

O'Donovan, B, Price, M and Rust, C (2001)

The student experience of criterion-referenced assessment (through the introduction of a common criteria assessment grid) .

O'Donovan, B, Price, M and Rust, C (2001) The student experience of criterion-referenced assessment (through the introduction of a common criteria assessment grid) . *Innovations in Education and Teaching International*, 38 (1). pp. 74-85. ISSN 1470-3297

doi: 10.1080/147032901300002873

This version is available: <http://radar.brookes.ac.uk/radar/items/8eda5cc4-c966-3fa3-112e-b60c52da767b/1/>

Available in the RADAR: October 2009

Copyright © and Moral Rights are retained by the author(s) and/ or other copyright owners. A copy can be downloaded for personal non-commercial research or study, without prior permission or charge. This item cannot be reproduced or quoted extensively from without first obtaining permission in writing from the copyright holder(s). The content must not be changed in any way or sold commercially in any format or medium without the formal permission of the copyright holders.

This document is the preprint journal article. Some differences between the published version and this version may remain and you are advised to consult the published version if you wish to cite from it.

THE STUDENT EXPERIENCE OF CRITERION-REFERENCED ASSESSMENT (THROUGH THE INTRODUCTION OF A COMMON CRITERIA ASSESSMENT GRID)

Berry O'Donovan, Margaret Price and Chris Rust

Oxford Brookes University

INTRODUCTION

This paper constitutes part of an ongoing research project into criterion-referenced assessment tools and processes. It is a project undertaken in the pursuit of a conceptually sound and functional assessment framework that would promote and encourage common standards of assessment. Commenced against a background of growing national concern in the UK about marking reliability, standards and calls for public accountability (Laming, 1990; Newstead and Dennis, 1994), a common criteria assessment grid was developed and piloted, and its effectiveness evaluated from the perspective of academic staff. These opening stages of the research project are summarised below (see Background), and have already been reported elsewhere (Price and Rust, 1999). The third stage, and subject of this paper, was to investigate the student perspective and to evaluate their experience of the use and effectiveness of criterion-referenced assessment and, in particular, the criterion assessment grid, and to identify the best ways of taking this project forward.

CONTEXT

The initial impetus to address the issues in this project came from an external examiner for the Business Studies programme at Oxford Brookes University who was a strong proponent of criterion-referenced assessment as a means of ensuring consistent standards between markers. Another external examiner was concerned to ensure common standards between modules. These concerns and recommendations were expressed at a time when, across the sector, a division of opinion on assessment was prevalent. The established normative, connoisseur model of assessment (“I could not describe it, but I know a good piece of work when I see it,” most often likened to the skills of wine-tasting or tea-blending) had been challenged by the emergence of the use of explicit assessment criteria and performance standards.

At a national level within the UK compelling pressure has been applied to higher education institutions to maintain high academic standards (Lucas and Webster, 1998). This pressure has been exacerbated over the last few years by an apparent fall in standards suggested by the rise from 25% to 50% in the proportion of good degree results (upper-second-class and first-class degrees). An increase in good degree results that has understandably met with public scepticism when viewed in the context of a rapid expansion of student numbers and a drastic cut in the unit of resource in UK higher education. In response to this, the Quality Assurance Agency (QAA) embarked on a new quality assurance system reliant on the establishment of explicit degree standards (so called ‘threshold’ standards or ‘benchmarking’). This move could be seen to be a logical development arising out of the national discussion on generic level descriptors (Otter, 1992; Greatorex, 1994; Moon, 1995; HEQC,

1996) promoting their use in higher education as a means of establishing common standards. These common standards can be particularly difficult to realise within a broad-ranging, multidisciplinary and discursive subject such as business and management in which many open form assessments, such as essays and reports, are set on topics which are integrative in nature.

It was in this context that it was decided to develop a programme-wide set of criteria 'referenced' against standards for use across the whole Business Studies programme at Oxford Brookes University.

BACKGROUND - THE INITIAL STAGES OF THE PROJECT

A common criteria assessment grid was developed within and for the Business School and first piloted in the academic year 97-98. The grid has 35 criteria plotted in matrix format against grades resulting in 'grade definitions' detailing acceptable performance for each criterion at each grade (one page of the grid detailing six criteria is reproduced in Figure 1). Staff select appropriate criteria for any given assessment to create a 'mini-grid' (see Figure 2 for an example). The main intention was to develop a comprehensive marking criteria grid to help establish common standards of marking and grading for Advanced Level undergraduate modules (those normally taken by second and third year students) across the Business programme enabling consistency in marking and easier moderation. Furthermore, it was hoped that the grid would have the additional benefits of providing more explicit guidance

to students (resulting subsequently in better work), and making it easier to give effective feedback to the students.

FIGURE 1 about here

FIGURE 2 about here

The results of this pilot were initially evaluated through the views of the staff who used the grid, gained from a combination of questionnaires and one-to-one interviews, and have been published (Price and Rust, 1999). The only student input into this initial evaluation was through the perceptions of the staff questioned.

The main conclusion of this initial paper was that, at least in its present form and usage, the grid had failed to precisely define a common standard for advanced level undergraduate modules. This was evidenced by the fact that different tutors had used the assessment grid (and consequently exactly the same grade definitions) for a basic module (one normally taken by first year students) and a Masters level module apparently without any difficulty. However, the paper further concluded that the findings had demonstrated that the use of such a grid could provide other real

benefits. It could help to raise the quality of marking through greater consistency in marking both for a team of markers and for an individual marker, but this was more likely to be the case if the tutors had discussed the grid together before using it. It could also help provide, from the tutor perspective, more explicit guidance to students and thus potentially improve the quality of their work, although it appeared that this was only likely to be true for the most motivated students unless time was spent by tutors discussing with students the meaning of the criteria terms and grade definitions. Using the grid could also raise the quality of feedback to students by making it easier to give and helping to focus the marker's comments.

The initial mixed findings reflected many of the issues associated with criterion-referencing in the marking of more qualitative and open form assessment. Whilst many would agree that criterion-referenced assessment appeals to our notion of equity and fairness, it is not without its pitfalls, not least of which is the potential for multiple interpretations of each criterion and descriptor by both individual staff members (Webster et al., 2000) and students.

Consequently, the next stage of the research was to investigate the student experience of criterion-referenced assessment and, in particular, the grid. It is these findings that this paper summarises and discusses.

METHODOLOGY

Within the context of multiple possible interpretations of assessment criteria and standards, our research aimed at revealing a rich understanding of the student experience of criterion-referenced assessment that would attempt to capture and convey a student's perception and understanding of the experience. A two-step process was employed - an initial broad sweep survey by questionnaire to check and identify the main issues, which were then used to structure conversational style 'prompts' given to a number of focus groups to gather more in-depth qualitative data. The questionnaire was initially sent electronically to students on two core 2nd year final term modules, but this elicited only three responses. It was then administered at the end of one week's seminars on one of those modules, and out of just over 100 given out, 89 were returned.

Perhaps predictably, the 92 completed questionnaires (89 paper, 3 electronic) did not provide the rich data we were seeking. In particular, the open questions asking for explanations had gleaned few responses. A number of issues relating to the use of the grid and its application were identified but most had already been anticipated from the previous study, nevertheless the data provided useful confirmation. One particular issue the survey did highlight was that a number of students (16, 24% of those that had used the grid) had apparently found difficulty using the grid.

Following the questionnaire survey three focus groups were conducted by an independent facilitator. There was a poor response to the request on the survey questionnaire for volunteers to join focus groups (one!) so direct (but not targeted) requests were made to groups of students on modules known to be using the grid.

The first two groups had seven participants and the third had four. The proceedings of each group were audio-taped and the tapes subsequently transcribed.

Data for analysis took the form, therefore, of questionnaire information and focus group tapes and transcripts. Interpretation and analysis of the meaning of the richer and more qualitative data from the focus groups was undertaken through the independent identification by each of the researchers of the themes and debates arising out of the alignment and grouping of individual elements from the textual transcripts. The objective was to attain a pattern of responses representing the subjects' sense of criterion-reference assessment through a 'sufficiency' of focus group data. Such sufficiency was gained through an interpretivist approach which simply required enough data for a rationality to emerge (Spender, 1989). The validity of interpretation was strengthened firstly, when a coherent rationality was considered to have surfaced when researchers pooled their individual interpretations and, secondly, when much of the findings confirmed prior published research.

The findings are reported below, grouped under the following headings:

1. usefulness as a guide and in improving the quality of work
2. usefulness in improving feedback
3. usefulness to improve standards and marking consistency
4. student suggestions for improvement

This is followed by a discussion on the overarching, fundamental and interesting themes concerning criterion-referenced assessment that also became apparent and consideration of the next steps to be taken in this research project.

FINDINGS

1. USEFULNESS AS A GUIDE AND IN IMPROVING THE QUALITY OF WORK

From the questionnaire, 68 (73%) of the 92 respondents said they had made use of the grid, and of these 35 (52%) had used it in ‘planning’ and/or ‘whilst doing’ and/or in ‘reviewing’ their work prior to submission.

From the focus groups, the general consensus was that the grid could be helpful in clarifying what a piece of work required

“It shows students what they should include...in their assignments and the relative weightings of the different components” (1E)

and could be especially helpful when it guided students to what they considered obscure and non-obvious and which they otherwise might have ignored. The proposition that students have a right to be informed and guided on the criteria and standards on which their work will be assessed is a compelling one (Brooker, et. al.,

1998). It is also difficult to argue with Sadler's point that without an understanding of criteria and standards students' efforts "*are likely to contain elements of random trial and error*" (1987, p. 196).

Over and beyond clarification, at least three different strategies for using the grid were identified in the focus groups

"refer to it straight from the beginning...base and structure it from the start" (3C)

"other way round, ..almost complete and then go and check the grid" (3D)

"I probably plan it and before I went to write it I would go and look at the assessment grid. Then when I think I've finished it, I would go back and use the grid again"
(3E)

There were a variety of criticisms of the grid and its usefulness, and reasons offered as to why the grid was not used more as a guide.

Lack of detail

A major criticism from most of the focus group members was the need for greater clarification of what is meant by many of the terms and phrases used in the grid.

“[the grid] *needs clarification, I mean ‘address them comprehensively’ what do you mean?*” (1E)

“*There is very little difference between a ‘C’ and an ‘A’. It is very hard to distinguish the differences*” (2G)

“*It is too broad and too vague*” (2E)

These charges of vagueness and imprecision echo one of the principal arguments against criterion-referencing. Criterion-referencing may be an ideal solution for simply defined competencies (e.g, ‘can swim 50 metres’) but is clearly problematic as assessment becomes more complex and grade descriptors are also used (Gipps, 1994). However, producing ever more numerous, detailed and complex criteria has proven to make their assessment unmanageable. “*A careful path needs to be found between the extremes of vague and nebulous criteria on one hand, and a proliferation of detailed and trivial objectives on the other*” (Shorrocks et al., 1992 in Gipps 1994, p. 93). Furthermore, these criticisms encompass two different viewpoints, ranging between the desire for better and clearer guidelines and the desire for more specific clarification of content. The distinction between making assessment *criteria and standards* explicit to students and precise guidance on *content* requires careful explanation. In the course of the discussion, some students demonstrated, if sometimes a little reluctantly, an understanding that the latter may not always be appropriate at university level.

“There’s a difference between making things clear and taking out some of these words that are very subjective, and telling you what to put which would obviously not be acceptable” (2A)

Subjectivity

Linked to the criticism of the lack of detail and precision was a further strongly voiced concern that interpretation of the wording and its meaning is always going to be subjective and open to multiple interpretations between individual students and staff. The wide-ranging different interpretations of staff members previously highlighted by the research undertaken by Webster et. al (2000) is compounded for students as relatively novice participants within the context of academic standards and criteria.

“as in comprehensive and imaginative you know, that’s all very subjective...” (1B)

“[it’s] open to interpretation what some might perceive as imaginative, other people may not” (3C)

Familiarity/negative experiences

Due to problems of subjectivity of individual interpretation (both staff and student), imprecision and lack of clarity of grade definitions and criteria, it was apparent that,

for many students, expectations of the grid had, over time, become disillusioned and frustrated.

“I think the reaction does get sort of more blasÈ and not bothering to look at it”
(1C)

“The first time...I was quite excited...if I do X, Y and Z I’ll get a good mark but when you do use it and the feedback doesn’t quite match, when you’ve had experience of it not working particularly well the second time you’re a bit numb to it and you think well I tried to do it this way last time and I didn’t do any better than I have done previously anyway so you use it either less or not at all.” (1G)

However at least one member of the focus groups had the opposite point of view,

“In my second year I didn’t actually look at these too much, but for the coursework so far [given in the third year] I have used the grids” (3C)

Implicit in their comments is that positive experiences from use of the grid is likely to encourage continued use whilst negative experiences can dash expectations. As Brooker et al (1998) identify, the importance of providing students with explicit and useful information on assessment in advance cannot be underestimated in terms of motivation.

Need for staff to explain/discuss the grid

Another common suggestion in the focus groups for why the grid was not used by students was that many staff appeared to disregard it. This clearly has implications that again link back to the questions of subjectivity and lack of detail. If more staff had introduced, discussed and explained the grid one would hope that a greater degree of (shared) understanding might have been achieved.

“I think the problem is sticking them [the grids] at the back of the course book because they get forgotten about” (3D)

“ It’s pointless giving someone something and then not explaining it to them” (3E)

“If it hasn’t been explained to you then I think if you get more and more you just think of it as another piece of paper” (1C)

Shared understanding, is more often achieved through a combination of tacit and explicit modes of knowledge transfer and can include such processes as observation, exemplars, imitation and practice (Nonaka, 1991). Sadler argues that owing to the inherent imprecision of verbal description that the way to specify standards is through a combination of verbal descriptors and exemplars. *“This dual approach has the potential not only for achieving comparability among schools, but also for helping students to acquire evaluative expertise themselves” (1987, p. 207).*

2. USEFULNESS IN IMPROVING FEEDBACK

The second direct benefit to students anticipated in the use of the grid was that it would be useful in improving the feedback they received on work once it had been assessed. Sadler suggests “*one of the conditions necessary for the intelligent use of feedback is that learners know not only their own levels of performance but also the level or standard aspired to or expected.* (1987 p.196). This could work either through staff making direct reference to the grid definitions in their comments, and/or possibly through using the grid itself as a feedback form with relevant grade definitions ticked or highlighted, or through students using the grid themselves to make sense of their mark and the tutor’s comments.

With regard to the first two of these possibilities, the focus groups were clear that used by itself the grid was not sufficient as a feedback tool.

“It may be very quick and easy for a lecturer to have a highlight pen and just highlight these boxes but ... it doesn’t tell me anything at all; it tells me no more than when I read through it initially” (1D)

This obviously links to the criticism of the general nature of the grid definitions. The fact that they were seen by some as unclear and vague in providing guidance means that they could be similarly seen as unclear and vague in providing feedback, if that was all that was provided.

The groups’ conversations around feedback also revealed that their prime concern was wanting feedback on the specific task undertaken rather than at the meta-level of the criteria used to assess the task provided by the grid.

There was widespread agreement in the focus groups that the grids could be useful in helping students make sense of their grades and the tutor's comment, and in the survey response, 22 (32%) of those using the grid said they had used it to help understand marking feedback. In particular, but not exclusively, data from the focus groups indicated that this was especially likely to be the case if the student had received a disappointing or bad mark.

"If you haven't done as well as you thought you did you can go back and see, and try to figure out why they think you've done and on what basis." (3C)

"Had I got a piece of work back and it was given a 'C' then I might go to this and start thinking, 'why?' ...I would use it to understand where the mark came from to try and work out in my own mind what I didn't do to get a better mark." (1F)

"I've just got a report back and I'm not very pleased with it [the mark], so I'm going to go back and have a look at this sheet." (2B)

In some cases there was also the suggestion that such analysis using the grid might prove useful in making a case to appeal marks. An interesting observation made by the students was that feedback tended to concentrate on weaknesses and did not sufficiently help to identify strengths. If we tend to advise students on what they have done wrong rather than right they can only learn what was right by a (possibly lengthy) process of elimination. There would seem to be similarities here with the connoisseur approach to assessment in that even when (and if) they reach this point

they may still not know or be able to articulate exactly what the positive features are that make it 'right'.

Several participants went even further and suggested that using the grid to analyse feedback for different assignments over time could help to identify patterns of performance. Including both strengths and weaknesses and in some cases this analysis might highlight persistent weaknesses, in others it might even identify previously unidentified weaknesses.

“It is interesting to see when you’ve done a lot of assignments, that when you receive the assessment grids back, if you are consistently doing badly in a particular area”
(3C)

“Can help to identify weaknesses you didn’t know you had” (3D)

3. USEFULNESS TO IMPROVE STANDARDS AND MARKING CONSISTENCY

We have already considered the students criticisms of trying to use the grid to improve and identify their own standards, which focused around the criteria and grade definitions being too general and vague, and the subsequent problems of definitional interpretation and subjectivity. They also identified that this was likely to be equally problematic for staff.

“Each one of them [grade criterion] is open to interpretation, one, between each student, and two, between each lecturer” (1F)

“because it is not specific enough , it is trying to regulate the marks among the lecturers but I’m sure they have trouble using it too” (2C)

“If I was a lecturer I think I’d sit there and wonder what’s the difference between all of these” (2G)

“I expect if you asked all the lecturers to define ‘polished’, I expect they’d be very different” (2D)

It was also suggested that other more personal factors would prevent the grid from making staff marking completely reliable. These ranged from cynical views about staff moods and individual perverseness through to recognition of the influences of norm referencing and the connoisseur approach to assessment.

“Marking is a subjective [sic] thing, your assignment is actually going to be tainted by the one it follows (2E)

“I don’t believe at all that markers would stick to the criteria” (1G)

“If you are the sort of lecturer that doesn’t much desire to use the grid and will look at a piece of coursework, formulate its mark in the first five minutes of reading it, ‘Right this piece of coursework is worth a B+, now what do I need to write on this

then? I need to tick this box, etc. because that's how this piece of work will fit into being marked at this grade'” (1D)

Students' acute understanding of the subjective and normative nature of many forms of assessment provides an interesting response to any argument that the questioning of assessment practices can potentially alarm and demoralise students, undermine professional expertise and open the flood gate to appeals. Realistically, however, the gulf may not be so wide between proponents of the connoisseur model of assessment, largely norm-referenced and based on 'professional expertise' and “*pretty much impenetrable to the non-cognoscenti*” (Webster et al., 2000 p.73). and those calling for more explicit, pre-determined standards and criteria and grading independent of the performance of other students. In practice, Rowntree (1989) argues that it is naive to believe that assessors can ignore normative data when assessing the performance of a particular student against a particular criterion.

4. STUDENT SUGGESTIONS FOR IMPROVEMENT

The overwhelming general response from the students was that despite the practical problems in using and applying the grid, it was considered to be a well-conceived idea. They welcomed a more systematic marking process and clearly recognised the potential of the grid and what it was trying to achieve.

“I think its aims and objectives are very good and it could help students a lot” (1G)

“I just want to say that this is the perfect method to a lot of people that I’ve spoken to are like ‘I really don’t know what they want us to do with this coursework’ so if there’s really any way of getting the whole group at once to understand what you want with your coursework then via this grid it would be an excellent way to explain.” (1B)

When confronted with the crude prompt of ‘does the grid work in practice’ the majority of the focus group participants said ‘no’ but it is clear from their other comments, already illustrated, that they recognised the potential benefits of the grid. However, students appeared remarkably cognisant of the implementation issues associated with criterion referencing and from this standpoint they made some pertinent suggestions for improvement.

Guidance on the grid’s application and use to both staff and students

Students expressed a strong desire to see the grid used on all modules. Linked to this, students clearly considered that the grid needed to have all the staff seriously committed to it and ‘trained’ to use it.

“I think the lecturers need to understand that they should take this seriously because they are the people who are going to be marking the work” (1B)

“They have to be taught the grid as much as we do” (1C)

It was also clear from some of the detailed descriptions of specific modules where the grid was used that the grid needed to be seen by the course staff as an integrated part of the course design and not just something tagged on at the end.

Clarifying the grade descriptions by providing examples

Echoing Sadler (1987) there was also a prevailing consensus that the grade definitions needed further clarification and many students felt that this could only be done through the additional provision of examples.

“If you would, for example, give an example...this is what I call imaginative and comprehensive and give an example” (1B)

“There should be some examples; it is hard to understand some of the criteria let alone apply it” (2H)

“Something needs to be written, ‘this piece of work would be given an ‘A’ for understanding of the question because’ and you could see how it’s written” (1D)

“You learn more from examples than you do from abstract explanations” (2H)

Where the focus groups did not agree (in fact there appeared to be two diametrically opposed views) is whether it is most helpful to have the consistency of always retaining the same wording from the main grid in every grid that is used.

Alternatively, whether it would be better to adapt the wording each time to specifically relate to the individual task. To an extent there may be at least a partial overlap here with the distinction between giving clearer guidelines and prescribing content which has already been identified.

Active engagement in seminars

Further to just the provision of exemplars, there was also definite support in the focus groups (and their suggestion) for seminars where the students could actively engage with the application of the grid criteria, both at the beginning of each module

“It could work, if they really would take you through it...I think every module should go through it quickly” (1B)

and at the end of modules, after the assessed work has been returned

“You should have a seminar which specifically gives us feedback on the assignment.”
(2F)

“[and tutors could] say, ‘Right, for the presentation what we really wanted is this’ and given you an example of a really good piece of work, but demonstrated that, ‘and then for the content and range...we wanted you to include’” (2D)

“If he’d said, taken an example of a ‘C’, not the person but taken real examples of the assignment that we’d done and actually referred them back to this [the grid] and said, ‘This is one that got an ‘A’ and here’s an excerpt from it and this is what we think, why it deserved an ‘A’,’ it would make the grid seem real because it still seems very abstract.” (2E)

Introduce the criterion-reference grid in the first year

A common and unanticipated suggestion was that the grid should be introduced in the first year. Not knowing that the original intention of the grid had been to establish a common standard for Advanced level modules, the students not unreasonably argued that this raising of awareness of assessment criteria and what they mean would be much more appropriate in the first year.

“It’s no use to us now in our second year to give us all this. If they did it in the first year ... it would be so much clearer” (2H)

“So you could start off on a good foot. So you could get into good habits.” (2B)

“if this is going to be key to your time at university certainly in the way things are marked and you are fed back then more time needs to be given to it in the initial stages so that you fully understand it” (1D)

“If they had it in the first year then you might have a better idea of what to aim for”

(2H)

COMMENT

Students were evidently keen to discuss assessment issues and thirsty for a more reliable and consistent marking process, and many of the criticisms discussed by the focus groups moved beyond the use of the grid to criticisms of assessment practices in general. Educational literature repeatedly emphasises assessment as the most powerful influence on student behaviour with the capacity to encourage either deep or surface learning. Assessment is a subject that students perceive to be of fundamental importance, about which they grasp the salient issues, and are openly dissatisfied about the miscommunication we practise on the subject.

Students clearly recognised and accepted that individual staff members have a greater or lesser inclination towards the connoisseur approach to assessment than others. Student dissatisfaction and frustration surfaced when students experienced a lack of match between the published criteria and the feedback received. This dissonance arguably promotes the creation and maintenance of a ‘hidden curriculum’ (Sambell & McDowell, 1998) and undermines trust in the assessment process. Brown et al (1997) suggest that trust is fundamental in achieving a developmental learning assessment process, and as such, damage to student trust can inhibit students’ capacity to improve and, consequently, should be of serious concern to us all.

A substantial hurdle inhibiting the effectiveness of criterion referencing in general, and the grid in particular, is the problem of subjectivity and multiple interpretations of criteria and standards. Individual students interpreted not only the purpose of the grid differently but also the criteria themselves. It is clear that students did not readily understand or identify with the terminology used in the criterion descriptions. It was also apparent that some focus group participants found the required level of abstraction to link criteria with task too great. Realistically, how can we expect students to develop a constructive understanding when they are the 'novice' to the academic's 'expert' (Eraut, 1994) and even the experts find it difficult to reach agreement?

Readily understood and recognised by the students the issue of subjectivity and multiple interpretations is not an easy one to overcome. Nonetheless, we would argue that the *process* of defining and refining of verbal descriptors enhances shared subjectivity between staff as well as resulting in the better articulation of criteria and standards. However, achieving a workable balance between precision and utility is a difficult feat. Realistically, a single-minded concentration on the construction of evermore comprehensive and precise verbal descriptors appears self-defeating, firstly, as the very precision of language and terminology progresses it away from common parlance and makes descriptors obscure and opaque to novice students. Secondly, increasing the *quantity* of explanation can make for unwieldy, less transferable and useable descriptors.

Consequently, we would suggest that developing a shared understanding of assessment criteria and standards requires a multifaceted approach. Accordingly, an

assessment criteria and standards framework encompassing both tacit and explicit knowledge transfer processes involving verbal descriptors, exemplars, imitation and practice appears to be worth pursuing. Such an approach seems to be sought and suggested by students as they struggle to find firmer footing within the changing tides of assessment criteria and standards.

CONCLUSION

Over the last decade calls for public accountability have led to increasing interest in developing robust assessment systems that portray actual rather than comparative student achievement. This research highlights student concern over, and awareness of, key assessment issues. Eager to find ways of achieving consistency and trust in assessment processes, students are keen to share understanding with assessors on standards and criteria, and as a consequence, improve their performance.

Students regarded the criterion-reference grid as a well-conceived assessment tool that is, however, of limited practical use if presented on its own without the benefit of explanation, exemplars and the opportunity for discussion. The imprecision inherent in verbal description requires consideration be given to other ways of achieving shared understanding on criteria and standards. The grid in its present quasi-scientific form has incorporated too great an assumption about the nature of the knowledge to be transferred and our ability to transfer it. Plausibly, this may involve looking at the grid as a process tool facilitating shared understanding between staff and students. Thus playing a key part, but only a part, in a multifaceted framework of assessment

processes rather than as a stand-alone tool for explicit verbal description of standards and criteria.

NEXT STEPS

In many ways this research has illuminated the complexity of the assessment process. However, whilst it seems naïve to anticipate a completely sound assessment framework it is possible to visualise a more definite framework of assessment processes encompassing a multiple, composite approach. An approach which not only enhances the specification of standards and criteria, but also facilitates the development of a common understanding not just between staff and students but all stakeholders in the assessment process.

Accordingly the next step is to introduce the grid within a framework of explanation, practise, discussion and exemplars on a large (300+ student) business module and monitor and evaluate student and staff reaction.

The authors would welcome correspondence with any others engaged in a similar venture.

BIBLIOGRAPHY

Brooker, R., Muller, R., Mylonas, A. and Hansford, B. (1998) Improving the assessment of practice teaching: a criteria and standards framework, *Assessment and Evaluation*. **23**, 1, 5-24.

Brown, T. & Atkins, M. (1988) *Effective Teaching in Higher Education*, Routledge, London.

Brown, G., Bull, J., Pendlebury, M. (1997) *Assessing Student Learning in Higher Education*, Routledge, London.

Eraut, M. (1994) *Developing Professional Knowledge and Competence* The Falmer Press, London.

Gipps, C. V. (1994) *Beyond Testing*, The Falmer Press, London.

Greator, J. (1994) The nature and future of academic levels; the implications for the accreditation of prior learning, presented at the *Nurse Education Tomorrow Conference*, St.Mary's College, University of Durham.

Higher Education Quality Council (HEQC), 1996, *Graduate Standards Programme Draft Report and Recommendations* HEQC, London.

Laming, D. (1990) The reliability of a certain university examination compared with the precision of absolute judgements, *Quarterly Journal of Experimental Psychology Section A - Human Experimental Psychology*, **42**, 2, 239-254.

Lucas, L. and Webster, F. (1998) Maintaining standards in higher education? A case study, in: Jary, D. and Parker, M. (eds) *The New Higher Education; issues and directions for the post-Dearing university* Staffordshire University Press, Stoke-on-Trent, 105-113.

Moon, J. (1995) *The Description of Levels in Higher Education: a discussion paper to support development of level descriptors for higher education in Wales* HECATS - Learning Outcomes Project, Wales Access Unit, Cardiff

Newstead, S. E. and Dennis, I. (1994) Examiners examined: the reliability of exam marking in psychology, *The Psychologist: Bulletin of the British Psychological Society* **7**, 216-219.

Nonaka, I. (1991) *The Knowledge-Creating Company*, The Harvard Business Review, November-December, 96-104.

Otter, S. (1992) *Learning Outcomes and Higher Education* Unit for Development of Adult Continuing Education, Leicester

Price, M. and Rust, C. (1999), The experience of introducing a common criteria assessment grid across an academic department, *Quality in Higher Education*, **5**, 2, 133-144.

Rowntree, D. (1989) *Assessing Students: How shall we know them?*, 2nd Edition, Harper and Row, London.

Spender, J. C. (1989) *Industry Recipes. The Nature and Sources of Managerial Judgement*, Blackwell, Oxford.

Sadler, D.R. (1987), Specifying and promulgating achievement standards *Oxford Review of Education*, **13**, 2, 191-209.

Sambell, K, and McDowell, L. (1998), The construction of the hidden curriculum: messages and meanings in the assessment of student learning, *Assessment and Evaluation in Higher Education*, **23**, 4, 391-402

Webster, F., Pepper, D., and Jenkins, A (2000) Assessing the undergraduate dissertation, *Assessment and Evaluation in Higher Education*, **25**, 1, 72-80