Student experiences of peer review marking of team projects

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

© Higher Education Academy

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Pond, Keith, Dave Coates, and Ofelia A. Palermo. 2019. "Student Experiences of Peer Review Marking of Team Projects". figshare. https://hdl.handle.net/2134/3040.

STUDENT EXPERIENCES OF PEER REVIEW MARKING OF TEAM PROJECTS

by

Keith Pond, Dave Coates and Ofelia A Palermo

The Business School, Loughborough University

Corresponding author:

Dr Keith Pond

The Business School

Loughborough University

LOUGHBOROUGH

Leics.

LE11 3TU

Tel: 10509 223290

Fax: 01509 223962

E.Mail: K.Pond@lboro.ac.uk

1

STUDENT EXPERIENCES OF PEER REVIEW MARKING OF TEAM PROJECTS

ABSTRACT

Peer review or peer assessment of students has been used in many places to motivate and focus students on their own development within a group work setting. Group work has its benefits for students as it allows many transferable skills to be developed. It also has benefits for tutors as it promises lower assessment burdens. However, critics of group work cite drawbacks including "free riders" and de-motivation of higher achieving students where their additional efforts are not rewarded. Peer review attempts to remedy such problems.

Pressures of time, budget and student numbers often obviate deeper study of alternative but beneficial assessment techniques. This can preclude their deployment or marginalize efforts to discuss and analyse their effectiveness when they are used. The study reported in this paper, made possible by a "Small Grant to make a Difference" from the HEA, allowed for analysis of a peer review mechanism that had been operating for a number of years and a new web based peer review data capture system from the student perspective.

The paper describes a specific module where peer review has been deployed and reviews the extant literature on peer review systems, paying particular attention to criticisms of such assessment techniques. The paper goes on to outline a research methodology whereby student perspectives and experiences of peer review were collected. The outcomes of the focus group methodology

are then discussed alongside a brief analysis of quantitative data from the peer review systems used.

Key conclusions from this research are that the method of data collection (paper based vs. web based) made no significant difference to the generally positive student experiences of the peer review concept. In addition peer review marks are not significantly affected by the data collection method either. Whilst much of the data collected updates, confirms and strengthens previous literature on this subject important new insights are gained into the emotional perspective of students, their desire to explain their marking of peers and their marking behaviours. The findings from this research are already being used to aid development of the web based data collection tool and to establish "good practice" guidance on the deployment of this valuable and innovative assessment technique.

Abstract 350 words

STUDENT EXPERIENCES OF PEER REVIEW MARKING OF TEAM PROJECTS

Introduction

This paper focuses on the student experience of peer review of team projects in a first year undergraduate module at Loughborough University Business School. Our peer review mechanism has been used for a number of years (Pond, ul-Haq and Wade, 1995) to help discriminate between team members' performance and has evolved from a paper based system of review following the design suggested by Goldfinch and Raeside (1990), used annually since 1998, to a web based mechanism (Willmot and Crawford, 2004) used for the first time in 2004/05.

The paper describes the specific module where peer review has been deployed and reviews the extant literature on peer review systems, paying particular attention to criticisms of such assessment techniques. The paper goes on to outline a research methodology whereby student perspectives and experiences of peer review were collected. The outcomes of the focus group methodology are then discussed alongside a brief analysis of quantitative data from the peer review systems used.

Background and deployment of peer review

The research took the opportunity to review student experiences before and after a switch from a paper based peer review to a web based data collection tool. The paper based data collection method took place at the end of the module and under near exam conditions (in an attempt to avoid collusion whereby students

4

agree marks for peers in advance) whilst the web based system was available to students registered on the module for a period of one week immediately after the end of the module and in a completely unsupervised atmosphere.

The module on which the observations in the research were based is called Personal Effectiveness. This is a core first year undergraduate module offered to all Business School students at Loughborough. Loughborough's five vocationally oriented business degrees all develop from a central core of common modules (Economics, IT and Quantitative studies, Accounting & Financial Management and Organisational Behaviour), many of these being delivered in a traditional lecture based way. Growth in student numbers in recent years also mean that many of these modules are delivered to class sizes of up to 300 students.

Whilst being resource efficient, the lecture based approach can nonetheless fail to overtly address the necessary development of management skills needed in today's workplace and classroom. Co-operative team coursework that students must research, present orally and in written format and physically deliver has the potential to allow these key competencies to develop, influence students' awareness of such skills and give students ownership of their experience and learning. The Personal Effectiveness module seeks to address these issues.

The Personal Effectiveness module fully recognises the need to influence students' awareness of and performance in key competencies in Business, Management and Accounting careers (BEST, 2004). The module welcomes invited industry and internal speakers (including Library, Professional Placements

and study skills) providing a forum for the introduction and discussion of key competencies. This approach also helps to contextualise the module. The module uses formative assessment of group coursework supplemented by individual reflection. The group coursework is based around a business related project, completion of which exposes students to and engages students in many key competencies. The project outcomes are assessed by a tutor. The final competency practised by students on the module is their judgement of others' contribution to their team project. The resultant peer review is then used to moderate the tutor awarded marks (as described in Pond and ul-Haq, 1997 and Willmot & Crawford, 2004 and 2005)

The Personal Effectiveness module provides "authentic learning" (Dochy et.al., 1999) as it combines development of cognitive and meta-cognitive competencies such as analysis and self-reflection within a team based project activity that adds social competency enhancement.

Our use of peer review in student assessment

Peer review focuses mainly on the learning process to achieve differentiated marks for individual students within project groups. It also focuses on the product and contributes not only to teamwork skills but also to a student's own self-evaluation. Employability criteria (see BEST, 2004, for example) stress the need for teamwork skills, judgement, influencing skills and leadership, all of which are affected by an honest and objective peer review process. Our use of peer review in the Personal Effectiveness module attempts to encourage participation in teamwork, improve group dynamics (Brown and Pendlebury, 1992), develop

reflective skills and higher levels of thinking (Falchikov, 1988) and provide more student-centred learning.

Although an anticipated outcome of peer review is to develop self and peer assessment skills this aspect is not marked or rewarded as part of the assessed work. Davies (2005) argues that a "mark for marking" approach should be considered within peer assessment although we have not included this feature in our model.

In our module the concept of peer review is introduced to students at the beginning of the module when sufficient detail is made available to ensure good dissemination of the criteria by which students will be asked to assess themselves and each other. Table 1 summarises the criteria used in the Personal Effectiveness module throughout the period of the study.

TABLE 1 ABOUT HERE

The design of the criteria acknowledge that student ownership of criteria and assistance in their design is beneficial (Pond and ul-Haq, 1998). However, with an annual student cohort of 300 the generation of unique criteria each year through discussion with students was felt to be logistically too challenging. In addition, observation and informal feedback indicated that there was no significant misunderstanding of criteria. It should be noted that the criteria used (see Table 1) were originally generated by students at a time when cohort numbers were smaller.

An explanation of the impact of peer review on allocated marks forms part of a lecture towards the end of the module and brief details of this are also available throughout the module on the intranet VLE that supports the module. The web based system has since been improved to incorporate a virtual tour of the data collection process to allow familiarisation with the software prior to use. Following submission of the group coursework, but before marks are known, students are required to use the web based system to review the performance of each member of their project group.

Student interaction with the web based system is regulated by general ID and password access and is synchronised with other University systems. Although students must identify other members of their group when making judgements all other aspects of the process retain anonymity. Students do not know from the web based system which of their colleagues awarded particular marks but (at the end of the process) become aware of their general position in the group by the impact that peer review has on their overall mark. Later in this paper a textual feedback procedure will be discussed but at the time of this project it was not in place.

Peer Review foundations

Peer review is not without its problems, however. Brown and Dove (1991) suggest that, in comparison to traditional assessment methods peer assessment can be too demanding of students, too time consuming and criteria setting can be problematic (see also Orsmond et al., 1996). It should be noted that Brown and

Dove were concerned with student assessment of others' work rather than our assessment model styled "peer review". In our scheme the criteria are tutor set and have been generated over time through interaction with the literature on group dynamics and the students themselves. Whilst most authors who have reported on peer assessment note general student acceptance of the methodology some question whether students have an appropriate understanding of individual assessment criteria (Lin, 2002).

Tutor reluctance to embark on peer based assessment can be because of fears of unreliability. Brown and Knight, (1994) suggest that student behaviours can make the validity of mark allocations questionable. Issues of friendship marking and collusion can result in over-marking (Pond, ul-Haq and Wade, 1995). Free riders or "parasites" can also find safety as students are often unwilling to "blow the whistle" on colleagues.

Other issues with peer assessment and review revolve around the perceived "fairness" of the system (Conway et al., 1993), acceptability to students since assessment is the tutor's job (Lin et al., 2002; Venables and Summit, 2003) and gender effects. Gatfield (1999) could detect no gender effects in Australian undergraduates whilst Menchaca et al. (2002) saw female scoring remaining high over time in Texas and male scoring falling. Menchaca et al. (2002) also noted a "training" effect as students were exposed to peer assessment on a number of occasions. Over time students became more comfortable and familiar with the practice of peer review and their self-assessment ability can improve (Griffee, 1995).

Where self-marking is used as part of the peer review, exaggeration of an individual's contribution to the group effort can also be seen (Lejk, 2002 and Willmot & Crawford, 2004). Whilst the impact of this can be moderated in larger groups it can still have an effect on an individual's mark. Davies (2006) noted that better students are more likely to be more critical of peers (under marking) whilst weaker ones less so. This is consistent with Boud and Falchicov (1989) who also noted that better students underrate themselves.

The successful deployment of any peer review mechanism will often depend on the purpose for which it is deployed and the clear links to the intended learning outcomes for the module. Lin (2002) underlines the importance of defining the issues, objectives and assessment criteria clearly. Moreira (2003) notes that the communication of the criteria or judging parameters must be very clear. The purposes for which peer review can be used and that are associated with the development of transferable skills include:

- Supporting group interaction
- Supporting group, self and peer reflection
- Aiding development of the skill of judgement
- Deeper learning

Without support for students in understanding the process of peer review deployment may not be effective (Conway and Kember, 1993). Conway and Kember (1993) also recognised that whilst students felt the allocation by a tutor of

a group mark to the whole group, without moderating for individual effort, was "unfair" there was "unfairness" and potential bias in any peer review system too. The sources of bias emerge from the question of ability of students to assess others, their reluctance to "mark down" (providing positive marks was felt to be easier) and the opacity of final grades. This final point relates to the need for students to understand how the peer review has impacted their individual mark.

Peer review can also be used to make mark allocation follow a more "normal" distribution in group work situations. Bunched group marks often show a low standard deviation and the use of peer review can help to spread this when marks are reviewed at the individual level. The administration of peer review probably does nothing to make the marking burden easier for the tutor, however, but can reduce transaction costs (Alexander and McKenzie, 1998). Our use of both paper based and web based data collection systems over time has seen the administrative effort shift from the end of the process (paper based) to the start (web based) but has done little to reduce the overall time spent on the administrative burden. A challenge for the web programmers is to reduce this burden.

Finally, in this section, the question of anonymity is raised by a number of authors (see Davies, 2000; Willmot and Crawford, 2004; Lin, 2002). From the student perspective almost all peer review systems studied for the purposes of this research were designed to retain anonymity for participating.

The review of key literature in this field has been very fruitful in revealing research and observations that combine to establish good practice in this area. If peer review is to work well and deliver benefits to students and to assessors in terms of learning outcomes, skill development and fairer marking the design of the scheme to be used must be considered very carefully. No peer review system can be perfect but it can certainly provide additional benefits to wholly tutor based marking, balancing these with additional transaction costs.

Project methodology

The overall aim of the project funded by the HEA grant was to provide a study of student experiences, acceptance and usage of paper based and web based peer review systems for group work in a Business School. The methodology used in the qualitative research element of the project was designed to reveal objective views of peer review from past users of both systems. The quantitative analysis of the datasets created by peer review use over a period of six years sought to reveal differences in marking behaviours and outcomes.

Qualitative research

Since an instrumental approach, such as a survey, might stifle objectivity and openness amongst respondents the key methodology for the qualitative research was the use of focus groups. Focus groups would allow informality of the setting (a staff lounge in the Business School), the provision of refreshments and the employment of a group facilitator, who had not been part of the original assessment team, to maximise objectivity.

Using focus groups is one of many tools available for carrying out social research. Merton (1956) was the first to use this tool in his research during the Second World War and he coined the phrase 'the focused interview' (Stewart and Shamdasani, 1990). In the time since Merton's pioneering work, focus groups have become an increasingly frequent research tool for applied social scientists. In the focus group, group interaction is employed to generate data and as a source of data for analysis (Morgan, 1988 and 1993). Market researchers have employed focus groups since the 1950s and during the 1980s there was a resurgence of interest in this methodology from social scientists (Catterall and Maclaran, 1997).

Although this research used both qualitative and quantitative analysis techniques there are important elements that distinguish focus groups from quantitative survey methodology (Grudents-Schuck et al., 2004). These two methodologies are different in their purposes, procedures, and results. In particular, focus groups differ from survey methods because they give insights rather than rules or measurements. This means that individuals who participate in focus group sessions are not constrained in their responses in the same way as respondents to a typical survey questionnaire. Participants are generally allowed to say anything they would like in focus group sessions. The group facilitator listens not only for the content of focus group discussions, but also for emotions, ironies, contradictions, and tensions. This enables the researcher to learn or confirm not only the facts (as in survey method) but also the meaning behind the facts.

Focus groups also differ from the survey method because they are social and not individual. Conversation among participants during a focus group session, results in verbal data. In this way, information that paints a portrait of a combined local perspective is elicited (Grudents-Schuck et al., 2004).

Procedurally, focus groups differ from surveys as participants should be homogeneous and not diverse i.e. participants have similar characteristics. If focus groups were made up by individuals with heterogeneous personal characteristics, status or education an individual could tend to censor ideas because of the perception of difference. Thus, multiple groups need to be conducted in order to get a cross section of views from a diverse population. Another aspect that differentiates focus groups from surveys is flexibility. In focus groups questions flow from the general to the specific and focus groups invite openness and avoid bias. Finally, in terms of data analysis and reporting, focus groups rely upon words spoken by participants. In general, the elements that are highlighted were common to a number of groups. Our focus groups sessions were tape recorded (no participants objected to this), with tapes being the key source of transcription notes, supported by the facilitator's contemporaneous notes.`

Focus groups are fundamentally a way of listening to people and learning from them. They create lines of communication. This is most obvious within the group itself, where there is continual communication between the facilitator and the participants, as well as between the participants. Facilitators should be motivated

to listen and learn from the participants focusing, when needed, on the topics of particular relevance to the research.

Despite the focus group aim to highlight the meaning behind peoples' opinions, this methodology is not a totally reliable technique for determining an individual's authentic point of view because of the influence of social norms in the group setting. Further limitations of this methodology are:

1. Generalisation of results of the focus group to a larger population is difficult because of the small number of respondents that participate even with several different focus groups.
2. The convenience nature of most focus group recruiting practices also makes generalisation difficult.
3. Respondents are not independent from one another and dominant or opinionated members may bias the results obtained because of the intense interaction that happens during sessions.
4. The researcher may place greater faith in the findings than is actually warranted because of the live and immediate nature of the interaction that takes place.
5. Interpretations of results may be difficult because of the open-ended nature of responses obtained; and
6. Results may be biased by the moderator not providing cues about what types of responses and answers are desirable.

Because of these limitations, focus groups are used most often as a preliminary stage in a larger research program that includes a larger, more representative survey of the population, or as a means for adding insight to the results obtained from a survey.

Follow-up research, funded by a grant from the Engineering CETL at Loughborough, has been designed to extend this work by a survey methodology.

Table 2 summarises the ways in which the above limitations were tackled in the current research:

TABLE 2 ABOUT HERE

Project focus groups

In the project we ran four focus groups composed of three to nine students each in order to collect qualitative insights on the main issues. Each focus group was homogeneous as each comprised either second year students or final year students at Loughborough University Business School. The second year students had experienced a web based peer review mechanism during their first year studies. The final year students had experienced a paper based peer review system during their first year studies. It was expected that the final year students would reflect more objectively with a longer time perspective and also because they had had a year-long professional placement as part of their studies. One consequence of the placement experience is that they would be more used to appraisal systems, including self-appraisal.

The main issues on which students had to express their opinions were drawn from the key concerns and criticisms of peer review systems recorded in the literature review section of this paper (above). The issues were:

- The ease of use of the Paper based or Web-PA peer review system;

- The understanding of what is being achieved through the peer review;

- The key benefits of the system to users; and

- Problems and hazards of the system.

Each focus group began with a brief reminder of the purpose of the group and circulation of an informed consent form that asked for consent to tape recording of the process. No students withdrew at this stage. The focus groups lasted between 25 and 40 minutes with the facilitator giving plenty of opportunity for everyone present to express their opinion.

Findings from the focus groups

The overall reaction to the opportunity that the focus groups gave to the students was very positive and the students engaged with the issues enthusiastically. One measure of this is recorded below in the discussions on improving the web based system in particular. Many comments garnered in this research have been passed to the web design team so that improvements to the software can be made. Summarising the transcriptions of the focus groups reveals the following observations and comments relating to the key issues:

Finding 1 - Ease of use

Participants agreed that both systems used were easy to understand and provided no problems in understanding how to mark other group members and themselves. It emerged from all focus groups that peer review would be more complete if students could explain the reasons why a certain mark is attributed. Although the general perception was that the marking criteria in the peer review were in tune with the ones used by the lecturer in marking the groups, the students were unhappy that the opportunity to leave textual comments to explain marks was missing. In a previous use of the paper based peer review system on a separate module this opportunity had been given where either bottom (0) or top (3) marks were given. No student had availed themselves of this opportunity and so the practice was abandoned. Perhaps it would be instructive to re-introduce it for a trial period.

Finding 2 - Understanding the purpose

Participants highlighted the point that peer review is a way to make people reflect when marking others; it is an activity that also involves emotions. One common emotion was the stress caused by the dilemma of truthful and objective marking versus the loyalty owed to group members. This is also echoed in the difficulty in marking an individual down, although complete free riders will be identified.

Some students considered how peer review would enhance deeper reflection if it were possible to receive feedback on the reviews from the group, even anonymously. A later version of the web based system has begun to offer feedback but deliberately stops short of actually advising peer marks.

Finding 3 - Benefits and hazards of peer review

When focusing on benefits participants underlined many aspects including the possibility to roughly differentiate between group members according to the level of their contribution and the possibility to learn skills for working in groups in the real world, the possibility to prevent (or even punish) "free riding". In addition it was felt that people who know that they will be peer assessed tend to be more motivated to work. Here it must be acknowledged that the self-selecting nature of the focus groups may well introduce bias from the perspective of the better motivated and engaged student.

Anonymity was also felt to be a major advantage of the systems used as this allowed the possibility to express judgements without being intimidated by those with strong personalities.

Participants also highlighted some hazards of peer review systems in general. The extreme subjectivity that students have in marking their friends and the influence of personal dislike on marking others was seen as a negative point. To reduce these problems, participants suggested that students need to justify their marks and say what they have done in order to address possible conflicts, perhaps by adding textual comments. There was no evidence from the focus groups relating to gender bias.

Finding 4 - Additional areas of discussion

Apart from the issues on which the researchers wanted to focus attention, the discussions within the focus groups addressed several other topics. These other topics related to ideas for improving peer review in general terms.

One popular idea was to use peer review in all types of modules that require group work rather than in isolation on individual modules. It was felt that, in order to develop self-evaluation, objectivity and skills for working in teams in the real world, students would need to be exposed to peer review mechanisms more frequently. A further thought related to the lecturer's explanation of the importance of peer review at the beginning of every module in order to shift some responsibility for learning outcomes onto students. A third point was that peer review could be used more than once during a module in order to influence behaviour during a module rather than simply reporting on behaviour after the module had finished. Each of these ideas would need consistently formed feedback following the peer review in order to make sense of it and not leave it as something to address in the overall grade achieved after the peer review moderation.

Secondly, feedback would, it was felt, help personal tutors support students in overcoming the problems that resulted in lower peer marks. It was also felt that selective use of textual comments from students on their reasons for low (or high) marks would also assist in this process. Clearly the narrative associated with each peer review criterion was felt to be insufficient to provide a useful

development tool for students and further thought will need to be given to this aspect.

Thirdly, focus group participants noted that all of the groups that they were allocated to in their first year studies were selected by the tutor, therefore they had no possibility to choose their friends. Friendship groups, it was acknowledged, would decrease the discrimination between peers. Some participants highlighted that peer review negatively affected their personal relationships with other members of the group, thus disrupting group harmony.

Isolated points from the focus groups included the tendency to "cheat" in marking by some students although it was not clear what form the "cheating" took. However, group collusion was sometimes apparent as certain groups agreed on how to mark each other (either equally or selectively by identifying one group member whom all others would mark down). The focus groups, themselves, were unable to assess the impact that collusion may have had and so this aspect will be considered by quantitative analysis of the datasets generated by the paper and web based systems over a period of six years.

Quantitative Data analysis

The literature review and the focus groups have identified and discussed three important themes. Firstly, is peer review too demanding for students? Secondly, what level of "collusion" can be expected in the peer review process? And thirdly, is peer review "fair"? The project was fortunate in being able to access and analyse several databases of student generated peer review marks alongside the

focus group methodology. The datasets reviewed were from the six academic years from 1999/2000 to 2004/2005. In the first five years the paper based data collection method was used and in the final year the web based collection tool was used. In all cases the criteria used by students were identical, group sizes were roughly equivalent (five to seven students) and the project tasks undertaken similar.

The resource limitations of the data collection methods meant that the data from the web based collection method was far richer and more detailed that the paper based method. The web based method captured every keystroke and individual judgement by students whilst the paper based method could only use summary marks for each student. Some sampling of individual student marking was undertaken, however, on the paper based datasets in order to test the observations from the richer web based data. Where more detailed data has been used this is indicated.

Quantitative research findings:

Finding 1 – not too demanding

It has already been noted under the findings from the focus groups that students wanted peer review in all modules that require group work and hence that students do not consider peer review to be too demanding. Further (practical) evidence of this comes from the fact that in the datasets there were very few instances of very low or zero standard deviations (all members of a group giving

the same mark). If the peer review process was "too demanding", then many instances of very low or zero standard deviations might be expected because students felt that the process was too hard for them.

It is true that very low or zero standard deviations could indicate collusion or laziness as well a reluctance to engage with the peer review process ("too demanding"). Without further survey based research it is impossible to determine the cause of very low or zero standard deviations in any specific case and impossible to quantify the proportions due to any of the possible reasons. However, there is anecdotal evidence that the incidence of very low or zero standard deviations is more marked where students are allowed to select their own group members (often resulting in "friendship" groups); where the opportunities for collusion are greater (for example where group sizes are smaller, or where students know each other better such as with "friendship" groups and final year students) and where cultural norms prevent objectivity in peer judgements (groups of far eastern students can generate this observation).

Finding 2 – level of collusion

Continuing to consider the standard deviations of peer review marks within a group suggests that there is little collusion and that students primarily assess their peers by making independent, individual decisions. The evidence for this is that standard deviations for individuals within groups were found to be significantly bigger than standard deviations of the cohorts taken as a whole. If there was considerable collusion, so that students were making very dependent, collaborative decisions, then the standard deviations for individuals within groups

would be very small and certainly smaller than the standard deviations of the cohorts taken as a whole

It is also the case that standard deviations within peer group marks are influenced by group size, the larger the group, the larger the variation in marks. This seems intuitively reasonable, since the larger the group the greater the chance of a particularly strong or (especially) particularly weak member of the group. In particular, with a larger group a student may feel there is more chance to "hide" and that their absence will not be noticed (at least in the short term). This supports the conclusion that there is little collusion; if there was substantial collusion then the standard deviations would be largely independent of group size.

Finding 3 – issues of fairness

There are four aspects of the quantitative analysis that suggest that marks obtained from the peer review process are "fair". Firstly, individuals do not always exaggerate their own mark. About 13% of students give themselves lower scores than their classmates. Further research is needed to investigate why this occurs and whether, for example, there is a link between gender or ethnicity and a student giving themselves a lower score. The observation that students do not always exaggerate their own performance has previously been commented on, for example by Falchikov and Boud (1989).

Secondly, at the level of the individual the average student gives themselves about 10% more marks than their class mates give them. If peer review marks

are simply obtained by averaging all marks within a group then larger group sizes will, of course, reduce the impact of this on the final mark awarded. However, even for small groups, the small size of this average would suggest that the mark impact for an "honest" self-marker is not very large. Further, in our systems, peer marks are normalised prior to generation of the peer review factor that will be applied to assessed marks and so groups that offer all members an "above average" ranking will not benefit. In other words there is a fairness to marks obtained from our peer review process, the marks for an individual student are not "exaggerated".

Thirdly, it is true that at the level of the individual there were some who hugely inflated their own marks. Although these represented a minority of students and groups (2% of students in 2002/03; 5.9% in 2003/04 and 1.7% in 2004/05) the marks they gave themselves were around 50% more than the average of their class mates gave them. Willmot and Crawford (2004) suggest that this is observed most readily amongst either weaker or stronger students as opposed to those of average ability although our research could not confirm this.

Finally, closer analysis was made of the 2003/04 dataset because this was the dataset that related to some of the focus group participants. This close analysis of the 2003/04 dataset revealed clear evidence that students are willing to deliver low marks to their peers. Based on focus group evidence it is known that participants do reflect on the "poorer" contributions of some group members and there is evidence from the quantitative analysis that there is an ability to identify certain students and a willingness to reward them appropriately (i.e. with lower

marks). Of students in the 2003/04 cohort (n. = 288), 13% received peer review factors of 0.8 or below. This indicates a noticeably worse than average mark (average = 1.0). Of these 13% of students who were marked down by their peers, 41% also attracted a penalty mark for not participating in the peer review process (against 11.8% of the whole cohort who did not take part in the peer review process). Thus there is clear evidence that many of the students who were marked down by their peers had failed to engage in a broader sense and again this suggests that there is a fairness to marks obtained from our peer review system. However, the analysis was not sensitive enough to link lower peer review scores with ultimate academic performance.

Finding 4 – other issues

In addition to the issues of "too demanding", "collusion" and "fairness" identified from the literature and from the focus groups, there were some other findings from the analysis ~~of datasets~~of datasets of student generated peer review marks.

Firstly, where a more detailed review of paper based datasets and web based datasets was possible (2002/03 and 2003/04 – paper; 2004/05 web) we found that at the level of the individual (i.e. the sorts of scores provided by an individual student) there were no significant differences across the datasets. This would suggest that the data collection method (both in terms of level of technology and context in which the data was collected) had no significant impact on the outcomes.

This is important when it is recalled that using the paper based method students were given a maximum of 30 minutes to complete the peer review paperwork in "near exam" conditions. On the other hand, for the web based data collection method students could select the time that they used the Web based system over a period of a week and that their data entry was unsupervised. The fact that there is no difference between these data collection methods is an important finding and one that may well indicate that the students produce objective and "fair" peer review marks that have been considered and reflected upon rather than swift judgements made in the heat of the moment. This latter point could have been a specific criticism of the paper based method, but such a criticism seems unfounded.

Secondly, and particularly from the tutor perspective, peer review clearly gives an opportunity to increase discrimination between students. For example, standard deviations of marks awarded are clearly greater after peer review than before peer review. Table 3 clearly indicates that peer review has a marked effect on individual student benefit from group marks:

TABLE 3 ABOUT HERE

Discussion and recommendations

Key objectives of the HEA funded research project on which this paper is based have been fully met as the project methodology has enabled us to review a vital element of any successful peer review system – the engagement of the students we purport to assess. The project has also provided rich data to support further

enhancements to the web based peer review tool used at Loughborough and to aid development of a good practice guide in this area.

The research methodologies used lead us to interpret the findings as objective and credible. Table 2 summarises the design factors used to reduce the incidence of bias in the focus groups.

Our findings and conclusions expand and extend the literature on peer review and peer assessment. This discussion highlights the findings in respect of key questions regarding peer review from the literature and goes on to indicate new issues arising. The paper concludes with a discussion of areas for future research.

The question of reliability of peer review systems is a consistent theme. Students taking part in this research confirmed that peer marking can be open to subjectivity and personal likes and dislikes (Brown and Knight, 1994). However, this research has shown that tutor selection of groups and assured anonymity do mitigate this effect from the student perspective. The difference between paper based and web based collection tools made little difference to this finding.

Similar reasoning can be used to show that systems can avoid bias and the perceived "unfairness" noted by Conway et al (1993). Students noted that subjectivity was possible as was collusion in marking. The extent of and nature of collusion was not measured, however, as this would have stifled openness in focus groups. Data analysis does show possible collusion revealed by low

standard deviations of group marks but this is not conclusive. Larger group sizes can make collusive behaviour less likely but these have implications for group dynamics and efficiency. Collusion is a continuing fear of tutors and students alike and, because of its nature, is unlikely to be revealed in a timely enough manner to affect assessed marks. Collusion can either help to hide "free riders" or punish them unduly harshly. In a similar way high performers can see marks reduced as the team marks converge towards the average or they can be inappropriately rewarded for their efforts. The focus groups felt that peer review was helpful as a rough guide to individual effort within a team

Under and over marking appears to be as a result of a number of factors, including the academic standard and skill levels of different students. Friendship (or antipathy) and collusion appear to have a minor impact. The data analysis showed that individuals do not always exaggerate their own marks, a finding consistent with some previous studies (Lejk, 2002; Davies, 2006; Boud and Falchicov, 1989).

Observation from the tutor perspective confirms that marginal transaction costs are lower using a web based system, thus refining our understanding of overall transaction costs (Alexander and MacKenzie, 1998). This effect is greater each time the software is used as the same templates for assessment can be carried forward. The set-up costs of such a system are, naturally, well in excess of paper based systems.

Finally the question of anonymity and the finding that this encourages more honest marking is noted. This, again, helps to build on earlier research (Davies, 2000; Lin, 2002). That no difference was noted between paper based and web based systems is encouraging, suggesting that either methodology can be used and that considerations of budget or access to the software need not be a barrier to innovation.

Areas for further research and possible further development of the web based system have also been highlighted by this research. A new project is currently underway to gauge, in part, the extent of emotional stress in the peer review exercise. A clearer view of how these stresses affect individual markers will advise deployment of peer review in the future. The new research also seeks to identify factors affecting collusive behaviour in peer marking.

The opportunity to add explanatory detail to justify marks given was discussed at length in focus groups yet the reasons for desiring this are mixed. No doubt such data collection is technically feasible but what purpose would collection serve? The purpose could be simply to achieve wider acceptance of peer review by students, by extending their engagement with and investment in the system. At another level the explanations could form the basis of feedback to the students so marked in order to support their development of specific skills. If the latter is the purpose then there would need to be considerable thought around the issues of data protection and student comments may need to be moderated prior to their use as feedback. Such comments would certainly be useful if peer review is followed up by developmental work with personal tutors. Counter to this,

however, is the potential disruption to group harmony and possible failure to meet other educational objectives if the peer review is perceived as an end in itself rather than a facilitator of development.

This paper has provided some evidence in both qualitative and quantitative forms to assuage doubts about the fairness and validity of peer review marking. In this way it lends weight to the arguments for adoption of this assessment technique and weakens common arguments against it. Whilst collusion was discussed there appears to be little evidence of it in practice amongst the first year students whose data we used. Students do understand the concept of peer review and the specific criteria used and show no significant differences between the paper based and web based data collection methods. Likewise the quantitative datasets reveal no differences in student marking and engagement between the two collection methods.

Word count: 6,686

Acknowledgements

References

Alexander S and Mckenzie J, (1998), *An Evaluation of Information Technology Projects in Australian Higher Education,* Australian Government Publishing Services, Canberra

BEST Subject Centre, (2004), *Student employability profiles - Business, Management and Accountancy*, BEST Subject Centre,October.

Boud D and Falchikov N, (1989), *Quantitative studies of student self-assessment in higher education: a critical analysis of findings,* Higher Education, Vol. 18, 529-549.

Brown G and Pendlebury M, (1992), *Assessing Active Learning*, CVCP/USDU., Sheffield.

Brown S and Dove P (eds), (1991), *Self and Peer Assessment*, SCED Paper 63, Birmingham.

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

Catterall M, Maclaran P, (1997), *Focus Group Data and Qualitative Analysis Programs: Coding the Moving Picture as Well as the Snapshots*, Sociological Research Online, Vol. 2, No. 1.

Conway R, Kember D, Sivan A and Wu M, (1993), *Peer assessment of an individual's contribution to a group project*, Assessment and Evaluation in Higher Education, Vol. 18, No. 1.

Davies P, (2005), *Weighting for computerized peer-assessment to be accepted*, http://www.caaconference.com/pastConferences/2005/proceedings/DaviesP2.pdf

Davies P, (2006), *Peer assessment: judging the quality of students' work by comments rather than marks,* Innovations in Education and Teaching International, Vol. 43, No. 1, Feb, 69-82.

Dochy F, Segers M and Sluijsmans D, (1999), *The Use of Self-, Peer and Co-assessment in Higher Education: a review,* Studies in Higher Education, Vol. 24, No. 3, 331-350.

Falchicov N, (1988), *Self and Peer Assessment of a Group Project Designed to Promote the Skills of Capability*, Programmed Learning and Educational Technology, Vol. 25, No. 4, pp 327-339.

Falchikov N and Boud D, (1989), *Student Self-Assessment in Higher Education: A Meta-Analysis*, Review of Educational Research, Vol. 59, No. 4, pp 395-430

Gatfield T, (1999), *Examining student satisfaction with group projects and peer assessment*, Assessment and Evaluation in Higher Education, Vol. 24, No. 4.

Goldfinch J and Raeside R, (1990), *Development of a peer assessment technique for obtaining individual marks on a group project*, Assessment and Evaluation in Higher Education, Vol. 15, No. 3, Autumn.

Griffee DT, (1995), *A Longitudinal study of student feedback: self assessment, course evaluation and teacher evaluation,* (Birmingham Alabama), cited in Dochy et.al. (1999).

Grudens-Schuck N, Lundy Allen B and Larson K, (2004), *Focus Groups Fundamentals*, Methodology Brief, Iowa State University.

Lejk M and Wyvill M, (2002), *Peer Assessment of contributions to a group project: Student attitudes to holistic and category-based approaches*, Assessment and Evaluation in Higher Education, Vol. 27, No. 6.

Lin SSJ, Liu EZ and Yuan SM, (2002), *Student attitudes toward networked peer assessment: Case studies of undergraduate students and senior high school students*, International Journal of Instructional Media, Vol. 29(2).

Menchaca M, Resta P and Awalt C, (2002), *Self and peer assessment in an online collaborative learning environment*, Proceedings of E-Learn 2002.

Merton RK, Fiske M and Kendall P, (1956), *The Focused Interview: A Manual of Problems & Procedures.*, Free Press, Glencoe, IL.

Moreira DDA and Da Silva EQ, (2003), *A method to increase student interaction using student groups and peer review over the internet*, Education and Information Technologies, Vol. 8, No. 1.

Morgan DL, (1988), *Focus Groups as Qualitative Research*, Sage, Newbury Park, CA.

Morgan DL, (1993), *Successful Focus Groups: Advancing the State of the Art*, Sage, Newbury Park, CA.

Orsmond P, Merry S and Reiling K, (1996), *The importance of marking criteria in the use of peer assessment*, Assessment and Evaluation in Higher Education, Vol. 21.

Pond K and Ul-Haq R, (1997), *Learning to Assess Using Peer Review*, Studies in Educational Evaluation, Vol. 23, No. 4, December, pp 331-348.

Pond K and Ul-haq R, (1998), *Assessing using Peer Review,* iSEDA paper 102, SEDA, April.

Pond K, Ul-haq R and Wade WB, (1995), *Peer Review - A Precursor to Peer Assessment*, Innovations in Education and Training International, Vol. 32, No.4, pp 314-323.

Stewart WD and Shamdasani N, (1990), *Focus groups: theory and practice (Applied social research methods series)*, Sage Publications, London.

Venables A and Summit R, (2003), *Enhancing scientific essay writing using peer assessment*, Innovations in Education and Training International, Vol. 40, No. 3.

Willmot P and Crawford A, (2004), *Online peer assessed marking of team projects*, International Conference on Engineering Education, Gainsville, Florida, October.

Willmot P and Crawford A, (2005), *Validating the assessment of individuals within undergraduate teams*, International Conference on Engineering Education, Gliwice, Poland, July.

Table 1   Peer review criteria

Criterion	*Explanation*
CO-OPERATION	*This covers attendance at meetings, contribution to meetings, carrying out of designated tasks, dealing with problems.*
COMMUNICATION	*This covers effectiveness in meetings, clarity of work submitted to the group, negotiation with the group, communication between meetings and providing feedback.*
ENTHUSIASM	*This covers motivation, creativity and initiative during the project.*
ORGANISATION	*This covers skills in self-organisation and the ability to organise others. It also covers planning, setting targets, establishing ground rules and keeping to deadlines.*
CONTRIBUTION	*This covers the overall effort put in by an individual during the Semester.*

Table 2   Overcoming the recognised limitations of focus groups

Limitation	Solution
1. Difficult to generalise	No attempt to generalise is made in this research
2. Convenience groups	Participants were offered a number of alternative dates and convenient times that would not clash with their academic timetable.
3. Dominant individuals	Sessions were taped and transcribed. No instances of dominance were observed.
4. Researcher bias in interpretation	Objectivity was ensured by tape transcription prior to reporting on outcomes.
5. Interpretation difficult	Sessions were semi-structured allowing open-endedness to be managed.
6. Researcher bias in group management	The semi-structuring by focusing on issues relevant to the research may have introduced bias but tape transcriptions reveal the facilitator's passive approach.

Table 3   Standard deviations of group and individual marks

Academic year	Group mark Standard deviation	Indiv. Mark Standard deviation
2002/2003	7.73	15.66
2003/2004	4.22	14.65
2004/2005	6.89	14.38
2005/2006	5.24	18.5