

Eight out of ten isn't good enough: challenging teachers' perceptions of assessment

by Jacek Brant, David Lines and Adam Unwin, Institute of Education, University of London

Abstract

One of the more challenging aspects of training teachers is demonstrating that assessment is a great deal more complex than simply marking students' work. It is necessary to show that assessment takes a number of forms, that it is used for a variety of purposes and that it must be applied in a systematic and structured way in order to be effective. This paper describes the techniques used at the Institute of Education, University of London, to encourage training teachers to look at assessment in a new way.

The whole-year programme is outlined, but emphasis is placed on an introductory session held at the start of the course. Ostensibly a simple ice-breaking exercise, the game in fact demonstrates many of the pit-falls of assessment and brings to life terms which the trainees will encounter more formally later on in the course.

ABSTRACT	1
INTRODUCTION	3
BACKGROUND.....	3
THE EIFFEL TOWER ICE-BREAKER	4
THE DE-BRIEF.....	6
MARKING CRITERIA	7
ALL OR NOTHING.....	8
APPLICABILITY TO BUSINESS AND ECONOMICS	10
ASSESSMENT AS A COURSE THEME.....	11
CONCLUSION	12
REFERENCES	13

Introduction

One of the more challenging aspects of training teachers, especially at the start of their careers, is demonstrating that assessment is a great deal more complex than simply marking pupils' work. It is necessary to show that assessment takes a number of forms, that it is used for a variety of purposes and that it must be applied in a systematic and structured way in order to be effective. This paper describes the techniques used at the Institute of Education, University of London, to encourage beginning teachers of business and economics to look at assessment in a new way.

The whole-year programme is outlined, but emphasis will be placed on an introductory session held at the start of the course. Ostensibly a simple ice-breaking exercise, the activity in fact demonstrates many of the pit-falls of assessment and brings to life terms that the trainees will encounter more formally later on in the course. Its experiential nature makes it all the more powerful as a result. In addition, since it is neither language nor culturally dependent it can (and has) been used to great effect with experienced educators, both in the UK and abroad, as well as those in initial training.

Evidence shows that the exercise challenges positions and encourages participants to re-think their perceptions and values. Specifically, it cause them to move towards a deeper thinking about assessment in particular, and more generally on what it means to become a truly reflective professional.

Background

The Post Graduate Certificate in Education (PGCE) course for teachers of Business and Economics at the Institute of Education, University of London, is the largest such course in

England and Wales. Each year there are approximately sixty people (known as Beginning Teachers, or BTs) training to becoming qualified to teach in those subjects. They come from a variety of academic and working backgrounds, but it is a deliberate policy to recruit those with commercial or industrial experience, in addition to a good degree in a relevant subject¹. As a result of this procedure, the age profile of business and economics BTs is somewhat different from other courses at the Institute, with a higher proportion of mature students.

The PGCE year is intensive. Under prevailing regulations (DfEE Circular 4/98) the training is strongly school-based, yet the Higher Education (HE) provider still carries ultimate responsibility for assessing the BTs' overall performance and awarding them Newly Qualified Teacher (NQT) status. For those teaching at the Institute therefore, it is necessary not only to blend the diverse elements that make up the cohort as early and as quickly as possible, but also to introduce difficult and challenging concepts such as pupil assessment, within a whole-course framework that will dovetail theory and practice.

The Eiffel Tower ice-breaker

On the first day of the course BTs are introduced to an ice-breaker exercise that involves them constructing a self-standing tower to a height of one metre, using for its construction only two broadsheet newspapers, some sticky tape and a pair of scissors (see Figure 1). As well as using the name as a heading, the task sheet quite deliberately contains an illustration of the Eiffel Tower. The reason for this will be explained below. Students are organised into three large groups in separate rooms and then into teams of four or five. They are then asked to start. They are given no initial briefing.

¹The degrees of the BTs are divided almost exactly between economics and business studies or management. A small proportion have qualifications in other related fields, such as Law and Accounting.

Over the years, a number of features have consistently emerged at this stage:

- Planning is often minimal, with the result that increasing volumes of tape are used to prop up inherently unstable constructions;
- there will always be one team that 'fails' (this group provides an important focus, as we shall see below);
- the height requirement is often exceeded, especially by those who plan ahead. On occasions towers have stretched from floor to ceiling;
- the 'Eiffel Tower' promotes imitations², even to the extent, one year, of a team producing paper 'tourists' and their 'pets' beneath the tower;
- personal relationships are quickly forged³ and often last for the rest of the year and beyond;
- the exercise is inherently 'fun', but can be turned into a deep-seated learning experience.

After the towers are completed teams are asked to 'mark' the other examples in their own room. Again, no indications are given to the BTs as to how this should be achieved, what criteria they might or should apply, the number it should be marked out of and so on. In other words, they are asked to assess the work intuitively. In practice most teams produce criteria against which they can judge the quality of the towers, but these always extend beyond the simple one metre requirement. They also have to learn about 'moderation', though they rarely use the term, because they have to agree a single mark for the whole team. Reaching this consensus requires negotiation and at times, diplomacy.

² Cultural values do sometimes emerge, however. When using the game with a group of educators from Greece, emerging towers had a distinctly Parthenon-esq quality to them!

³ In fact this notion of teamworking is another theme in the course, one that is started at the interview stage where prospective candidates have to work together to devise a 'lesson plan'.

Having marked the work in their own room, the teams move into the other rooms to assess the towers produced there. This is important because even this small degree of 'distance' from the people who constructed the towers has an impact upon their marking; it is generally somewhat 'harder', in the sense that it results in a lower mean score. Thus one lesson about external as opposed to internal assessment can be drawn out later.

The de-brief

At this stage the teams return to their original room and are handed out a de-briefing sheet (figure 2). Each team then has to explain in written form, why they chose to mark out of ten, twenty, or something else; what criteria they applied in assessing the towers; and finally what they learnt about assessing students' work from the exercise. This period is significant, for it enables participants to reflect on the experience and examine their own tacit assumptions about assessment. An apparently fun ice-breaker activity suddenly has new meaning and purpose. Pedagogic issues become relevant with many of the reflections being drawn out and developed in the ensuing discussions. Emerging issues have included the effectiveness of group work, student centred learning, managing 'excited' children and the possibility of having 'fun' whilst engaging in meaningful learning experiences.

Following this period the plenary session imitates an external examination moderation meeting, where the chief examiner looks at the marks awarded for identical pieces of work and attempts to draw conclusions about the skills of individual markers. In this case it is conducted in a non threatening, light-hearted way. It becomes evident how wide the range can sometimes be in the marks awarded for students' work and this carries clear implications as far as consistency, pupil motivation and reinforced learning are concerned. The session also introduces some straightforward terms, such as formative and summative assessment, as well as others of a quite technical nature, such as regression to the mean.

Marking criteria

Because the BTs decide on different criteria in each of their teams, the marks can vary quite widely. It is pointed out, however, that the only criteria mentioned in the work-sheet are that the towers should be free-standing and a minimum of one metre in height. Therefore any towers that meet these should 'pass' and any that don't should fail; if marks *are* awarded they should either be zero or one hundred per cent - there should be nothing in between. Of course this fact is often recognised by some of the BTs, but they argue, with some justification, that the very fact that they are asked to mark the work *demand*s a particular response. 'The mark' is very powerful psychologically, as the BTs themselves know. The first thing any student looks for when an assignment is returned, is the mark. Thus, this part of the exercise immediately brings home the lesson that it is necessary to declare criteria in advance of any test; indeed the need to see the question and the marking guide as an integral unit to be produced together is emphasised at this point.

Often towers are marked on quite imprecise criteria such as 'appearance' and that can provide a parallel with work produced in the classroom that is word-processed as opposed to hand-written or neat as opposed to tidy. It depends upon what the assessment task is trying to achieve of course, and that is an important point to draw out, for it could be that neatness is part of the learning process to be revealed by the assessment. Clearly there is nothing wrong in that so long as it is declared in advance to those doing the work. Furthermore, when it comes to setting vague criteria such as 'appearance' some of the difficulties of criteria based assessment become apparent, in particular the need to break them down into smaller and smaller, more closely defined ones until the whole exercise becomes excessively bureaucratic (Wolf, 1995). On the other hand, judging the towers also demonstrates some of the well-known problems of normative assessment, including the one that the 'best' sets the standard

for the rest. Thus the tower with people and dogs beneath was both amusing and certainly added to the aesthetic appeal of the tower, but if that was given 100% because it went well beyond the brief, no other team could achieve the same without similarly raising the stakes. Since this has only happened only once in all the years the game has been played, a normative assessment of that year group would be assessed differently from all the rest.

All or nothing

The very idea of awarding either zero or maximum marks is explored. Invariably the team that fails to build the tower to one metre is awarded 'something': four marks out of twenty is a typical score. Generally, no reasons other than what might be termed 'humanitarian' ones are advanced for this, although sometimes it is justified on the grounds of a reward for effort, or for some of those features other than height, that figure in the teams' criteria. The reaction of members of these 'failing' teams is both interesting and instructive. Usually they recognise it for what it usually is and are grateful for the support, but sometimes they regard it as being condescending or tokenistic. This is an important lesson for the classroom. No doubt we have all generously given marks as a way of rewarding and encouraging the individual to keep up the effort, knowing that in a sense the work was 'not worth it'. However, whilst this might be justified in a formative setting, it cannot be justified in a summative one. There, such 'charity' marking can have a profound and harmful impact on grades awarded to candidates, because in that context a good candidate may have to fight hard to gain one more mark, which in the end would count the same as the 'charity' mark. This in turn would have the effect of bunching the marks, thereby making grading and differentiation more difficult.

There is a similar reluctance to award maximum marks. Again this is often justified in the feed-back session because awarding eighteen out of twenty allows "room for improvement" or, perhaps it reveals an underlying belief that no work, especially produced by a student, can

possibly be 'perfect', though what 'perfect' means in this context is never made entirely clear. This also introduces the notion of negative marking, because sometimes marks are deducted for some real or perceived failure of design, construction or aesthetic appeal in the towers. Close questioning of participants usually indicates that such negative marking is unevenly applied; if more 'positive' marks are available, it is more likely that others will be deducted. Thus, where a tower might have received twenty out of twenty, but the markers disliked the fact that it didn't resemble the Eiffel tower, they took two marks off. In contrast, a 'failure' i.e. one that did not even meet the only declared criterion of height, was more likely to *receive* ('charity') marks than have them deducted! The overall effect, once again, is a bunching around the mean.

An interesting avenue for discussion concerns the reasons for marking out of 10, 20, 100, or even A, B, C. Often there is no rationale for this, but frequently it is something that has been thought through. Ten is often regarded as offering insufficient differentiation, whilst a percentage figure offers such a range that it becomes unworkable. More sophisticated versions offer marks according to the criteria, as one might hope, and these marks are then aggregated and turned into percentages or some other denominator. It is clear, however, that in British educated people, certain 'numbers' stand for certain things. Thus, 70% and above is 'excellent', 60 - 70% good, around 50% OK (not 'average') and 40 - 50% just about acceptable. Less than 40% is regarded as poor and would probably fail. This 'signposting' is very powerful, so that teams can talk about "six out of ten" or "eight out of ten" and it is more than a quantitative measure, it is saying something qualitatively as well. For moderation purposes this is fine so long as it is made explicit.

Applicability to business and economics

Clearly whilst assessment may be discussed in a generic sense, there may be specific considerations for the teacher of business and/or economics. How different is it assessing these subjects than, for example, art or mathematics? Both parallels and differences are explored in discussion. Since the BTs often choose aesthetics as one of their criteria for judging the towers, might it be appropriate to consider coursework in a similar way. In the context of England and Wales, that is a particularly apposite question given the importance of assignments within General National Vocational Qualifications (GNVQ). Should a *well presented* GNVQ assignment attract a higher grade than one that is not as neat? At first sight the answer might appear self-evident, but since the course is designed, at least in part as a preparation for the office context where the layout and appearance of a document is often a significant factor (White Paper, 1991), then it may not be so straightforward.

Similarly, the discussion may explore the justification or otherwise of awarding a word-processed economics essay a higher mark than for a 'scruffily' hand-written one. Experienced teachers are all aware of essays that always end exactly at the bottom of a page, as if by magic, having completed the precise number that are mysteriously deemed 'acceptable' by both teacher and student, albeit with the option of greater length offering the possibility of a higher grade. As Stephen Barnes so strikingly puts it, students are rather "... like peasants paying rents to a distant but powerful landlord....measuring their academic remittance in terms of 'sides' completed" (Barnes, 1991). So, are *longer* business and economics assignments better? Do teachers have an inclination to reward the effort of writing *more*, when in fact the business skill of writing succinctly might suggest that shorter but better focused work should carry higher marks? After all there can be little doubt that well-argued and developed assignments generally display more worthy characteristics than long descriptive ones.

Success in answering business and economics questions often depend on appreciating the subjects' holistic' nature. Reference to parts of the syllabus other than the one that is obvious often enriches a response and puts it into a meaningful context. Examination boards now routinely use level of response mark schemes that reward depth as opposed to breadth. For beginning teachers it is important that an early, explicit analysis of assessment methodology is made, so that once they go into the classrooms they will possess a deeper, more reflective stance on the aims and purposes of assessment and as a result will use it more effectively to inform and evaluate their own performances as well as that of their students.

Assessment as a course theme.

The theme of assessment is revisited a number of times during the course. About half way through the first term there is a session on formative assessment, which looks at homework, question and answer and other techniques designed to reinforce learning and guide teacher effectiveness (Lambert, 1995: Black, 1998).

Later in the term summative assessment in business and economics is examined through the marking of candidate's scripts using an examination board's marking scheme. A simulated moderation meeting is convened and the beginning teachers are confronted with the reality of the impact of 'high stakes assessment' (Gipps, 1994) on student performance. This work is reinforced at the Institute of Education through sessions run centrally on assessment theory and practice.

One should not forget that the BTs are themselves part of an assessment process and so in addition to written assignments and practical observations of their teaching the final week of the course they are also required to focus on their career entry profile (CEP) and their

personal development folder (PDF). Both provide an opportunity for them to reflect on their PGCE year and to write more formally about their experiences of assessment in relation to the standards as outlined in DfEE Circular 4/98. By then the experiences of the tower game may seem little more than a distant memory, but one that is generally quite firmly embedded and fundamental to their development as teachers of economics and business.

Conclusion

Experienced teachers assess their students virtually non-stop (Shon 1983, 1987; Black, 1998). Often this is an instinctive response, but it also recognises that a range of methods can be applied, from the simple question and answer session in class to the kind of 'high stakes' assessment that appears to determine life chances. Especially in a system where summative assessment is so strongly embedded as it is in England and Wales (Eckstein and Noah, 1993) but in a wider context as well, it is essential that people who are training to become teachers recognise that assessment takes a variety of forms, that it offers positive benefits as well as potential risks, but that above all it needs to be thought through and carefully considered if it is to be truly effective and educationally beneficial.

References

- Barnes, S. (1991) *A is for Alienation, Art and A level*, Paper presented at Langrish House, Hampshire, 5 October.
- Black, P. (1998) *Testing: Friend or Foe? Theory and Practice of Assessment and Testing* London: Falmer Press
- Eckstein, M and Noah, H (1993) *Secondary schools examinations : international perspectives on policies and practice*; New Haven: Yale University Press
- Gipps, C. (1994) *Beyond Testing*, London: Falmer Press.
- DfEE (Department for Education and Employment) Circular 4/98
- Lambert, D (1995) "Assessment and improving the quality of pupils' work" in Capel, S., Leask, M., and Turner, T *Learning to Teach in the Secondary School* London: Routledge
- Shön, D., (1983) *The Reflective Practitioner: How Professionals Think in Action*, New York: Basic Books
- Shön, D., (1987) *Educating the Reflective Practitioner: Towards a New Design for Teaching and Learning in the Professions*, San Francisco: Jossey-Bass.
- White Paper (1991) *Education and Training for the 21st Century*, London: HMSO.
- Wolf, A. (1995) *Competence-based Assessment*, Buckingham: Open University Press.