



“

Strategies such as getting students to assess each other have been shown to be effective and work in the classroom

”



In an essay on assessment, Paul Dressel described a grade as “an inadequate report of an inaccurate judgment by a biased and variable judge of the extent to which a student has attained an undefined level of mastery of an unknown proportion of an indefinite material”.

This captures some of the inadequacy of an activity that is often a taken for granted. Teachers assess every day, but how much of what they do is because it has always been done that way and how much can really be defended?

What do the results of assessment actually tell us? Can assessment do all – or indeed any – of the things that we expect of it? To answer these questions we need to clarify a few definitions about the nature of different kinds of assessment.

The key distinction that is usually made is between summative assessment, which records achievement and formative assessment, which informs the teaching and learning process.

However, as Paul Newton has pointed out recently, this distinction is neither clear nor really helpful. Instead, he lists 19 different

purposes of assessment, which include its uses in:

- selection;
  - qualification;
  - identifying learning needs.
- But he adds others such as:
- diagnosing specific learning difficulties;
  - licensing professional or other practice;
  - holding schools to account;
  - monitoring standards;
  - informing decisions such as school choice;
  - target setting;
  - placing students or allocating resources.

He makes the point that although an

assessment can often be designed to fulfil a single purpose satisfactorily, expecting it to do too much may lead to it not really doing anything very well.

#### Controlling influence

Newton’s list is concerned with the ways assessment results can be used. A slightly different kind of “purpose” of assessment that he does not mention, however, and that is seldom mentioned in discussions of this type, is its role in controlling the curriculum.

On the basis that “what gets tested gets taught”, it could be argued that one of the main functions of assessment, at least within schools, is to control what does get taught. Teachers may also use assessment to control their students. For example, knowing that a piece of work is going to be assessed can make them more likely to complete it.

Another distinction that is commonly made is between criterion-referenced and norm-referenced assessment. In the former, a particular level of achievement is defined in terms of criteria to be demonstrated; if the criteria are met, the candidate has

reached the level. In the latter, a level is defined in terms of the standard reached by a particular proportion of the population.

However, the problem with criteria is that they always have to be interpreted, and the difficulty of a particular interpretation depends on many factors other than the criterion it aims to assess. Even in a subject such as maths, where specific learning criteria are relatively easy to define precisely, apparently superficial features such as the complexity of the language in the question and its context can affect its difficulty.

Moreover, the precise level of difficulty of a particular question can be very hard to judge *a priori*, even for an expert. For this reason, supposedly criterion-referenced assessments in practice always depend on some kind of knowledge about how hard a particular group of candidates have actually found them (or questions like them) to be – which is, in effect, a kind of norm-referencing. Hence the distinction between the two is not as clear as may sometimes be suggested.

One implication of this is that using assessments to judge whether standards of achievement are being maintained is actually very difficult. Unless we can use the same assessment each year it is far from straightforward to ensure that the level of difficulty is the same.

If we add in strong political pressures

to demonstrate that standards are in fact rising, then it becomes even harder to find trustworthy data. Two examples may illustrate this. In the US, dramatic rises in scores on a state test, particularly for black and latino students, were celebrated as the “Texas Miracle”. However, other “low-stakes” tests, taken as part of a national monitoring project, showed the gains made by minority students in Texas were generally not even as good as those in other states.

A similar story emerged in England, when Peter Tymms compared the dramatic rises in key stage 2 scores with changes on a range of other tests. All the independent tests showed at best tiny increases in reading performance, in contrast to the rising percentages reaching level 4 at KS2.

#### Largely ‘illusory’

As he has written: “Neither the impact of Ofsted, nor the national curriculum, nor target setting, nor homework clubs, nor the national literacy strategy, nor the plethora of other initiatives, which together have cost hundreds of millions of pounds have had an impact. To use the word from Qualification and Curriculum Authority’s own evaluation, the apparent rises in reading scores seen in national tests were largely ‘illusory’.”

Part of the problem with interpreting the changes in assessment results in both Texas and England is that these tests are

# What is assessment for?

Words **Robert Coe** Photographs ??????

**Our new series debating what assessment should be and how it should be used begins with an examination of assessment for learning**

used to support systems of accountability. The performance of teachers and schools is judged by the achievements of students on the tests. Given that assessment becomes a real focus of attention in such systems, it is worth considering what effects this kind of accountability may have. It seems obvious to many that holding people accountable will improve their performance, though the evidence for this is pretty thin, and there may be a number of reasons why such systems can be counterproductive.

One problem with the use of assessments for high-stakes accountability is that teachers are tempted to teach to the test. Of course if the test is really good this may not matter too much, since it can measure valued learning outcomes such as deep understanding, application of ideas and creative thinking. Too often, however, such assessments are routine and relatively predictable tests of recall, so that the high stakes attached to them serve to narrow and impoverish the curriculum.

Another problem is that such systems create a market in “desirable” students at the expense of others. If high achievement is what counts, then every school will want the kinds of students who are most likely to do well in the tests. Attracting such students, along with discouraging or removing those who are unlikely to do

well, may become a more profitable – and certainly easier – strategy than genuinely improving the education of all. Those who are most educationally at-risk are unlikely to be wanted by any school, so the system acts to compound their exposure.

### Teaching difference

A third problem is that we are by no means certain how much difference teachers and schools can in fact make. Of course, most people would agree that you can't hold schools responsible for the raw achievements of their students; differences in ability and achievement on entry predict these well and are clearly beyond the control of the schools. So most accountability systems adjust for these and perhaps other intake characteristics, using some kind of value-added model. However, different models can produce quite different results and many estimates of “value-added” correlate highly with raw results anyway.

Overall, the use of assessment for accountability seems rather questionable, therefore. A much more welcome development is the recent focus in many countries on “assessment for learning”. Of course, the use of assessment to inform teaching is not new, but the explicit focus on the ways feedback from assessment can support learning, backed up by solid

research, is. This development owes much to the work of Paul Black and colleagues, who not only collected the evidence about the potential benefits for learning, but also worked with teachers to develop strategies to achieve them in classrooms.

These ideas have been taken up by governments and teacher groups around the world to introduce policies to encourage “assessment for learning”. Strategies such as getting students to assess themselves and each other, widening teachers' opportunities to receive feedback about their students' learning, and optimising the kinds of feedback they provide for students about their learning have all been shown to be effective although, as with most strategies that work, the difficulties of changing to this way of working should not be underestimated. Nevertheless, widespread promotion of an approach with such a sound basis in research is quite unusual in education, so this certainly seems to be a positive development.

### Meet the expert



**Robert Coe** is director of secondary projects at the Curriculum Evaluation and Management Centre at the University of Durham



## Sources

**Black, P and Wiliam, D (1998)** *Inside the black box: Raising standards through classroom assessment* (King's College London School of Education)

**Dressel, P (1983)** “Grades: One more tilt at the windmill” in *Bulletin*, edited by A W Chickering (Memphis: Memphis State University, Center for the Study of Higher Education, Dec 1983, p12)

**Klein, S P; Hamilton, L S; McCaffrey, D F; and Stecher, B M (2000)** “What Do Test Scores in Texas Tell Us?” Education Policy Analysis Archives, Vol 8 No 49 ([epaa.asu.edu/epaa/v8n49](http://epaa.asu.edu/epaa/v8n49))

**Newton, P E (2006)** “Clarifying the purposes of educational assessment” *Assessment in Education*, in press.

**Tymms, P (2005)** “Dyslexia Is Not a Disease to be Cured” *Times Educational Supplement*, 4 November 2005 ([www.tes.co.uk/section/story/?story\\_id=2156860](http://www.tes.co.uk/section/story/?story_id=2156860))