught averages to his class he must have been off with a cold. Expecting every child and every school to be simultaneously above average is a typical obsession of politicians.

Then, to compound this nonsense, his successor, John Patten, allowed KS1 to go up to Level Four. Lo and behold, some children (the ones who had been banging their heads on the ceiling of Level Three) went up to that level. He declared that because of government policies there had been an improvement. You could not make it up.

So how would I respond to an invitation by David Laws to show how Initial Teacher Education can contribute to closing the gap between the performance of the 'disadvantaged' and the rest? I would 'box clever': very clever. It is politicians that create grading systems. Schoolteachers and academics did not create the gaps that they wish us to close.

And if I got the chance I might sneak in the revolutionary point that one way to improve the performance of the disadvantaged is to stop ensuring that they remain disadvantaged. I doubt, however, that a millionaire minister would know what I was talking about.

As Snoopy of the Peanuts cartoon once said:

**'It matters not if you win or lose
Its how you place the blame'.**

May 2013

For a comment by Kathleen Tattersall on recent proposals by Michael Gove to change public examinations click on the link below.

<http://www.guardian.co.uk/education/2012/sep/18/exam-define-failure-not-success?newsfeed=true>