

An apple for the inspector

It is impossible to rid assessment of inconsistency and personal preference, says **Andrew Davis**, so politicians really must put an end to using test results as a means of judging schools

For years teachers have begged government to reduce the burden of testing, and to stop using exam results to judge the quality of schools and teaching. While some tests have gone in the last year or so there are plenty left. So, with a general election looming, perhaps it is worth making the effort once again to persuade those in power to mend their ways.

Markers grading national tests are supplied with assessment criteria. 'For 12 marks, the story has a beginning, a middle and an end.' 'For 15 marks the writing is lively.' All this ensures we can meaningfully compare schools, teachers and pupils on the basis of test performance. Without it, school league tables would make little sense. Anyway, that's the official story...

There are some cans of worms lurking around here. To discover them consider the following activity. Over the years I've tried it with many students and teachers. You are to supply criteria for grading apples on the following scale: Excellent Apple, Good Apple, Sound Apple with some Weaknesses, Failing Apple. My student victims usually wade in with initial enthusiasm. However, after a few minutes they begin to protest. 'Some people like them crisp! But not everyone! My mother prefers them mushy.' 'It depends on the variety!' 'What are you going to do with them? Cooking apples are different.' I nod soothingly, and suggest something similar for pears. Finally I request new criteria to apply to both fruits. 'But this is impossible – our descriptions are so bland now that they are almost meaningless! We're leaving out what makes an apple an apple!'

The students sometimes smell a rat. 'Oh no! You're not implying that our fruit criteria resemble criteria for school tests! That's ridiculous. Obviously fruit preferences are just opinions. Grading exam answers isn't like that! Or if it is, someone should put a stop to it!'



‘How certain are we that everything of any educational importance can be captured by consistent grading?’

Andrew Davis, Research Fellow, School of Education, University of Durham



‘We create a special food regulator, Ofgrub. Ofgrub lays down what counts as excellence whether in terms of juiciness, pip colour, sheen or whatever... Fruit graders learn to set their personal preferences to one side.’

Yes – criteria for assessing school learning differ from the infamous fruit grade descriptions. All the same, let’s stay with fruit just a little longer. It would be possible to train several people to judge fruit in the same way. But what criteria would work here? Just how is consistency going to be achieved? You might think that the solution would be to exclude any descriptions linked to personal reactions. So ‘delicious’, ‘refreshing’ and so on would be banned because different people find different things delicious.

However, we couldn’t stop there. For even colour, size and shape, seemingly objective features, can be linked to personal sentiments – some people like green fruit and others do not. So if we went down this road we would eventually run out of descriptions for our fruit criteria. Prohibiting descriptions tied to people’s opinions isn’t going to work.

The alternative is an exercise in power over fruit thinking. We create a special food regulator, Ofgrub. Ofgrub lays down what is to count as excellence, whether in terms of juiciness, pip colour, sheen or whatever. Fruit graders are trained accordingly. So the cost of this exercise is that graders learn to set their personal preferences to one side. Ofgrub imposes its views on everyone.

Some readers will react enthusiastically to this result when transferred to school tests. ‘That’s great! We certainly don’t want mere personal opinions influencing how markers assess. Whatever next!’

This response is too quick. Opinions in educational assessment arguably have a different status from reactions to fruit. Consider the criteria for writing tasks included in national curriculum English tests for 11-year-olds. They include things like ‘Length and focus of sentences varied to express subtleties in meaning and to focus on key ideas’ and ‘All aspects of the story are consistent and contribute to overall impact’.

A striking number of phrases here relate to opinion – and could theoretically lead to disagreement between markers – what one person might deem an effective variety of sentences expressing subtleties in meaning, another could find irritatingly artificial. Yet the opinions are not mere opinions in quite the sense that reactions to fruit are just personal in the end. Can’t equally well-informed professionals legitimately disagree here on occasion and isn’t it possible that each of them could support their judgements with reasons?

At the same time, we cannot live with significant discrepancies in judgements between different markers when the test results are so important for schools, teachers and pupils. Such differences would drive a coach and horses through the fairness of the process. Hence anything liable to provoke disagreement will have to be ironed out of the criteria themselves or excluded during the training of the markers. And certain opinions will have to be endorsed officially. Markers will have to follow suit, or find another occupation.

Now it would be foolish to encourage inconsistency. Yet at the same time we should be very concerned about what might be sidelined in a single-minded pursuit of consistency. We are told time and time again that what is not assessed will not be taught. So what cannot be assessed consistently will not be assessed at all. How certain are we that everything of any educational importance can be captured by consistent grading? Whose values are being imposed in the drive to ensure consistency, and with what justification? What is being excluded from education because of the way assessment is still being used? Can education survive assessment criteria in our high-stakes system?