

Perceptions of Quality in Educational Assessment

Colin Robinson, Education Consultant, UK

Email: colinrobinson@tinyworld.co.uk

Abstract

Huge effort is made by those constructing and delivering educational assessments to ensure that they are as valid and reliable as possible. However, those who receive the results of their efforts – the candidates, teachers, employers, university admissions tutors and the general public – have very differing views of the quality they represent. Candidates are keen to obtain the highest grade they can, but how does the grade they receive affect their perception of the quality of the assessment itself? If the grade is lower than they were expecting, they may well conclude that the assessment was flawed. But are they similarly sceptical of the quality of the assessment if the grade they received was higher than they expected? Are teachers more critical of the quality of the assessment if the grades their students receive are lower than they had hoped? Are they equally likely to raise concerns over the quality of the assessment if the results are higher than they expected? The paper explores the factors that may influence different stakeholders' judgments of assessment quality and discusses what response those involved in educational assessment should be.

Keywords

reliability, validity, stakeholder perceptions

Perceptions of Quality in Educational Assessment

Colin Robinson, Education Consultant, UK

Introduction

For everyone involved in the delivery of assessments, quality is of prime importance. Rigorous systems are put in place to ensure that the source materials are appropriate, the questions are unambiguous, the examiners and markers are properly trained and the results meticulously scrutinised.

Yet despite all our efforts, every year we see complaints in the media – from teachers, business people, university tutors and from politicians – that standards have not been maintained or that the results are unsatisfactory.

One of the problems with all of these concerns is that they are primarily subjective. Very few are based on a scientific analysis of the data – which is what the awarding bodies and the regulators rely on. In this paper, I shall look at some concerns discussed in the media in the UK and try to distinguish between what are legitimate concerns and what are baseless fears.

Students

Of course students want to receive the best grade they can so, to an extent, the higher the grade they receive, the “better” they will regard the assessment. But in fact they are capable of much more sophisticated judgements regarding the underlying quality of the assessment itself.

Immediately after taking the assessment, they will probably have a good understanding of the relationship between what they have been taught and what they are assessed on. If these match reasonably well, they will accept that the assessment is, at least, a valid reflection of the curriculum. If there are aspects of the assessment that they do not recognise, it will raise questions in their minds as to whether the assessment was fair, though of course, there are plenty of reasons why that may still be the case even if everything on the paper was taught: the student may have missed some lessons or may have been daydreaming. In these circumstances, the student will probably check with others to see if they have also had a problem before raising it as a concern.

Students are also in the best position to gauge how well the result reflects their actual attainment, though they may not wish to admit it even to themselves. Of course they will be disappointed if they get a lower grade than they were expecting and elated if the grade is higher, but they will judge its accuracy by a range of criteria:

- was I realistic in my expectations? (Girls tend to underestimate their potential, whereas boys tend to exaggerate.)
- was I (un)lucky in the choice of questions?
- were there aspects of the assessment that I had not really mastered?
- did I answer all the questions as fully as I should?

Students are willing to accept a degree of imprecision in the results. If they expected a C, they'll not be too surprised with a B or a D, though those results may have very different consequences for them. If they receive an A, they are more likely to regard it as a happy accident but an E as a failure of the system.

Students rarely take examinations in total isolation, so they will discuss their result with their friends and classmates. This adds a further perspective to their judgement of the assessment as a whole. They have a clear view of their own ability, attainment and dedication in comparison with their colleagues. Their view will be influenced by how far the assessment corresponds with the rank order they have in their heads. If people they regard as less able (or less worthy) have received higher grades than they think they deserve, they will start to doubt the reliability of the assessment. If those they admire as “high fliers” receive low grades, their sense of fair play will reflect badly on the quality of the assessment.

Teachers

The majority of criticisms tend to come from teachers and schools, both from individuals and from their associations. Obviously teachers are concerned that the results are accurate, but they frequently look at them from their own perspective: what does it say about my teaching?

What is it that teachers expect from assessment?

Teachers use assessments in their everyday teaching. They are constantly checking to ensure that their pupils have learned what they are trying to teach. They recognise that this feedback is variable: they know that their pupils have good days when they appear to have grasped what they have been taught and bad days when they seem to have forgotten everything. When it comes to external assessments, however, they have different expectations. Deviations from their own assessment are regarded as indications of failings in the external assessment rather than as inconsistency on the part of the pupils or problems with their own assessment.

In 2009, a great deal of concern was expressed about the results of the writing task in the national assessment of 11 year olds. A number of schools returned their pupils’ scripts and asked for them to be remarked. The main source of their complaints was that pupils that teachers had assessed at a higher level received a lower level in the tests. They also identified discrepancies between the reading and writing results for individual pupils.

Many different factors may have contributed to the results. “Some blame the question; others say their marker was not up to scratch; others have said the mark scheme itself was poor.”¹ Of course each of these aspects would have been checked during the process. The question and the mark scheme were discussed by the examiners, trialled with a sample of pupils and those results analysed before it was used in the test. Of all the scripts submitted in 2009, schools asked for a remark of 50257 – a large number, but still only 3% of the total. And of those reviewed, only 13% - 6,532 scripts - received a revised grade. So 87% of the scripts the teachers returned were not regarded as inappropriately graded, and presumably, most of the teachers responsible for the 97% of scripts that were not returned, were reasonably satisfied with the result.

As Professor Roger Murphy, has indicated, assessment is not – and cannot be – exact. ‘Give the candidate a different sample of tasks and almost certainly they will produce a different performance. ‘Give them the tasks on a different day and their performance may vary again. ‘Give their responses to a range of different examiners to mark and again you will probably get different judgments.’²

Schools

A further complication for schools arises from the Government’s decision to use the results from the GCSE as the basis of school performance tables – so-called “league tables”. These tables are supposed to give parents an objective picture of the performance of their children’s schools. All state-funded schools are included in these tables but it is also open to independent schools to have their figures included, provided they use approved qualifications.

¹ Times Educational Supplement (TES) 16 October 2009

² <http://www.dailymail.co.uk/news/article-1268199/A-levels-GCSEs-HAVE-got-easier-says-Cambridge-exams-chief.html#ixzz0rCzOhJf4>

One of the issues frequently raised is not so much about the quality of the assessment and more a political issue of what should be taught and how pupils should be assessed.

When the General Certificate of Secondary Education (GCSE) was introduced in 1988, there was a compulsory element of coursework in almost every subject (in mathematics it did not become compulsory until 1991). Previously, the General Certificate of Education (GCE) O-level examinations, which catered for the top 20% of the attainment range, were assessed almost exclusively using paper and pen. On the other hand, many subjects in the Certificate of Secondary Education (CSE), which was designed for the next 40% of the range, included a substantial proportion of assessed coursework. The rationale for the inclusion of coursework in the new GCSE – designed for the full spectrum of ability – was the belief that some aspects – in all subjects - where

This requirement for assessment of coursework, to be undertaken by teachers throughout the two-year course, required considerable resources to train the teachers to carry out the assessment and to ensure that their work was properly monitored and moderated. In England, Wales and Northern Ireland, the boards had no choice, but examination boards that offered their examinations overseas sought a mechanism by which they could continue to offer qualifications that resembled the new GCSE, but did not have the coursework element. Also, because they were international, they did not need to conform to the requirements of the National Curriculum.

So since 1988, qualifications such as the IGCSEs – the I standing for “International” - have been very successful around the world, and have been accepted by organisations in the UK as being the equivalent of their UK counterparts. However, the only schools that were able to offer them in the UK were independent schools and colleges that did not receive public funding; state funded schools were barred from using them.

And because the IGCSE was not approved, state-funded schools could not provide courses leading to the IGCSE and independent schools could not include IGCSE results in the tables.

Last year, the then Labour Government relaxed the rules and allowed the IGCSE to be included for subjects other than English, mathematics and science because approval “would present a risk to the focus on the vital curriculum elements. For instance, young people would be able to opt out of answering questions on Shakespeare.”³ The new Coalition Government, which came into office after the indecisive elections of May 6th has promised that all IGCSE syllabuses would be accepted in future.

The point here is that this is not an issue of the quality of the examinations; it is just a question of what you expect from the examinations. The responsibility of the examination board is to assess validly and reliably. If the examination boards are given total freedom, they must specify what they are assessing so that teachers and their students can choose the assessment that matches what they are teaching and learning. If the examination boards are required to assess a government-determined curriculum, judgements on the quality of their assessment must be based on how validly and reliably they do that.

There have been claims in recent years of a decline in the quality of science examinations in the GCSE. Some of these are justified, others based upon a difference of opinion as to the purpose of the assessment.

³ Iain Wright, Schools Minister in the Labour Government, quoted in TES Nov 6th 2009

When the GCSE was introduced, there were great concerns that a high proportion of students were leaving school without any qualification in science. In an increasingly technological world, it was regarded as essential that *all* students should study *all* three areas of science. There was also a tendency for even the most able girls to opt for biology rather than chemistry or physics. To combat this, a new type of science was introduced into the GCSE. For most students, an integrated approach to science was offered, rewarding them with the equivalent of two GCSE grades. Those who wished to do so, however, could study the separate sciences – but in state-funded schools this was only allowed if they studied all three, a particularly refined group.

This resulted in a massive increase in the number of students achieving a science qualification. However, there was a problem: there was a decline in the numbers of candidates opting to take the sciences in the sixth form. Was the new Double Award qualification the cause?

Of course the new qualification had to cater for a very wide spectrum of attainment. To cope with this a differentiated approach was used, with different combinations of papers allowing access to only a small range of grades, some excluding the higher grades, others the lower ones. Inevitably some of the assessments in the lower tiers were much easier than the earlier examinations – the new qualification had to reward candidates for more limited attainments - but this led to claims that the examinations had been “dumbed down”. Again, the issue is not really one of quality. We can have a debate as to whose responsibility it should be to determine what is taught or what children should learn, but it surely is not the examination boards’ – their responsibility lies in providing an examination syllabus that accurately reflects the curriculum, and then making the assessment of it as valid and reliable as they can.

It is clear that not enough students were progressing from science at GCSE to one or more of the sciences at A level, but it is difficult to pinpoint the cause. It may be that the study of a more wide ranging science curriculum did not prepare students adequately for the separate sciences at A level. But this should have resulted in a drop in attainment at A level, not in a reduction of the numbers opting for the subjects.

Another explanation may be that other subjects were perceived as easier. To students, this may be attractive as they seek to maximise their results. Teachers may also wish to maximise the results in order to enhance the school’s overall performance in performance tables.

Employers

Employers are major users of the results of assessments and frequently have strong views of the quality of the results of candidates they employ or whom they have interviewed. However, one of the criteria they use to judge the quality of the assessment is the match between what is assessed and the skills needed for the job. But this is really a comment on the curriculum and a potential disagreement between what schools teach and what employers would like them to concentrate on. The examination boards can only assess what teachers teach and it is the validity of that assessment that they should be judged by.

Universities

A similar issue arises when the assessment is used to select students for courses in higher education. Up to the early part of the last century, universities provided their own entrance examinations. These had a single function: to identify those students who would benefit from a university education. In other words, the assessment was not primarily designed to certificate the achievement of the student; the important requirement was that it had predictive validity. In the middle of the last century, as part of a government initiative to reduce the number of examinations students had to face in the later years of schooling, universities gradually – and reluctantly – agreed to make use of a more general certificate whose primary function was to certificate the student's achievement in relation to a particular course. This has always been a bone of contention and there have been various attempts to try to make the assessments more relevant to universities without reducing their efficacy as certificates of achievement.

One aspect that has recently been of concern is the growing number of candidates that have been eligible for the highest available grade. There are two possible explanations for such a change over the years. Some argue that this is more likely to be “grade inflation”- the impact of very slight changes from year to year, but all in the same direction. Tim Oates, Director of Research at Cambridge Assessment has been quoted as saying that 'Giving the benefit of the doubt to pupils - consistent with the general moral sense of "access" and "best chance" which was foremost in the political agenda - can result in subtle grade inflation.'⁴

Another explanation is that candidates are indeed getting better and that a larger number deserve the highest grade. Improvements in sporting achievements are cited in support of a general improvement: athletics records are frequently broken but, because they are judged against an absolute standard such as time, height or distance, they are generally accepted without question. In fact, in many sports, improvements in equipment and training could be regarded as no different from the modifications that have served to make examinations more accessible. It may not be the case that increased accessibility makes the assessment easier.

A new attempt is being made to satisfy the needs of those who wish to differentiate amongst the highest achievers. A new grade of A* (“A star” has been introduced into the A-level examinations, designed to reward the highest achievers. However, even this is not without controversy.

For a start, the grade lies at the extreme end of the distribution. There is always concern that candidates for high grades obtain many, if not most, of their marks from questions that are at lower levels. Undoubtedly this will be the case again. The award is targeted at such a small proportion of candidates that it will be impossible to include sufficient questions for a really reliable assessment without including the candidates' performances on other, easier questions. But this does not invalidate the assessment.

Conclusion

So what is the relevance of these different perceptions of the quality of assessments for those of us who are required to provide or regulate them?

It is of paramount importance that we explain what it is reasonable to expect assessment to deliver and what it cannot. Examination bodies are not responsible for determining the curriculum. Teachers choose the examination their students will take on the basis of its match to their teaching syllabus or to the government determined curriculum they are required to follow. Awarding bodies should not accept criticism for things over which they have no control.

⁴ <http://www.dailymail.co.uk/news/article-1268199/A-levels-GCSEs-HAVE-got-easier-says-Cambridge-exams-chief.html#ixzz0rCzOhJf4>

We must also be wary of taking criticism at face value; the critics may have their own agendas.

When universities complain that the GCSE is not producing sufficient science specialists, we must recognise that there is a problem. The universities are looking at a particularly narrow range of students. GCSE science was designed to ensure that the general population was better informed about science. Most 16 year olds now leave school with a GCSE in science. In that aspect, the programme has been a success. In tackling the problem of the lack of science specialists, we must not throw out the gains made by the population at large.

When leaders of some schools rail at the quality of the assessment, we must ask whether they have a particular reason. Excluding the IGCSE from the league tables means that the results of independent schools that use it are depressed and, as they are competing with state schools for students, they do not regard this as fair. But it may also be that the IGCSE is particularly suited to the style of education that independent schools provide and for the students for whom they cater. We must monitor its introduction to ensure that they do not obtain an unfair advantage.

Employers have a right to expect that their prospective employees are educated to a satisfactory standard. In listening to their concerns, however, we must remember that not all employers have the same requirements. Schools are required to provide a general education – they are not responsible for training young people for specific employment. If the general education is not a suitable basis for the employer to undertake the necessary on-the-job training, they have a right to complain, but they may need to direct their complaints to the designers of the curriculum rather than those responsible for the assessment.

And we must heed Roger Murphy's caution about allowing people to read too much precision into the results of our assessments. When we use marks as the basis of our assessments, the difference between one grade and the next MUST be – at some point in the scale – the difference of a single mark. Some candidates - on a different day, on a different task, with a different examiner, or even when the same examiner re-reads the same scripts – WILL have got different results. What we must not allow is for the assumption to be made that the first results were therefore wrong.

Examination and assessment bodies, and the organisations charged with regulating them and ensuring quality, have a responsibility to ensure that the assessments are fit for purpose, a valid sample of the curriculum they are intended to reflect, and the processes lead to reliable results. If we achieve this, we can rebut almost all other criticisms.

Colin Robinson, June 2010

Address for correspondence: 26 Manor Park Close, Tilehurst, Reading, RG30 4PS, UK