Are we all agreed it's a 2.i? Making assessment criteria explicit

Practitioner Research
in Higher Education

Vol 2 (1): page 36-42

Darrin Beattie, Sue Gill, Kelley Wallace and Jim Wood
Newcastle University
Northumbria University
sue.gill@ncl.ac.uk

Abstract

Using our experience of a work-related learning project, run jointly between Newcastle and Northumbria Universities, in which 600 students annually undertake work placements, this paper focuses on the inherent challenges of using over 30 assessors to assess students. The assessors come from a wide variety of backgrounds and we have had to find solutions to the issue of consistency of making across all assessors. Inevitably this has led to a process designed to make explicit the tacit assumptions which underlie our individual experiences of assessment and forge a common, clearly articulated understanding across all the assessors. We also recognise that this is an ongoing activity which needs continual attention to maintain and develop the community of practice and adapt to the changing environment in which we operate.

Keywords
Assessment; placement; employability; authentic assessment; innovative; transparent.

Introduction

This paper is based on the results of a three-year project to develop an effective assessment model for employability modules. In particular, we consider how the twin challenges of moving to the use of over 30 assessors, working in pairs; and rethinking the assessment model were addressed. The majority of the marks are based on an assessed interview in which students articulate and evaluate evidence of their competence against our graduate skills framework.

Background

Newcastle University and Northumbria University have been collaborating since 1993 on a work-related learning project that involves students gaining credit for taking modules in which they work for the benefit of others in the local community. This involves semester-long placements in local schools, colleges and community learning centres, as well as volunteering in the University Union Society and within the community. Subject to permission from their course, students may use their paid term-time work as a context for their learning.

Six hundred students from diverse degree programmes take the modules each year, and come with different experiences and expectations of 'assessment'. It has been challenging to find a method for assessing students that is authentic, aligned to learning outcomes, drives the students' learning and performance throughout their placement, and is cost-effective.

> *Assessments that help students to identify and then present their achievements effectively are invaluable... However, assessment is partly a matter of pragmatics (how much assessment it is reasonable to impose on learners and teachers)... Teachers may still innovate in individual modules, while being prepared for objections from students who prefer the familiarity of established methods and are suspicious of new ones.*
>
> (Yorke and Knight, 2006)

The aim of the modules is:

> *… to develop students who can independently self-manage, proactively interact and ethically apply their knowledge and skills in a work-related context.*

This reflects the Newcastle University employability statement and is a significant aspect of the 2007–2012 University Learning and Teaching Strategy. Students attend a compulsory induction workshop at the start of the semester, which introduces them to the module. During this workshop they are given a handbook containing the assessment criteria and the set of possible questions for the assessed interview. Using principles of experiential learning (e.g. Kolb, 1984), the students are asked to maintain a reflective diary evidencing the skills they have demonstrated and developed which will form the basis of the evidence to be articulated at the assessed interview. The learning log is not formally assessed, but is used formatively through a dialogue between the student and the module leader. During the semester, students are expected to manage their own learning. They are strongly encouraged to attend workshops, a personal tutorial and a seminar, arranged at critical points, in order to develop their ability to learn reflectively and to help them prepare effectively for their assessed interview.

The assessment model has evolved over time and we moved to an assessed interview at the end of the module in academic year 2006/2007. Although called an interview, the assessment takes the form of a 20-minute session where the student is asked a set of seven questions which cover the learning outcomes of the module and which are designed to guide the student learning while they are on the module. For each main area there are a number of questions and the student does not know which one will be used in their particular assessment. The assessment is audio recorded, and this recording acts as the exam script and is available for the external examiner. The interview has a rigid structure and does not currently allow scope for the assessors to explore specific issues raised by the student. We believe it offers a reasonable balance between the rigour of an examination framework while offering an authentic assessment experience.

Assessment is against a set of criteria drawn from the learning outcomes with carefully defined standards for each criterion which provide useable guidance for both student and assessor.

In each interview there are two assessors, a 1st and a 2nd marker – all 1st markers (6 from Newcastle University and 1 from Northumbria University) work as part of the same unit, are module leaders and may be known to the students through this role. 2nd markers are recruited from across the two universities (see Box 1).

Box 1 Recruitment of assessors

2nd markers are drawn from:
Non-academic staff (includes academic-related):
- Careers Service
- QuILT
- Education Liaison
- Student Recruitment (Newcastle & Northumbria)
- Student Progress

Non-academic staff are reached through department heads through an expression of interest supported by their line manager (cascaded to all staff with tacit approval for participation)

Paid postgraduates
Postgraduates are recruited by informing all postgraduate students at both institutions. They go through a formal application and selection process – CV and cover letter, interview and testing of their experience and ability to apply assessment criteria.

Academic staff
Academics are encouraged to participate through personal contact/invitation.

Integrating assessors

All assessors go through the same training process where our aim is to achieve consistency of interpretation of the assessment criteria and standards, build confidence and encourage full participation in the assessment process. New markers have a two-hour introductory training session, which includes the practical and operational aspects such as collecting written evidence on the pro-forma assessment document. The main focus is on practising making assessment judgments and discussing the process with at least one experienced assessor. For experienced markers there is the option of attending one of the assessment seminars with the students, or a one-hour recap session. All markers have access to the student training and support materials and resources through our respective virtual learning environments (VLEs) (Blackboard).

One important purpose of the training for all assessors is to ensure all markers are aware of their role and the range of behaviours we expect (student to marker, marker to marker and marker to student). Possibly the most important issue is to get across our insistence on criterion-based assessment and the training provides the opportunity for markers to really get to grips with how seriously this is taken and have the opportunity to use the criteria in practice. This is further developed in the assessment where the 1st marker has the responsibility for mentoring less-experienced 2nd markers. Listening to the deliberations of markers, they come back time and again to the criteria – 'Let's go back to the criteria again' or 'How does this fit with the criteria?' Markers who are unreliable are not used again (a rare occurrence). 2nd markers are also provided with access to an adjudicator, external to the 1st markers, if they are unhappy with the assessment decisions being reached, although to date this has not been used.

In general, and subject to the demands of their roles, markers are retained from semester to semester and year to year. For those staff who act as 2nd markers, who are not paid and have to make time for this activity in their busy schedules, taking part provides:
• a face-to-face link to students
• the opportunity to gain a better understanding of the employability of students,
• the chance to participate in an interesting activity
• the chance to join a community with shared interests.

Theory

Throughout the development of these modules we have aimed to produce assessment which was the best possible for developing student understanding and learning, drawing on the wealth of literature. Our move to an interview was a response in part to our frustration with students who produced pedestrian portfolios but who had, when you spoke to them, insightful and perceptive reflections on their development in the placement. Bringing together a community of markers was a pragmatic response to the need to assess around 350 interviews each semester rather than a deliberate attempt to model a specific theoretical standpoint. In particular it made necessary the articulation of many aspects of the assessment process which were previously known only by members of the careers service curriculum unit and forced us to refine and review our understandings in order to share them with the new assessors. Despite this pragmatic approach, the theoretical models do throw light onto the factors behind some of our success and offer guidance as we consider future changes.

The challenge of a new assessment technique meant the team were already primed to review their practice and remake it in line with good practice (e.g. Gibbs and Simpson, 2004). In particular they were keen to embed the concept of assessment for learning rather than assessment of learning (see Northumbria University Assessment for Learning CETL).

We must recognise that the individual staff who act as assessors bring their own specific professional context (Trowler, 2008) to the process of marking students and unless opportunities are provided it is too easy to stay within that context, rather than testing views against the developing team paradigm. For example, within the team of assessors, careers advisors had a different view of what comprised and defined an assessed interview compared with an educational developer. Additionally we had to recognise the team paradigm was (and is) not fixed, rather it is continually developing through the testing of staff and university processes and has evolved as the project evolves.

Trowler suggests that groups that are able to see themselves through looking at their regimes are better able to look inward and outward and see good practices within their relevant contexts, making it easier to reflect on themselves and their practices and change. One of the outcomes of changing the assessment has been to increase the ability and frequency of group reflection on processes and practices.

One underlying theory, which has relevance, is that of 'communities of practice' (Wenger, 1998). These are said to grow out of shared practice where we learn from ourselves and each other and become part of the community. By providing the social and academic norms and the standards against which we judge the work, the community provides moderation for its members as individuals and aids consistency and fairness across all work assessed. It is the community which accepts or rejects the individual judgements we have made when assessing (Morgan, 2004).

How we learn to be an assessor

In work undertaken at Oxford Brookes University, students were encouraged to attend an intervention, very similar to the assessment workshop we provide, where they undertook marking exercises based on the types of material they would have to submit themselves (Rust et al., 2003). These processes of 'socialisation' help transfer tacit knowledge about assessment and students who attended the workshops were found to obtain a mark 6% higher than non-participants. In our work we use this model with both students and assessors. The process of socialisation (discussion, exemplars and peer review) is vital to help 2nd markers gain familiarity and confidence with the assessment practices and with our existing tacit assumptions around assessment.

Tacit knowledge is generally acquired from personal experience, it is associated with its use in that situation and it is valuable to the person. Thus tacit knowledge can be defined as knowing how rather than knowing why (Smith, 2003). Through undertaking the same exercises as the students and discussing their marks in small groups and a plenary session, new markers begin to accumulate that experience. Through this process a shared understanding is developed across the markers. This understanding is reinforced during the ongoing discussions 1st and 2nd markers have when considering a student's performance in the oral examination. These conversations allow markers to reference against each other and help bond new markers into the community of markers through space for discussion and elaboration of specific points. This process could be characterised as intended to develop meaningful knowledge (Price, 2006) – that combination of tacit and explicit.

For assessors to assess, they need to have a sense of the quality which is appropriate to the task and be able to measure the student work against that concept (Sadler, 1989). Sadler has recently suggested that two factors are uppermost when judging student work, marker expectations and the performance in relation to other students – norm referencing (Sadler, 2005). The training provided to assessors and the continuing dialogue around the assessment of individuals should help to mitigate these two factors. By marking together, the markers can check their expectations against each other and challenge their assumptions; and by having a 2nd marker who is unconnected with the student they can act as a check to prevent the 1st marker allowing their knowledge of the student within the cohort to colour the mark they award.

In contrast to some of the conversations around marks recorded by Orr (Orr, 2007) our assessors are encouraged and expected to discuss the students' performance in terms of the assessment criteria and the ways in which the assessors noted agreement with the various standards within the criteria. Only after this discussion has taken place is it usual for details of the personal knowledge of the individual to be raised, e.g. in relation to any personal circumstances. Overall compliance with the criteria is followed by a more detailed breakdown of the individual elements to ensure recall of as much evidence as possible and consider how the evidence fits the standards. Like the academics recorded by Orr, when coming to a mark, it was usual to agree a grade (2.1 or fail), then, if this was agreed, move to look at the criteria in more detail to agree a specific numerical mark; finally the assessors consider what feedback would be appropriate to the student in the light of their assessment. In all cases, the feedback is drawn from the assessment criteria and in line with best practice guidelines within the two host universities.

Notions of apprenticeship and connoisseurship

The processes used to induct markers could be seen as an **apprenticeship** system leading to a **connoisseurship** model of assessment – only after training are you able to mark, paired with an experienced colleague. This process of acclimatisation to the norms of the group can be defined as connoisseurship (Morgan, 2004). Ecclestone, quoted in Morgan, cautions against the belief that the expert implicitly knows the mark:

> *Experts become more intuitive and less deliberate, and less able to articulate the tacit knowledge on which much of their decision-making has come to depend.*

The processes that encourage the discussion of marking within the team work to militate against the development of 'experts'. By encouraging and expecting discussion, those colleagues who are working to improve assessment practices are continually being held to account by the wider assessor community to describe and make explicit the continuing changes in the assessment model. This process helps define the shared values and assumptions of the assessment process and the employability content and the preferred ways of working to assess students' work. That this takes place face-to-face is important, as it provides a very personal support for 2nd markers and immediate response to questions, overcoming some of the barriers to assessment reported in the literature (Black and Wiliam, 1998). This raises another challenge for the continuation of the assessment using many markers – the more familiar we become with marking our students, the less able we are to articulate the criteria and processes we use to new markers.

The training process for students and assessors draws on Bandura's **social learning theory** (Bandura, 1977) with its emphasis on the importance of the social context for learning.

Discussion

While this paper has attempted to make clear the various interventions and activities we undertake to make the assessment criteria explicit, there are a number of ongoing challenges for the team.

Firstly is the challenge of maintaining focus on the assessment criteria and avoiding the growth of group think (Janis, 1972), where maintaining agreement between markers becomes more important than critically examining why and how we assess. Our work to ensure we have a shared understanding of the assessment processes and criteria is an absorbing activity, which develops a deep understanding in the majority of markers and a passion for the application and quality of the process.

The relatively stable pool of 1st markers coupled with the fact that the majority of them are located in one unit can work to create a situation which encourages the development of group think. This cohesion may also be encouraged by the need to defend the modules from challenges that they are in some way less rigorous than traditionally (i.e. unseen written examinations) assessed modules.

Overlapping with this cohesion, some 1st and 2nd markers used to be part of the same unit, creating another set of shared expectations and experiences. So while the markers may not all inhabit the same physical location they do inhabit a series of overlapping spaces through work and social familiarity.

Many 2nd markers stay with the project, some withdraw and some are found to be unsuitable. In the very few cases of the latter we must have clear, rational explanations to explain why we had specific issues with individuals. We need to be aware that we may be influenced in our judgements where people do not fit easily into the group norms and that this may encourage us to find fault.

The potential for group think is mitigated against by the wide range of backgrounds from which 2nd markers are drawn, with groups coming from the Careers Service (Newcastle University), Student Recruitment (both Universities), QuILT (Newcastle University) and academics; and the robust personalities of the individuals.

The work of Shay argues that the:

> *… assessment of complex tasks is a socially situated interpretive act…*
>
> (Shay, 2004, quoted in Shay, 2005)

That is, the individual assessors bring their own experiences mediated through their involvement in specific communities of practice. This means that our training processes have to allow space for new markers to reflect on their potentially varied experiences and become receptive to the practices of the group. Making the space and time for this is challenging for all of us and we may need to evaluate more systematically how our induction process for markers works.

Additionally we need to be sensitive to the **past assessment experience** of new markers. For example, within both our institutions it is difficult to choose words to describe the assessment vehicle which are not value-laden and open to a wide range of sometimes subtle, sometime less subtle, interpretations. Is it an assessed interview, an oral examination, a viva, or do we end up with some over-lengthy but politically very correct phrase which gets shortened anyway?

One particular consequence of the diversity of backgrounds has been the amount of time spent discussing whether or not we should **prompt** students who were struggling in the interview and how this would be reflected in the mark given. There were sharply differing assumptions made about the interview process, thus a member of careers staff viewed interviews as a very specific way of allowing the interviewee to provide information, while an educational developer specialising in assessment viewed them (in this case) as a much less flexible assessment vehicle. While we have reached a consensus not to prompt, this is still a topic of discussion and debate within the team and is raised anew every time a new marker joins the team. This is a good example of the benefits of constantly being re-challenged by new markers joining the team and, providing we keep discussing the topic, one key way in which we constantly keep our norms under review.

The link between our universities also opens possibilities for **fragmentation** of the group due to the challenges of operating across two different institutional and operational cultures. Additionally, as Newcastle is the 'organising' university and one where the majority of 1st markers are based, it may be possible that this negatively impacts on the integration of 2nd markers into the pool of assessors. We may need to consider how often we mix a 2nd marker from Northumbria with a module leader from Newcastle? Can we honestly say that the distribution is due wholly to when markers were available, or from a wish to put markers with people they know? May we need to ensure markers work across the two universities as much as possible to generate the shared experiences and expectations we say we value?

Within our institutions there is an increasing emphasis on **anonymous marking,** which is difficult, or impossible using assessed interviews. We try to take sensible steps to mitigate the potential impact of an assessor knowing a student. For example, we specify that the assessors must identify if they know the student in any way (except in their role as module leader, usually the 1st marker) at the start of the recording of the examination. Secondly, the assessment discussion between the markers after the student has left the room has a protocol where the 1st marker always defers the initial assessment summary to the 2nd marker, then adds their comments to the summary provided by the 2nd marker.

It is also customary that the 1st marker does not divulge any information about the student (apart from details of the placement) until after the assessment discussion is over. For example, in one instance where in the assessment the student was fluent and poised, the 1st marker only revealed how hesitant and lacking in confidence the student had previously seemed after the mark had been agreed by both. Through formatting the discussion in this way, we try to minimise the impact of any expectations the 1st marker may have of the student and to see the student against the assessment criteria rather than our preconceptions.

For this and other aspects of the assessment to work we must have robust 2nd markers who deliberately **avoid over-compliance** with 1st markers. In training they are reminded that their role is to provide an almost 'external' view of the student and they are encouraged to stand up for their points of view and are reassured there are mechanisms if the two markers cannot agree a mark via a third party. To date this has arisen very few times. One instance was where a student performance was particularly poor and the markers sought a wider discussion to place the student mark at the appropriate point in the 'fail' category. This discussion provided the team with the impetus to reconsider the criteria for a fail mark to distinguish across the full range of the 40 marks. Through discussions during training it is emphasised to 2nd markers that they are not there to 'rubber stamp' the decisions of the 1st marker but to provide their own take on the performance of the student against the criteria. This is not to say that we don't have disagreements between markers over an individual student but these are resolved through discussion and listening to the recording again, if needed, and always referring back to the criteria and how we can map the student performance against that criteria.

Acknowledgements
The authors acknowledge the receipt of two Newcastle University Innovation Fund awards which have helped with the development of aspects of this work, and the support of the CETL Assessment for Learning, University of Northumbria.

References

(all websites accessed 19.12.08)
Bandura, A. (1977) *Social Learning Theory.* Prentice Hall, New Jersey
Black, P. and Wiliam, D. (1998) *Inside the black box: raising standards through classroom assessment.* Kings College, London

Gibbs, G. and Simpson, C. (2004) Conditions under which assessment supports students' learning. *Learning and Teaching in Higher Education* 1: 3–31. http://resources.glos.ac.uk/tli/lets/journals/lathe/issue1/index.cfm
Kolb, D.A. (1984) *Experiential Learning Experience as a Source of Learning and Development.* New Jersey: Prentice Hall.
Janis, I. L. (1972) *Victims of Groupthink.* Boston: Houghton Mifflin.
Morgan, D (2004) From Competence to Connoisseurship: Challenging Assessment Processes in Post-compulsory Education. Research Brief Series 6, Research Centre for the Learning Society, University of Exeter
Northumbria University AfL CETL (2008) Assessment for Learning web page. http://northumbria.ac.uk/cetl_afl/
Orr, S. (2007) Assessment moderation: constructing the marks and constructing the students. *Assessment & Evaluation in Higher Education* 32(6): 645–656.
Price, M. (2006) I know a 2i when I see it. *Challenging Perspectives on Assessment.* http://stadium.open.ac.uk/stadia/preview.php?s=39&whichevent=736
Rust, C., Price, M. and O'Donovan, B. (2003) Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education* 28(2): 147–164.
Sadler, D. R. (1989) Formative assessment in the design of instructional systems, *Instructional Science,* 19:119-144.
Sadler, D. R. (2005) Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education.* 30(2): 175–194.
Shay, S. (2004) The assessment of complex performance: a socially situated interpretive act *Harvard Educational Review,* 74(3), 307-329.
Shay, S. (2005) The assessment of complex tasks: a double reading. *Studies in Higher Education* 30(6): 633–679.
Smith, M. K. (2003) *Michael Polanyi and tacit knowledge.* www.infed.org/thinkers/polanyi.htm
Trowler, P. (2008) Beyond epistemological essentialism: academic tribes in the 21st century, in Kreber, C. (ed) *The University and its Disciplines: Teaching and learning within and beyond disciplinary boundaries.* London: Routledge.
Wenger, E. (1998) *Communities of practice: Learning, meaning, and identity,* Cambridge University Press
Yorke, M. and Knight, P. (2006) *Judging and communicating achievements.* York: Higher Education Academy www.heacademy.ac.uk/assets/York/documents/ourwork/tla/employability/id459_employability_%20judging_and_communicating_achievements_337.pdf