Using peer- and self-assessment to engage with assessment criteria and learning outcomes: a case study from a course for lecturers

Steve Wilson
Centre for Academic & Professional Development
London Metropolitan University

Key words: *peer-assessment, self-assessment, assessment criteria, groupwork, feedback, validity/reliability (of assessment)*

Introduction

Much recent literature on assessment in higher education (HE) gives strong support to the use of both peer and self-assessment (Boud, 1995; Brown & Knight, 1994; Gibbs, 2001; Brown, Bull & Pendlebury, 1997; Brown & Glasner, 1998; Brown & Dove, 1990). Self-assessment is regarded as a transferable skill and a principal part of the student learning experience.

> "*Students will be expected to practice self evaluation in every area of their lives on graduation and it is a good exercise in self-development to ensure that their abilities are extended.*" (Brown & Knight, 1994)

Peer assessment is the assessment of the work of others with equal status and usually has an element of mutuality. Underpinning a peer-assessment process is the giving and receiving feedback from which continued reflection and perhaps dialogue may continue. Brown *et al* (1997) draw a distinction between 'peer assessment', 'peer marking' - the process by which someone makes an estimate of another's work - and 'peer feedback marking' which involves students deriving criteria, developing a peer assessment form, providing anonymous feedback and assigning a global mark. In the context of this study, the term "peer-assessment" is more closely allied to this concept of "peer feedback marking".

The use of peer and self-assessment carries a number of perceived advantages:

- Students have more ownership of the assessment process (i.e. it is not just being "done" to them);
- It can involve students in devising and understanding assessment criteria and in making judgements;
- It encourages formative assessment – learning through feedback;
- It encourages the reflective student (autonomous learner);
- It has validity - it measures what it is supposed to measure;
- It can emphasise the process not just the product;
- It is expected in working situations;
- It encourages intrinsic rather than extrinsic motivation;
- It challenges the role of the tutor as the sole arbiter of assessment.

Toohey (1996) argues that self (and peer) assessment by students is generally trustworthy provided that the criteria have been made explicit, students have had the opportunity to practice assessment skills, the rating instruments are simple and a second marker is used to moderate the assessment. In essence, if a student

understands the learning requirement, the process is managed appropriately, with opportunities for giving and receiving feedback, then it is likely to be a positive and constructive process for all concerned.

This study explores an approach to involving students (called "participants" on this course) in the assessment process by using parallel processes of self and peer-assessment. The University operates a postgraduate course in learning and teaching for its teaching staff. The Postgraduate Certificate, which is part of a full Masters scheme and is accredited by the UK Institute for Learning and Teaching in HE (ILTHE), contains three modules; one centred on teaching and facilitating learning, one on assessment, the third focused on curriculum evaluation and development. One of the programme aims is for participants to experience and examine a variety teaching, learning and assessment methods.

Case study

The module entitled 'Managing the Assessment Process' focuses on the role that assessment plays within the overall learning process by examining assessment purposes and practices in terms of supporting and enhancing the learning experience of students. Of its two assessment components, one is in the form of a group presentation.

This assessment entails a number of features:

i) Each group is required to identify the assessment criteria by which its presentation will be assessed.
ii) Groups must also decide which of the module learning outcomes are applicable to their presentations.
iii) The 15-minute presentation (with 5 minutes for questions) is tutor, peer and self-assessed using the assessment criteria set by the group and against the selected learning outcomes.
iv) Assessors (tutor, peer and self) recommend an overall mark (percentage) using the grade descriptors provided in the scheme.
v) Assessors provide written feedback to the presenting group in which comments refer to the match with the assessment criteria and learning outcomes, as well as providing general remarks relating to the presentation.
vi) The final percentage is the average of all the recommended marks unless the overall average is sufficiently different to the average of the tutor marks i.e. falls into a different grade category. If this is the case, a further moderation by tutors will determine the final mark. Given that the majority of assessors will be peer participants, this variation occurs only when there is a significant difference between the tutor and peer averages.

The purposes underpinning what may seem like a fairly complex and elaborate process are specific. Creating greater student involvement in and ownership of the assessment process is intended to sharpen their awareness of the importance of learning outcomes and assessment criteria. In recent years, the use of 'learning outcomes' have replaced 'objectives' as the means of identifying the learning that needs to be demonstrated for credit to be awarded. At the same time, if appropriate judgements are to be made, then the criteria by which this occurs need to be "explicit, available, open to interrogation and shared" (Brown & Glasner, 1999:11). The exercise also seeks to provide a process that has a measure of reliability, validity and is transparent.

Presenting groups were self-selecting (group size varied from 2 to 5, most commonly 4) and were formed around a declaration of interest in specific topic areas. Guidance to assist presentation preparation was provided, in particular, in devising appropriate assessment criteria and the selection of learning outcomes. Each group was required to submit a presentation outline early on in the module, and, later, a final summary.

Research Methodology and Findings

The study focused on two cohorts, each comprising twenty-five participants, the majority of whom had little (<3 years) teaching experience in an HE context. The process of collecting and analysing data was undertaken in two parts:

1. A comparison was made of the marks and written feedback produced by each assessing group (tutors, peers, self). Each assessor had anonymously completed a feedback form, which outlined the project title, assessment criteria (with weightings, if deemed applicable) and learning outcomes for the presentation.
2. A follow-up questionnaire sent to all participants four months later to ascertain their experience of the exercise and its usefulness and impact. . The questionnaires were anonymous but where respondents opted to give their names, they could be interviewed, if necessary. Responses were received from 30% of participants, and among one third of these further issues were discussed in short interviews.

Comparison of marks

An analysis of the marks awarded reveals, firstly, some notable differences between tutor and peer assessments. In the first cohort, tutors marks were consistently lower than peers in all but one case (see table 1).
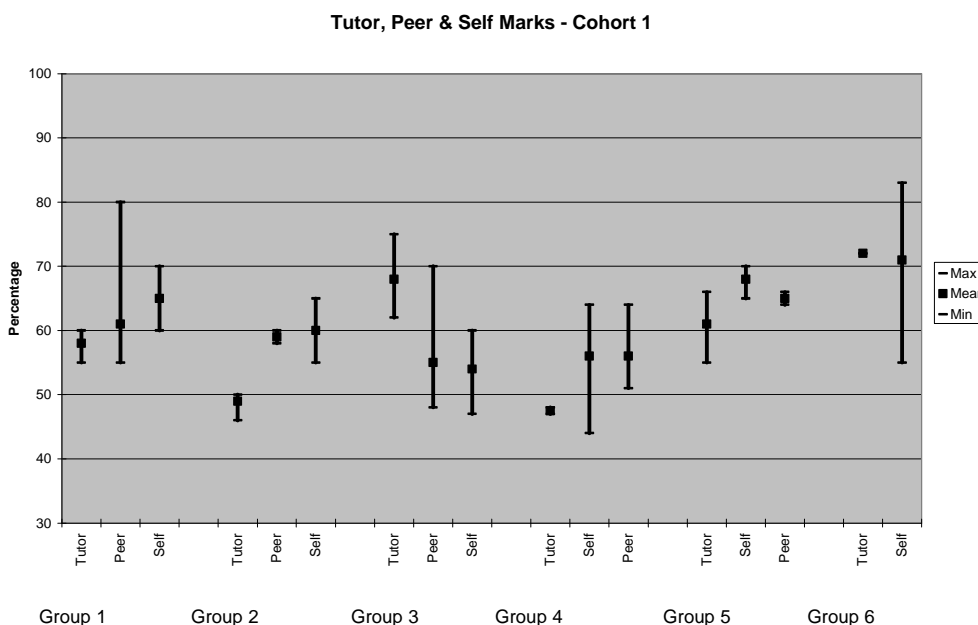

Table 1

Comparisons in cohort 2 show an equal number of tutor averages higher than peer averages as the reverse (see table 2). However, tutor marks display a significantly greater range of marks. This outcome is, in some part, due to the addition of two new course tutors who had not previously undertaken assessment on the programme.
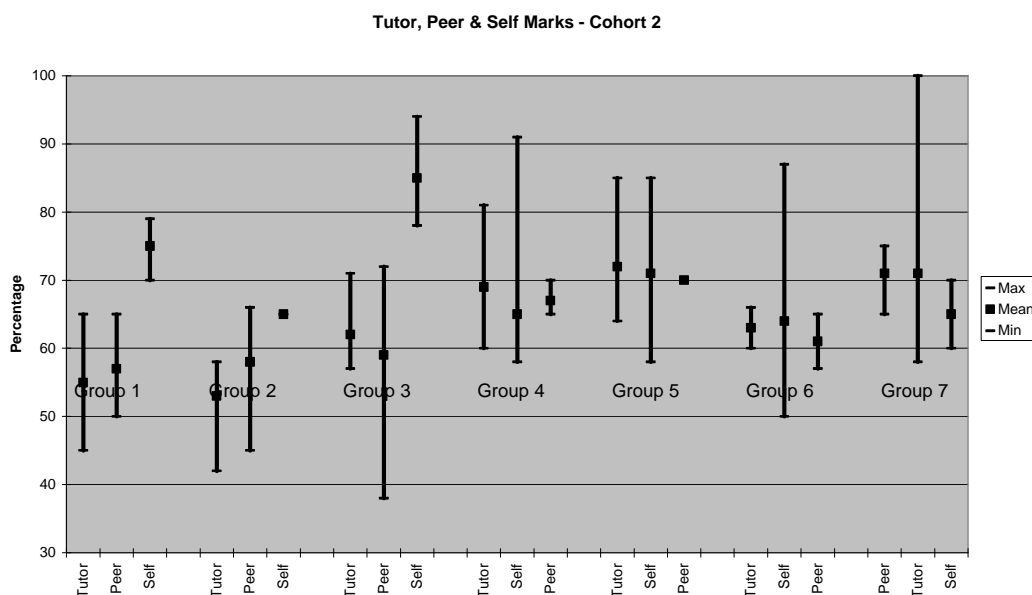
Tutor, Peer & Self Marks - Cohort 2

Percentage

100
90
80
70
60
50
40
30

Group 1   Group 2   Group 3   Group 4   Group 5   Group 6   Group 7

Tutor Peer Self Tutor Peer Self Tutor Peer Self Tutor Self Peer Tutor Self Peer Tutor Self Peer Peer Tutor Self

Max
Mean
Min

Table 2

The peer averages might appear to reflect a regression to the mean that can occur in a larger group size. Closer scrutiny, however, demonstrates an increased range in the marks awarded by peers in cohort two. Significantly, the preparation of these participants included greater attention to the grade descriptors for the scheme. Although few peer marks fell below the pass mark (40%), peer-assessors seemed more comfortable in using the full marking range.

It is perhaps understandable that the range of marks from self-assessors is less. Questionnaire and interview responses indicated that the process of self-assessment was felt to be less problematic than peer-assessment. While self-assessors often commented on how they had "run out of time" or not presented points clearly enough, because considerable time and thought had been invested in the preparation, the presenting group already had an "inner feeling" of the worth of their presentation. It was just a matter of delivering the material and "releasing the potential". This may provide some explanation as to why average self-assessment marks were consistently higher than average tutor marks, and often the average peer mark. Apart from one group, the difference between the tutor and self-assessment in cohort 2 is significantly less. Again, the additional attention given to grade criteria that occurred with this cohort may be a factor.

A comparison of the average peer and self-assessment marks reveals some unanimity in cohort 1, but in cohort 2 the self-assessment marks are higher. Participant feedback highlighted the challenge that peer assessments had to be undertaken "at short notice" and using assessment criteria derived by the presenting group. Interpreting the criteria within the time-pressured context of the presentation was problematic, a comment also made by tutors.

Some presenting groups decided to provide a weighting for each criterion, which was deemed helpful by some peer assessors. However, it also proved to be a double-edged sword when criteria given a high weighting were adjudged to be less fulfilled in the presentation or others weighted lower were strongly evident but could not be given any more credit.

Analysis of participants' comments

Analysis of the responses from the questionnaires and interviews revealed the following:

i)   *Negotiating and agreeing assessment criteria*

Both cohorts referred to the increased appreciation of the importance of the role assessment criteria, which were deemed crucial for assessing presentations objectively and fairly. Setting one's own criteria, with guidance and tutor approval, did increase a sense of validity.

Assuring reliability was perceived as more problematic by some participants. Interpreting the assessment criteria, making judgements using the grade criteria (used throughout the course), doing so "on the spot" and to "co-colleagues" were factors identified. However, reliability was thought to be largely assured through use of the guidance provided within the module and the requirement for tutor ratification prior to the group presentation. This supports arguments made by Toohey (1996) about proper assessment management, as outlined earlier.

ii) *Identifying applicable learning outcomes*

Responses indicate that the determination of applicable learning outcomes was undertaken in group discussions, making use of tutor and written guidance available. Debates within groups centred on whether the pre-determined learning outcomes would be addressed sufficiently. Consequently, groups omitted some learning outcomes because these were deemed to be addressed only tangentially. However, it appears that presentations were also modified with the intention of better matching the learning outcomes. Paying attention to learning outcomes at this juncture had a beneficial effect on the planning and delivery of the presentation, as is borne out in questionnaire and interview feedback. Many respondents noted, in hindsight, how it had helped them to focus on what was central to the module.

> "*I found this to be valuable in that assessing other groups and our own group in the extent to which they/we met the learning outcomes provided me with a greater awareness and understanding of the learning outcomes. In turn this helped our group to focus more on trying to meet the learning outcomes when doing our group presentation.*" (Cohort 1 participant)

Matching the presentation to the learning outcomes did cause some difficulties. For some participants, learning outcomes were too vague and required greater clarity before they could be effectively matched. This provides a salutary reminder to ensure that learning outcomes are produced with the intended students in mind rather than other readers (e.g. QAA, course approval boards, fellow lecturers).

iii) *Usefulness of feedback*

In conventional tutor-led assessment, feedback, however comprehensive, mostly is in the hands of one or two people. This tutor, peer and self-assessment process provided in excess of twenty five sets of feedback. The sheer volume of response is considerably more than a student might normally expect to receive. Furthermore, respondents were overwhelmingly struck by the range of feedback received and of the informative opportunities it provided. Only one participant felt

the self-assessment comments to be of little use. For a few, the discrepancy and apparent contradictions in some feedback made it difficult to determine what was appropriate. However, for a significant majority this variation, combining a mix of summative comment and formative advice, was welcomed.

iv) Continuing views on peer and self-assessment

It is evident that peer and self-assessment were two forms of assessment that, hitherto, had been little used by most participants and certainly in the formal assessment of their own work. Creating an opportunity to "try on" these ideas in an arena "that counts" has, for many, provided a genuine reflective opportunity for enhancing their own understanding and to apply it in their own teaching.

Some concerns were also expressed. The preparation time required to induct and prepare participants was noted. Other research (Brown & Knight 1994, Jordan 1999, Roach 1999) also comment on this. However, quantity of time needed should not be confused with the quality of the experience. If students (participants) are to engage in the process of peer and self-assessment and to appreciate fully the interrelationship of assessment criteria and learning outcomes, then adequate time and preparation is needed for what should be highly transferable learning.

Participant concerns also focused on potential bias or lack of objectivity that might occur through a peer or self-assessment process. Similar arguments have been made in other research (Brown, Bull & Pendlebury, 1997; Falchikov, 1986, 1996; Boud *et al*, 1997). It is acknowledged that these participant groups were mostly colleagues working in a supportive environment. Questionnaire responses referred to "giving the benefit of the doubt". Others commented on being influenced more by the quality of the presentation rather than by its content. As previously stated, peer- and self-assessments were overall higher than the tutor averages. However, there were some notable exceptions and overall differences were not always significant. Students can be highly critical of themselves and their peers and this was reflected in some quite frank remarks (both positive and negative) in written feedback.

Conclusion

Much is postulated about the advantages and disadvantages of using self and peer-assessment instead of or as an adjunct to tutor assessment. For Race (2001:5):

- Students are already self-assessing and peer-assessing quite naturally;
- Tutor assessment is not sufficiently valid, reliable or transparent;
- Peer and self-assessment lets students into the assessment culture;
- The process deepens students' learning experiences;
- Students gain much more feedback than would otherwise be possible;
- It helps students become autonomous learners.

Outcomes from this study lend support to many of these claims. Participants initially entered into the assessment process because it was required of them. Experiencing different assessment mechanisms was an intended goal of the programme. Drawing attention to the match with the learning outcomes was meant to highlight their role in curriculum design and delivery. Requiring groups to devise their own assessment criteria was aimed at exploring their use and importance within the assessment process. Feedback from participants confirms that these objectives were achieved,

and that the exercise has had a continued impact on their thinking and practice as teachers in HE, as illustrated in the following observation:

"My positive experience of peer assessment and self-assessment in the MAP module has encouraged me to experiment with the use peer assessment in my own teaching. I have implemented peer assessment as part of the assessment process in a third year undergraduate module I teach. Student feedback suggests that students, in the majority, have also found this to be a positive and valuable experience in enhancing their own learning." (Cohort 1 participant)

References

Biggs, J. (1999) *Teaching for Quality Learning at University.* Buckingham: Open University Press/SRHE.

Boud, D. (1995) *Enhancing Learning through Self- Assessment.* London: Kogan Page.

Boud D, Anderson G, Cohen R, and Sampson J (1997) *Developing Assessment for Peer Learning*, Research and Development on Higher Education, 20, p. 117-125

Brew A (1999) *Towards Autonomous Assessment: Using Self-Assessment and Peer Assessment,* in Brown, S. and Glasner, A. (eds.) (1999) Assessment Matters in Higher Education - Choosing and Using Diverse Approaches. Buckingham: OUP.

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

Brown, S and Dove, P (Ed) (1990) *Self- and Peer-Assessment* SCED

Brown, S and Knight P (1994) *Assessing Learners in Higher Education* Kogan Page

Brown, S. and Glasner, A. (eds.) (1999) *Assessment Matters in Higher Education - Choosing and Using Diverse Approaches.* Buckingham: OUP.

Brown, S., Rust, C. and Gibbs, G. (1994) *Strategies for Diversifying Assessment in Higher Education.* Oxford Brookes University: Oxford Centre for Staff Development.

Falchikov, N (1996) Involving Students in Feedback and Assessment in Hounsell D (ed) (1996) Rethinking Assessment, Kogan Page.

Falchikov, N (1986) *Product comparisons and process benefits of collaborative peer group and self assessments*, Assessment and Evaluation in Higher Education Vol 11 No 1, p146-165

Fry, H., Ketteridge, S. and Marshall, S. (1999) *A Handbook for Teaching and Learning in Higher Education - Enhancing Academic Practice.* London: Kogan Page.

Gibbs, G. (1992) *Assessing more students.* Oxford Brookes University: Oxford Centre for Staff Development.

Gibbs, G. (1999) Using assessment strategically to change the way students learn. In: S. Brown and A. Glasner (eds.) *Assessment matters in higher education-choosing and using diverse approaches.* Buckingham: Open University Press.

Heywood (2000), Assessment in Higher Education: Student Learning, Teaching, Programmes and Institutions, Jessica Kingsley Publishers, London

Jordan, S (1999), *Self-Assessment and Peer Assessment*, in Brown, S. & Glasner, A. (eds.) Assessment Matters in Higher Education - Choosing and Using Diverse Approaches, Buckingham: OUP

Miller A, Imrie, B & Cox K, (1998), *Student Assessment in Higher Education, A Handbook for Assessing Performance*, Kogan Page

Race, P (2001) *LTSN Assessment Series No 9: A Briefing on Self, Peer and Group Assessment*, LTSN Generic Centre

Roach, P (1999), *Using Peer assessment and Self-Assessment for the First Time*, in Brown, S. & Glasner, A. (eds.) Assessment Matters in Higher Education - Choosing and Using Diverse Approaches, Buckingham: OUP

Biographical note

Steve Wilson is the Head of the Centre for Academic Professional Development at London Metropolitan University. Having started academic life as a mathematics teacher, he has worked in higher education since 1987, the past ten years within academic and professional development, including developing and teaching on masters degrees in learning and teaching at two universities. His research and professional interests concern all matters related to teaching, learning and curriculum development and in particular, assessment. Steve is an experienced subject reviewer. [email: s.wilson@londonmet.ac.uk]