

Improving Students' Learning by Developing their Understanding of Assessment Criteria and Processes

CHRIS RUST, MARGARET PRICE & BERRY O'DONOVAN, *Oxford Brookes University, Oxford, UK*

ABSTRACT *This paper reports the findings of a two-year research project focused on developing students' understanding of assessment criteria and the assessment process through a structured intervention involving both tacit and explicit knowledge transfer methods. The nature of the intervention is explained in detail, and the outcomes are analysed and discussed. The conclusions drawn from the evidence are that student learning can be improved significantly through such an intervention, and that this improvement may last over time and be transferable, at least within similar contexts. This work is a development within a longer and ongoing research project into criterion-referenced assessment tools and processes which has been undertaken in the pursuit of a conceptually sound and functional assessment framework that would promote and encourage common standards of assessment; that project is also summarised.*

Introduction

Within Higher Education there is an increasing acceptance of the need for a greater transparency in assessment processes, and moves have been made to make methods of assessment clearer to all participants. This paper is concerned with the extent to which students understand these processes and how we might improve their understanding of them. It presents the development and planning of a two-year project involving the transfer of knowledge of the assessment process and criteria to students in a variety of ways; in particular, through a structured process involving both tacit and explicit knowledge transfer methods. The aims of this project were to improve the students' performance through enhancing their ability to assess the work of others and, in consequence, their own work, against given marking criteria. The initial findings of the first year of the project, the methodology and its background were first reported at the

8th Improving Student Learning Symposium in Manchester, England, and first published in the conference proceedings (Price *et al.*, 2001). The success of the project, and a replication of the exercise with a second cohort the following year, has now been evaluated from a number of perspectives, the most important of which being by gauging the subsequent effect on the students' performance. A further evaluation of the longer-term effect on performance has also been carried out on the first cohort.

Background

This work is a development within an ongoing research project into criterion-referenced assessment tools and processes, which has been undertaken in the pursuit of a conceptually sound and functional assessment framework that would promote and encourage common standards of assessment. The earlier findings from this larger project have informed the development of this research and have already been reported elsewhere (Price & Rust, 1999; O'Donovan *et al.*, 2001), and are summarised below.

Context

The research project into criterion-referenced assessment tools and processes commenced in 1997 against a background of growing national concern in the UK about marking reliability, standards and calls for public accountability (Laming, 1990; Newstead & Dennis, 1994). At a national level within the UK compelling pressure was beginning to be applied to higher education institutions to maintain high academic standards (Lucas & Webster, 1998). This pressure has been escalated 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). This trend has been compounded by the rapid expansion of student numbers and a drastic cut in the unit of resource for UK higher education. The debate about standards was further informed by a national discussion on generic level descriptors (Otter, 1992; Greatorex, 1994; Moon, 1995; HEQC, 1996) which were seen by some as a means of establishing common standards. The focus of this discussion tended to be on the need for explicitness, with the implication that if all were made explicit this would be sufficient to establish standards. Little, if any, mention was made about involving students in the process.

In response to this, the Quality Assurance Agency (QAA) embarked on a new quality assurance system, with three distinct elements—benchmark standards, programme specifications, and a national qualifications framework—all intended to bring about the establishment of explicit degree standards. However, it is interesting to note that when the benchmarks were published in May 2000 they were retitled benchmarking *statements*. Arguably, this change recognised the failure of the process to clearly define explicit standards for all subjects. At a conference on Benchmarking Academic Standards (Quality Assurance Agency, 17 May 2000), Chairs of the QAA subject panels commented on the difficulties of defining threshold standards and using language which meaningfully conveyed level. However, the benefit realised by the academic community from the process of drawing up the statements was emphasised. Professor Howard Newby stated:

I would certainly want to assert the value to self-understanding in disciplines of debating the basis on which the discipline is conducted and what the

students need in order to be able to participate in the community of scholars who practise it. (QAA, Benchmarking Academic Standards Conference, 17 May 2000)

First Steps

The initial impetus to address the issues in this project came from an external examiner for the Business Studies undergraduate 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. As a consequence of this, a common criteria assessment grid was developed for the Business School and first piloted in the academic year 1997–98. The grid has 35 criteria plotted in matrix format against grades resulting in 'grade definitions' detailing acceptable performance for each criterion at each. Staff select appropriate criteria for any given assessment to create a 'mini-grid' (see Figure 1 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 in better work), and making it easier to give effective feedback to the students.

Staff and Student Views

The use of the grid has been evaluated through the views of staff and students as well as noting the feedback from external examiners.

The main conclusion of the initial paper (Price & Rust, 1999) was that, at least in its present form and usage, the grid failed to establish a common standard—different tutors having taken the grid and used exactly the same grade definitions for a basic module (one normally taken by first-year students) and an MBA module apparently without any difficulty. However, the paper further concludes 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. However, 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 and assist in focusing 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 grade definition by both individual staff members (Webster *et al.*, 2000) and students.

The views of students were sought when they had experienced the grid on a variety of modules, and more detailed findings have been reported elsewhere (O'Donovan *et al.*,

7029 Placement Search and Preparation — Feedback Sheet
ASSIGNMENT 1

Student Name: Student Number:

CRITERION	A	B +	B	C	Refer/Fail
1 Presentation of assignment	Shows a polished and imaginative approach to the topic	Carefully and logically organised	Shows organisation and coherence	Shows some attempt to organise in a logical manner	Disorganised/incoherent
7 Attention to purpose	Has addressed the purpose of the assignment comprehensively and imaginatively	Has addressed the purpose of the assignment coherently and with some attempt to demonstrate imagination	Has addressed the main purpose of the assignment	Some of the work is focused on the aims and themes of the assignment	Fails to address the task set
27 Self-criticism (include reflection on practice)	Is confident in application of own criteria of judgement and in challenge of received criteria and begins to develop own criteria and judgement.	Is able to evaluate own strengths and weaknesses; can challenge received opinion strengths and weaknesses.	Is largely dependent on criteria set by others but begins to recognise own weakness.	Dependent on criteria set by others. Begins to recognise own strengths and	Fails to meaningfully undertake the process of self criticism.
28 Independence/Autonomy (include planning and managing learning)	With minimum guidance can manage own learning using full range of resources for discipline; can seek and make use of feedback.	Identifies strengths of learning needs and follows activities to improve performance; is autonomous in straight forward study tasks.	Can work independently within a relevant ethos and can access and use a range of learning resources.	Can undertake clearly directed work independently within a relevant ethos and, with some guidance, use the standard learning resources.	Unable to work independently, needing significant guidance on methods and resources.

(Please tick boxes)

Comment:

Marker: Mark:

FIG. 1. Sample marksheet.

2001). The students felt the criterion-referenced grid to be a well-conceived assessment tool and clearly recognised the potential of the grid and what it was trying to achieve. However, it was also seen as of limited practical use if presented in isolation without the benefit of explanation, exemplars and the opportunity for discussion. The need for such aids resulted from the identification of several issues undermining the easy application of the grid. These issues included the need to clarify the meaning of terms and phrases; subjectivity and multiple interpretations of criteria and standards; a lack of match between published criteria and the feedback received, arguably, suggesting the presence of a 'hidden curriculum' (Sambell & McDowell, 1998; Webster *et al.*, 2000).

Implications

A common view of both staff and students was the need for discussion to support the use of the grid, between staff in the marking process, and between staff and students to enhance students' understanding of the assessment process and as a result to improve their performance. Students also identified exemplars and further explanation as useful in making the assessment criteria more comprehensible.

When QAA experts are unable to make standards explicit after months of learned debate, we should, perhaps, begin to question the single-minded focus on explicit articulation of standards. It is difficult to relinquish the notion that academic standards can be documented and codified in such a way that they may be available for the passive consumption of all stakeholders in higher education. However, our research experience has been that, without active involvement through discussion and debate, the development of a common view on standards and level is problematic, if not impossible—even within a close-knit community of a single academic department. Obstacles to the transfer of knowledge about standards and assessment requirements are accentuated when such transference takes place with more 'novice' students undertaking modular courses in which they have very limited time to construct cohesive, 'objective', interpretations of assessment standards. Obstacles which are further heightened in a broad-ranging, multidisciplinary and discursive subject such as Business and management—a discipline in which many open-form assessments, i.e. essays and reports, are set on topics that are integrative in nature.

Consequently, we suggest that the imprecision inherent in passively presented verbal description requires that consideration be given to other ways of achieving shared understanding of criteria and standards. Arguably, in its present quasi-scientific form, the grid incorporates too great an assumption about the nature of knowledge to be transferred and our ability to transfer it.

Transferring Tacit Knowledge

Polanyi (1998) and Tsoukas (1996) among others, argue that the transfer of *useful* knowledge involves the transmission of both explicit and tacit knowledge. Consequently, a single-minded concentration on explicit knowledge and careful articulation of assessment criteria and standards is not, in itself, sufficient to share useful knowledge of the assessment process.

We conjecture that more complete or 'useful' knowledge of a particular assessment starts in the mind of an individual assignment writer. However, in the process of transferring this knowledge to others, parts of the knowledge can be difficult to articulate and consequently go 'missing' from the final communication. In Polanyi's words "we

can know more than we can tell” (Polanyi, 1998, p. 136). Such missing knowledge can be described as tacit—tacit knowledge in this context being defined as something that we know but we find impossible or, at least, extremely difficult to express. Unlike explicit knowledge “that can be expressed in formal and systematic language” (Nonaka *et al.*, 2002, p. 43) tacit knowledge is highly personal and hard to formalise. Deeply rooted in action and often in an individual’s commitment to a profession, tacit knowledge consists partly of technical skills based on professional experience, and in a more cognitive dimension, in our ingrained mental models, beliefs and perspectives (Nonaka, 1991). A type of knowledge that can be said to underpin the established normative, ‘connoisseur’ model of assessment—illustrated by the phrase “I cannot describe it, but I know a good piece of work when I see it”, most often likened to the skills of perfume or tea-blending and “pretty much impenetrable to the non-cognoscenti” (Webster *et al.*, 2000, p. 73).

Knowledge of this kind is experience-based and can only be revealed through the sharing of experience—socialisation processes involving observation, imitation, dialogue and practice (Nonaka, 1991; Baumard 1999). So over time discussion and shared experiences of marking and moderation among staff enable the sharing of tacit knowledge, resulting in more standardised marking (Saunders & Davis, 1998). It follows that inviting students into this shared experience should also enable more effective knowledge transfer of assessment processes and standards to them.

The Research Project

Aims of the Project

The primary aim of the project was to improve student performance, through a process that would confirm our understanding that the value of tacit knowledge and its role in the learning process has been relatively neglected in the current climate of higher education, and that this neglect has been detrimental to student learning. We aimed to increase students’ knowledge of assessment criteria and processes, and as a consequence, improve student performance through a structured intervention devised for a large (300 +), basic (first-year), undergraduate, Business module; an intervention that placed increased emphasis on the tacit knowledge transfer processes of practice and imitation to achieve shared understanding by engaging students in a series of activities that combined discussion of exemplars and marking criteria, marking exercises and self-assessment. The effectiveness of activities involving marking exercises and peer marking have also been supported by claims from the USA (Nelson, 1994) and studies from the UK (Forbes & Spence, 1991; Hughes, 1995); studies that have shown significant subsequent improvement in the work of students involved both in marking exercises and peer marking.

Intervention Design

The intervention took place in the final three weeks of the students’ first term on a Business degree programme. It involved students in preparation work, participation at a workshop, and the submission of a self-assessment sheet, along with their coursework to be handed in at the end of the first term (three weeks after the workshop).

The detail of this intervention was as follows:

- (a) A week before the workshop all students on the module were provided with two sample assignments and mark sheets including assessment criteria and grade

definitions. Students were asked to individually complete the mark sheets, providing a grade, marks and rationale/feedback for each of the assignments before coming to the workshops.

- (b) Workshops (90 minutes long) were offered to all students in groups of 40. (Student attendance at these workshops was optional.) The workshops were structured in the following way:
- (i) student discussion in small groups of their initial individual marking of the two sample assignments;
 - (ii) feedback of small groups' agreed grades and rationale to plenary;
 - (iii) tutor-led comparison of provided rationales with criteria;
 - (iv) tutor explanation of each criterion;
 - (v) small groups review assessment and grade in light of tutor explanation;
 - (vi) final report from small groups to plenary of grade for each piece of work;
 - (vii) tutor provided annotated and marked versions of samples and discussed tutor assessment and mark.

The small group discussions allowed the student to compare and justify their initial assessment of the work against that of others as well as allowing the declared grade to be the responsibility of the small group. However, the students were asked explicitly not to change their initial grading on their individual sheets.

- (c) Three weeks later, students submit their coursework along with a completed self-assessment sheet. There was no indication on the coursework whether students had participated in the intervention or not.

The mark sheets used for the sample assignments, the tutor's actual assessment of the students' submitted work and for the students' self-assessment were the same, incorporating comments, a module assessment grid, a grade and a mark.

Data Collection

The intervention was designed to run twice, in successive years, with two different cohorts on exactly the same module. Each year it was carried out in exactly the same way, so that it would provide comparisons between student participants and non-participants at the workshops for quantitative and qualitative analysis. Further data were also collected from two other modules. The three modules concerned had similar assessment methods and were as follows:

- Module 7009—a first-year Introduction to Business module on which the intervention took place;
- Module 7508—a large first-year Introduction to Business Economics with assessment taken by the students prior to the assessment on Module 7009;
- Module 7026—an advanced second-year Business module taken one year after 7009.

Measuring and Comparing Student Performance

Quantitative measurement of student performance was undertaken using three comparative statistical measures.

- (i) A baseline comparison—comparison of the assessment performance of the participants at the assessment workshop with the non-participants on a module (Module

7508) taken before the intervention was carried out. This comparison was carried out in two successive years (1999 and 2000) with successive cohorts;

- (ii) A treatment comparison—comparison of the assessment performance of the participants at the assessment workshop with the non-participants on a module (Module 7009) within which the intervention was carried out. This comparison was also carried out in two successive years (1999 and 2000) with the same successive cohorts as in (i);
- (iii) A transfer comparison—comparison of the assessment performance of the participants at the assessment workshop with the non-participants on a module (Module 7026) taken by both groups one academic year later than Module 7009 (in 2001)

Data preparation was carried out in Microsoft Excel. The between-group comparisons used independent group t-tests and were carried out in *AlStats*. Because of the number of t-tests used ($n = 5$) a Bonferroni correction was applied to convert the 5% significance level (one-tailed) to the 1% level (one-tailed). In advance of the study a significance criterion of 0.01 (1%) was set. The effects ratio (mean difference divided by standard deviation) was also calculated for each comparison with 0.5 considered a moderate effect (Cohen, 1969).

Gauging Student Knowledge of the Assessment Criteria and Processes Prior to the Intervention

In order to test out the students' initial knowledge of the assessment criteria and processes, it was necessary to test their application of the criteria. Through the use of the sample assignments the students' initial attempts at marking were collected via the completed mark sheets from those students who participated in the workshop. The mark sheets provided the grade (A–C or F) and mark (%) awarded, their justification for the mark, which identified strengths and weaknesses of the work, and a completed assessment grid showing their application of the marking criteria and grade definitions in marking.

At the workshop student contribution was monitored through non-participant observation. Notes were taken on the grades awarded by the students for the sample assignments following small group discussion as well as the justification provided for those grades.

Gauging Student Perceptions of the Workshop

Using a questionnaire with position statements and Likert scales of agreement/disagreement, students were asked to evaluate the workshop in terms of its effect on their understanding of criteria and assessment processes and their levels of confidence in completing their assessed work and applying the criteria to their own work.

Student Ability to Self-assess

On Module 7009, after the intervention, when students submitted their coursework for assessment they were also required to submit a self-assessment of their work. It was the intention of the researchers to compare the student's self-assessed grade overall and for each criterion with that of the marker to provide an indication of the student's understanding of level. This was only in fact possible with the second year cohort (see below).

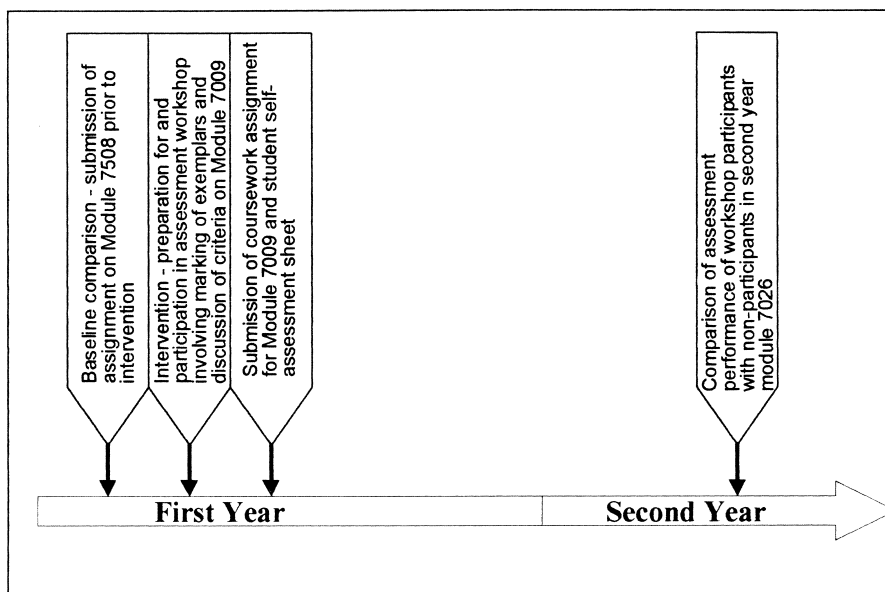


FIG 2. Timeline for one cohort.

Staff Perceptions

In the first year of the project, using informal interview/unstructured discussion, the seminar tutors (also the first markers of the assessed coursework) were asked about student response to the intervention and their perceptions of how well the students had done the assessed coursework. In particular, those who had taught on the module in the previous year were asked to comment on how the work compared.

Methodological Issues

Participant Self-selection

The participating students were self-selected due to the non-compulsory nature of the assessment workshop. There was therefore a concern that these student participants might not be representative of the cohort population. Results could be easily skewed with the participant sample differing from the population in key attributes such as ability or motivation.

The results were tested in two ways: firstly, the standard deviation of the marks for each group was calculated and compared to surface any anomalies. Secondly, the marks achieved by the participating students were compared against those of the non-participating students on a piece of coursework on another module (Module 7508) submitted prior to the workshop sessions (a baseline comparison).

Data Contamination from Exemplars

Identifying the sample of students that had taken part in the assessment workshops was straightforward. However, it must be recognised that the non-participant remainder of the

cohort were not completely isolated from the process. All students were issued with the original sample assignments for their initial preparatory assessment and, more importantly, it is inevitable that the tutor-marked and annotated versions of the samples will have been widely circulated (although only given directly to workshop participants). So although many of the cohort did not take part in the knowledge transfer process within the workshop, many may have had the benefits of exemplars of the standards. The effect of this may well have been to diminish any differences between the performance of the participating sample and the rest of the cohort. However, this serves to heighten the significance of the active involvement of the participating students in the difference that did emerge between the two groups.

Norm-referencing in Marking

There is a potential problem in using student assessment results as a measurement of student performance due to the tendency of assessors to norm-reference which remains the dominant model within higher education and ‘naturally’ preferred by most markers (Professor A. Wolf, SHRE Assessment Network Meeting, 13 March 2000). Even though the application of criterion-referenced marking should supplant the application of norm-referenced marking it seems that the latter remains an influencing factor in marking whichever is used (Professor A. Wolf, SHRE Assessment Network Meeting, 13 March 2000).

The difficulty posed by norm-referencing within quantitative analysis of assessment results is that if assessors mark comparatively to an average, say of 55%, then inter-module comparison and longitudinal comparison of results can be meaningless. The numerical description of 55% simply becomes the portrait of average whilst the ‘real’ average performance of two different cohorts could be very different. Consequently, numerical descriptions can confer delusions of accuracy. Qualitative data from the markers about their perceptions of the quality of the students’ work help counter this effect of norm-referencing.

Student Self-assessment

In the first year of the project, because of what with hindsight can be seen to have been an error in the methodology (i.e. the assessment sheets on which both the students self-assessed and then the tutor assessed were photocopied for use on this research project, prior to their return to the student with their work), it was not possible to tell which assessments had been made by the student and which by the tutor, making any analysis impossible. This error was rectified in the second year.

Findings

The Effect on Performance Standards—Results of Statistical Analysis

The major findings of this study, repeated over two years, are that the participants at the assessment workshop subsequently achieved significantly better results in their assessed coursework. Whilst comparison of the performance of participants and non-participants on a module prior to the intervention showed no significant difference in the performance of the two groups. Furthermore, one year later, the participants of the first cohort of the study were still showing significantly better results than the non-participants.

TABLE 1. Cohort 1

Module	Participant (mean)	Non-Part (mean)	df	p	Mean difference	Std dev.	Effects ratio
7508	58.6	57.8	191	0.56	0.7606	9.08	0.08
7009	59.78	54.12	292	0.00001	5.649	9.49	0.6
7026	57.91	51.3	182	0.004	6.604	11.91	0.55

Note: Participants = 151, Non-participants = 143.

TABLE 2. Cohort 2

Module	Participant (mean)	Non-Part (mean)	df	p	Mean difference	Std dev.	Effects ratio
7508	58.67	55.75	133	0.06	2.92	9.32	0.31
7009	59.86	52.86	319	0.00001	6.9716	10.03	0.69

Note: Participants = 152, Non-participants = 169.

As can be seen from the tables of results (Tables 1 and 2), with both cohorts there was no significant difference between the participant and non-participant groups in the baseline comparison on Module 7508, where the assessment was submitted before the training took place. (And power analysis of these figures provides power estimates in both cases which show the design was almost certainly adequate to have identified differences should they have existed—Power (5%) = 93% and 76% respectively.)

After the intervention, there is a significant difference between the results of those participating in the workshop and for those who did not, for both cohorts ($p < 0.01$) on the module (7009) on which the intervention took place, and that significance can still be identified one year later for the first cohort ($p < 0.01$) (on module 7026) although the figures may suggest, not surprisingly, a minor reduction in the effect. This is also reflected in the effects ratios, 0.6 and 0.69 in the module on which the intervention took place (7009), and 0.55 in the module one year later (7026).

The Results of Student Self-assessment and Comparison of their Accuracy

All students were asked to complete and submit a marking criteria grid, grading their work for each of the five criteria as well as giving an overall grade, but in the first year the data was not usable (as has already been explained above). In the second year, however, 140 (92%) workshop participants and 150 (89%) non-participants complied with the request and gave themselves an overall self-assessment grade. Only a somewhat disappointing 68 (45%) participants and 38 (22%) non-participants additionally graded themselves for each of the individual criteria. The students' self-assessments were then compared with the grades given by their tutors, and a comparison made between the workshop participants and non-participants to see if the former were able to be more accurate in their self-assessment (Table 3).

To make a straight comparison of the accuracy of the self-assessments, a simple numerical system was devised, whereby zero indicated that student and tutor had put the same grade, one indicated a one-grade difference (plus one if the student's grade was higher, i.e. an overestimate; minus one if the student's grade was lower, i.e. an

TABLE 3. Self-assessment accuracy

Overall grade diff.	Female Participant (n = 83)	Male Participant (n = 57)	Total Participants (n = 140)	Female non-participants (n = 52)	Male non-participants (n = 98)	Total non-participants (n = 150)
+ 3	1 (1%)	0	1 (0.7%)	0	8 (8%)	8 (5%)
+ 2	7 (8%)	4 (7%)	11 (7%)	3 (6%)	15 (15%)	18 (12%)
+ 1	17 (20%)	10 (18%)	27 (19%)	12 (23%)	18 (18%)	30 (20%)
0	31 (37%)	23 (40%)	54 (39%)	24 (46%)	44 (45%)	68 (45%)
- 1	23 (28%)	15 (26%)	38 (27%)	10 (19%)	12 (12%)	22 (15%)
- 2	4 (5%)	5 (9%)	9 (6%)	3 (6%)	1 (1%)	4 (2.5%)

underestimate), two indicated a two grade difference, etc. (the range of possible grades being A, B +, B, C, F).

Comparison of Overall Self-assessment Grades with Tutor Grades

An initial comparison of the overall grades given by the students in the two cohorts is disappointing, because there does not seem to be any great difference; in fact, with 54 (39%) participants and 68 (45%) non-participants accurately predicting their grade and 27 (19%) participants and 30 (20%) non-participants only over-predicting by one grade, it looks as if, if anything, the non-participants are more accurate. However, on closer examination it is interesting to note that if we compare those either overestimating by two or three grades, and those underestimating, there is a clear difference. Although these are quite small numbers, especially in the case of non-participants, there is a distinct pattern in these results, with a higher percentage of non-participants overestimating the grade and a higher percentage of participants underestimating the grade. Given the suggestion that has been made in some of the literature on self-assessment that women may be more likely to underestimate their true worth while males may be inclined to overestimate (Thomas, 1990; Gibbs, 1991), it is also interesting to note that while amongst the participants there appears to be no identifiable difference between the males and females, with the non-participants there does seem to be more male overconfidence, 41% compared to 29% overestimating and 23% compared to 6% overestimating by two or more grades. Comparison of the two male groups shows male non-participants more confident than participants—41% overestimating their grade compared to 25%, and 23% compared to 7% overestimating by two or more grades. Conversely, 13% of male non-participants graded their assignment lower than their tutor in comparison to 35% of male participants. Females showed less of a startling difference—25% female non-participants underestimated their work compared to 33% of female participants—but nevertheless, participants were still more likely to underestimate.

Comparison of Individual Criterion Grades

As has already been stated above, the numbers of students who graded themselves for each criterion was disappointingly small [68 (45%) participants and 38 (22%) non-participants] which makes it impossible to claim any statistical significance for the results. However, the same pattern of more overestimation by non-participants and underestimation by participants is evident. And it is perhaps interesting to note that the

criteria for which this is most marked are criteria 4 and 5 (analysis and evaluation)—the arguably more difficult ‘invisible’ criteria.

The Extent of Student Understanding of Assessment Criteria and Processes

Applying Criteria

The students’ initial attempts at grading, and the workshop discussion, showed that they exhibited more confidence in applying certain ‘visible’ criteria, such as structure, presentation and referencing. These criteria were used extensively as justification of the grade awarded by individual students on their mark sheets and by the small groups in the workshop. Conversely, the students found difficulty in applying more ‘invisible’ criteria such as analysis, evaluation, etc. There was a deep reluctance to use these criteria at the start of the process. Even when they were mentioned in the justification of the marks, their application was weak. Following explication of the criteria the students in the workshops then began to apply the ‘invisible’ criteria but still found it difficult to use them to justify marks. Many students commented on how difficult they found the marking task and their fear at exposing a lack of ability to assess.

Understanding of Level

In the first year, the evaluation of students’ ability to assess was based on an analysis of 116 mark sheets and marking grids completed and handed in by 116 of the 151 students who participated in workshops.

The initial grades taken from the mark sheets for each of the sample assignments showed that almost every student had correctly identified the excellent and poor piece of work relative to each other. However, for each piece of work there was a range of grades awarded.

Assignment One had been graded as an A by the tutors, while the profile of marks awarded by the students is as shown in Table 4.

Assignment Two was graded as C by the tutors and the students’ assessment showed slightly more alignment, with the profile of marks awarded by students as shown in Table 5.

After initial group discussion in the workshop, several groups were in line with the then unknown tutor grading (it should be noted that one might expect the more ‘extreme’ individuals to be moderated out by the group process). Later in the workshop, following tutor-led discussion and group review, a further 8 out of 39 groups shifted their grades

TABLE 4. Student grading of Assignment One

A 34 students (29%) (3 students did not grade this assignment)	B + 63 students (54%)	B 16 students (14%)
--	--------------------------	------------------------

(Graded A by the tutor)

TABLE 5. Student grading of Assignment Two

B + 1 student (0.9%) (3 students did not grade this assignment)	B 22 students (19%)	C 87 students (75%)	F 3 students (3%)
---	------------------------	------------------------	----------------------

(Graded C by the tutor)

to the still unknown tutor grading. None of the groups shifted away from tutor grading. It is interesting to note that students graded more conservatively than tutors did, even after involvement in the series of structured activities. This was more noticeable at the higher level than the threshold pass. This conservatism may have been as a result of a lack of confidence in applying the criteria or an expectation of higher standards.

Student Perceptions

Feedback from students indicated that they viewed the workshop very positively. They felt that the activities and discussion had contributed ‘a lot’ to their understanding of marking criteria and their assignment. Also, generally, they felt more confident about their assignment preparation, although a small minority stated that they felt less confident after the workshop, because although they better understood the level required to pass, they were concerned about their ability to meet it. Many students requested that the workshop be scheduled earlier before they undertook any coursework on other modules.

Markers, Perceptions

In the first year of the project, markers perceived the standard of student coursework on the module to be higher than that of previous years but there were indications of convergence of style and structure of the student work. The scripts were not separated for marking and the markers were not aware which of the students made up the participating sample. The markers could only form an overall impression of standard rather than improvement of particular students’ work.

Discussion

Countering an Over-reliance on Explicit Knowledge

The move towards greater transparency in assessment processes has been founded on the articulation of standards, levels and criteria for assessment in written format. The provision of information in such a format was considered sufficient to increase participants’ understanding of the processes. This assumption also underpinned our initial attempts to achieve more effective and resource efficient knowledge transference of assessment standards and criteria solely through explicit articulation—embodied in a common criteria assessment grid.

However, in light of the limitations and difficulties encountered by both staff and students in their use of the grid (Price & Rust, 1999; O’Donovan *et al.*, 2001) which arguably echo the experience of the QAA in seeking to establish benchmark standards, such assumptions about the transfer of knowledge of assessment processes need to be questioned. Both the QAA’s failure to establish (through subject benchmarking) explicit standards, and the grid’s limitations, may lie with the assumption that all aspects of assessment standards can be articulated and made explicit. An assumption that furthermore does not fit comfortably with the application of standards through the use of the traditional assessment model that relies on a normative, connoisseur approach—a connoisseur approach undertaken by those that regard “assessment as akin to wine tasting” (Webster *et al.*, 2000 p. 73). Such an approach appears to rely on a relationship between student and tutor developed over time to achieve the transfer of knowledge, both explicit and tacit, from novice to expert (Eraut, 1994); a transfer process that takes

place gradually, if at all, through complex social processes relying on feedback and discussion. However, even if one were to concede that in the past it may have been reasonably successful with a majority of students (and some may wish to challenge that), it is surely questionable whether in these times of increased student numbers, an increasingly diverse and 'part-time' student body and diminished resources all leading to reduced staff–student contact, such a process can be relied upon to take place automatically—and certainly not for all students.

Consequently, we suggest that in the context of today's higher education we must move away from sole reliance on the explicit articulation of assessment standards and criteria. To transfer useful assessment knowledge on which students can construct improved performance we must also involve the tacit domain.

Improving Student Performance Through the Use of Explicit and Tacit Knowledge Transfer Processes

Improved performance of participants. The intervention supported the transfer of tacit knowledge through the use of exemplars, marking practice and the opportunity for dialogue between staff and students to compliment explicit knowledge provided through the verbal explication of assessment criteria by staff and in written format embodied within the grid.

Whilst the intervention was constructed to use both explicit and tacit knowledge transfer processes, the findings from this project point to the significant factor in knowledge transfer and sharing being the socialisation processes focused on in the workshop. Given that all students were provided with samples of work (to mark) prior to the assessment workshop and annotated versions, given out at the workshop, were widely circulated among the whole student group, the workshop remains the distinguishing aspect of the process. Only those students taking a full part in all the activities were seen to perform to a significantly better standard than the rest of the cohort. Evidence from the literature on peer-marking using model answers (Forbes & Spence, 1991; Hughes, 1995) would suggest that it is being engaged with the process of marking as well as seeing examples of other work that significantly contributes to the students' subsequent improvement in performance.

It could be argued that the significant improvement in the participant group performance when compared with the non-participant group resulted because of their voluntary participation in the intervention process. Undertaking the baseline comparison using a module assessed prior to the intervention has provided assurance that there was no significant difference between the participants' and non-participants' performance prior to the intervention, thereby eliminating ability as a basis for self selection. Also, the repetition of the research with almost identical outcome eliminates a chance outcome. The remaining major variable that affects all the students is the intervention.

Improved performance of whole cohort. As noted in methodological issues, the documentation that was provided and circulated to the whole student group is likely to have benefited the non-participants as well as the participants. This supposition was borne out by the staff perceptions of the work of the whole cohort when they signalled that the standard of student coursework as a whole had risen from standards prior to the introduction of the intervention on the module. This being the case, the treatment analysis (comparing performance of participants and non participants) will have understated the effect of the whole intervention on the participants. The treatment comparison

of two groups will have been using data showing improved performance by both groups, albeit of differing value. So the improvement in the participants' performance was probably greater than shown and the non-participants' performance probably also improved. The difficulty in verifying improvement to the whole cohort lies within the problems associated with norm-referenced marking (as discussed earlier).

Long-term Effects

Albeit that evidence is based on only one cohort so far, there is encouraging evidence that a relatively simple intervention, taking a relatively small amount of course time, can have an effect which can last over time and be transferred. Admittedly the follow-up module was deliberately chosen because of the similarity of the nature of the assessment task and the assessment criteria used—so no grand claims for transferability can be made on this evidence, but it would be interesting to extend this research project to look at the performance of these students on very different types of assessment.

Self-assessment Accuracy

Disappointingly at first glance the data suggest that the intervention does not appear to make those who participated any better able or more accurate in their self-assessment compared with those who did not participate. However, on closer inspection this appears to be only part of the story. Plausibly, rather than making students more accurate in their self-assessment it may be that exposure to the exemplar assignments opened up their horizons to what really good work could look like and the potential for improvement. As a consequence of such exposure, we conjecture that participants underestimated the quality of their own work and that this seems to have had more of an effect on the previously over-confident male students. This supposition receives tentative confirmation from student perceptions of the intervention—although believing they subsequently had a greater understanding of the level required of them, they also indicated ambivalence about their ability to achieve the standard.

Conclusion

The continued emphasis on explicit articulation of assessment criteria and standards is not sufficient to develop a shared understanding of 'useful knowledge' between staff and students. Socialisation processes are necessary for tacit knowledge transfer to occur. The traditional methods of knowledge transfer in higher education placed reliance on complex socialisation processes based on practice, imitation, feedback and discussion, often on a one-to-one or small group basis. For most institutions, reliance on these resource-intensive methods is difficult, if not impossible, in the context of today's rapid expansion of student numbers and cuts in the unit of resource. It does appear, however, that through a relatively simple intervention incorporating a combination of explicit articulation and socialisation processes a considerable amount may be achieved in developing shared understanding and, consequently, in improving student performance—and that this improvement may last over time and be transferable, albeit possibly only in relatively similar contexts.

Acknowledgements

The authors would like to thank the editor and reviewers for their time and effort, and to acknowledge their debt to Roger Lindsay for his invaluable help and advice on the statistical analysis of the quantitative data.

Notes on Contributors

CHRIS RUST is Head of the Oxford Centre for Staff and Learning Development and Deputy Director of the Human Resource Directorate at Oxford Brookes University. He was Course Leader for the university's initial training course for new teaching staff for six years, and along with eight colleagues he helps to provide both staff and educational development support to the university's academic schools and support directorates. He has researched and published on a range of issues including the experiences of new teachers in HE, the positive effects of supplemental instruction, ways of diversifying assessment, and the effectiveness of workshops as a method of staff development. *Correspondence*: Oxford Brookes University, Oxford OX3 0BP, UK. E-mail: cirust@brookes.ac.uk

MARGARET PRICE holds a National Teaching Fellowship with a project focusing on the development of assessment strategies, and is a Principal Lecturer in Management in the Business School, Oxford Brookes University. She has researched and published on a range of issues related to peer-assisted learning, motivation to learn and assessment processes and standards.

BERRY O'DONOVAN is a Principal Lecturer (Learning and Teaching) at the Business School at Oxford Brookes University. Her research interests over the past few years have focused on knowledge transfer processes, both within higher education and commercial organisations.

REFERENCES

- BAUMARD, P. (1999) *Tacit Knowledge in Organizations* (London, Sage).
- COHEN, D. (1969) *Statistical Power Analysis for the Behavioural Sciences* (New York, Academic Press).
- ERAUT, M. (1994) *Developing Professional Knowledge and Competence* (London, The Falmer Press).
- FORBES, D. A. & SPENCE, J. (1991) An experiment in assessment for a large class, in: R. SMITH (Ed.) *Innovations in Engineering Education* (London, Ellis Horwood).
- GIBBS, G. (1991) Reflection and self-assessment for new lecturers, in: S. BROWN, S. & P. DOVE (Eds) *Self and Peer Assessment (SCED Paper 63)* (Birmingham, Standing Conference on Educational Development).
- GREATOREX, 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* (London, HEQC).
- HUGHES, I.E. (1995) Peer assessment, *Capability*, 30 (1), pp. 39–43.
- 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), pp. 239–254.
- LUCAS, L. & WEBSTER, F. (1998) Maintaining standards in higher education? A case study, in: D. JARY & M. PARKER (Eds) *The New Higher Education: issues and directions for the post-Dearing university* (Staffordshire University Press, Stoke-on-Trent).

- 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).
- NELSON, C. E. (1994) Critical thinking and collaborative learning, in: K. BOSWORTH & S. HAMILTON (Eds) *Collaborative Learning and College Teaching* (San Francisco, CA, Jossey-Bass).
- NEWSTEAD, S. E. & DENNIS, I. (1994) Examiners examined: the reliability of exam marking in psychology, *The Psychologist: Bulletin of the British Psychological Society*, 7, pp. 216–219.
- NONAKA, I. (1991) The knowledge-creating company, *The Harvard Business Review*, November–December, pp. 96–104.
- NONAKA, I., TOYAMA R. & KONNO N. (2002) SECI, *Ba* and Leadership: a unified model of dynamic knowledge creation, in: S. LITTLE, P. QUINTAS & T. RAY (Eds) *Managing Knowledge* (London, Sage).
- O'DONOVAN, B., PRICE, M. & RUST, C. (2001) The Student Experience of Criterion-Referenced Assessment through the Use of a Common Criteria Assessment Grid *Innovations in Learning and Teaching International*, 38 (1), pp. 74–85.
- OTTER, S. (1992) *Learning Outcomes and Higher Education* (Leicester, Unit for Development of Adult Continuing Education).
- POLANYI, M. (1998) The tacit dimension, in: L. PRUSAK (Ed.) *Knowledge in Organization* (Boston, MA, Butterworth Heineman).
- PRICE, M., O'DONOVAN, B. & RUST, C. (2001) Strategies to develop students' understanding of assessment criteria and processes, in: C. RUST (Ed.) *Improving Student Learning: 8 Improving Student Learning Strategically* (Oxford, Oxford Centre for Staff and Learning Development).
- PRICE, M. & RUST, C. (1999) The experience of introducing a common criteria assessment grid across an academic department, *Quality in Higher Education*, 5 (2), pp. 133–144.
- SAMBELL, K. & 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), pp. 391–402.
- SAUNDERS, M. N. K. & DAVIS, S. M. (1998) The use of assessment criteria to ensure consistency of marking: some implications for good practice, *Quality Assurance in Higher Education*, 6 (3), pp. 162–171.
- THOMAS, K. (1990) *Gender and Subject in Higher Education* (Buckingham, SRHE/Open University Press).
- TSOUKAS, H. (1996) The firm as a distributed knowledge system; a constructionist approach, *Strategic Management Journal*, 17 (Winter Special Issue) pp. 11–25.
- WEBSTER, F., PEPPER, D. & JENKINS, A. (2000) Assessing the undergraduate dissertation, *Assessment and Evaluation in Higher Education*, 25 (1), pp. 72–80.

