

Can students assess students effectively? Some insights into peer-assessment.

Dr A. Mark Langan and Dr C. Philip Wheeler
Department of Environmental and Geographical Sciences

Can naïve markers evaluate their peers? What factors influence student assessors? Is it fair to 'pass on' your marking to students? This article provides an opinion of the current thinking about peer-assessment by describing potential benefits, considering some of the limitations, and colouring discussions with preliminary findings from two projects being undertaken in the Department of Environmental and Geographical Sciences, in the Faculty of Science and Engineering.

Current thinking

There is a growing volume of pedagogical and practical arguments that has been advanced to support peer-assessment by students in higher education (e.g. Falchikov 1995; Magin & Helmore 2001). Many of us have encountered such schemes both at school and in subsequent education, and may have mixed feelings about their place in H.E. In some cases the learning experience of H.E. can fall short of expectations for rational debate, independence of thought and awareness of alternative approaches to a subject (see Stefani 1994). Boud & Falchikov (1989) commented that Australian graduates considered "evaluating one's own work and the work of others" in the top skills required by graduates. They also found that most graduates did not believe that University made a great contribution to acquiring such skills. It was suggested that effective programmes of self and peer-assessment must be based on a sincere commitment to "the encouragement of student autonomy in learning and student responsibility for critical evaluation of their own work". As a Department we have supported inclusion of

Table 1. Potential benefits of peer-assessment

- ✂✂ An educational process that installs autonomy in learners.
- ✂✂ Empowerment of the learner in a learning environment.
- ✂✂ Development of learner confidence in assessing/ marking peers (through practice).
- ✂✂ Development of learner ability to self-evaluate and reflect.
- ✂✂ Greater understanding of what is required by tutors for assessments at degree level.
- ✂✂ Interactive classes for marking/feedback.
- ✂✂ Reflection on recently completed assessments with full explanation of the answer (improving information and understanding).
- ✂✂ Clear, open marking systems (seeing what is required and improving work).
- ✂✂ Seeing standards set by peers as well as mistakes of others (and avoiding them in the future).
- ✂✂ Gaining an ability to 'stand back' from own work for assessment purposes (an essential ability of an 'objective', 'unbiased' scientist).
- ✂✂ Rapid way for a tutor to assess a large amount of student

learners in the assessment process, the rationale for this was to try to encourage independence in their learning and to connect them to the assessment of their academic progress. It has been suggested that classes where students mark a colleagues assignment can initiate an ability to self evaluate and reflect on their own work. This can lead to a greater understanding of what is required by tutors for assessments at degree level (Stefani 1994). An overview of potential benefits of peer-assessment is shown in Table 1. The classes that are needed to support such initiatives can lead to interactive lessons with detailed reflection on recently completed assessments. Hopefully the detailed explanation of answers leads to improved understanding. Of course, this type of assessment necessitates an open marking system (as each assessor needs to see what is required and how to improve the work in front of them) and this provides an opportunity to see standards set by peers as well as their mistakes. There is also the hope that assessors gain an ability to 'stand back' from their own work and assess objectively. It is possible that as students become familiar with the way in which marking criteria are implemented they improve their understanding of assessment procedures. Since this process provides a rapid way for a tutor to assess a large amount of student work and provide specific feedback, the pressure to mark assessments is also reduced (in the second case study below this equated to over five batches of 250 assessments, submitted every 4 weeks). Therefore, peer-assessment can lead to interesting, interactive lessons and less marking.

So why aren't we all doing it?

Although peer-assessment may be a quicker, more comprehensive learning process in some ways, there are of course pitfalls. Many of the associated problems may occur because it is a more complex assessment procedure (compared to tutor marked assessments) and the tutor has to manage a group of mostly very inexperienced assessors. Some tutors are reluctant to introduce peer-assessment due to concerns about the validity and reliability of peer-assessments, leading to the problem of inaccuracy/low precision of naïve markers and variability of marking standards of groups of peer assessors (e.g. Swanson et al. 1991). However, there is considerable evidence that students can peer assess effectively (e.g. Topping 1998; Hughes 2001). Concern has also been raised that the confidence in University assessment practices of 'external' communities (particularly those associated with employment of graduates for professional careers) may be negatively influenced (Freeman 1995). We think this is of minor relevance if peer-assessment is integrated into a diverse portfolio of assessment, on the understanding that learning outcomes are being achieved. Although, there may be little debate about positive reasons to develop skills associated with peer-assessment in learners, there is debate about whether marks generated by peer-assessment should be used for formative assessment only or can be used summatively. The answer to this, and the ultimate success of any such scheme, is largely dependent on the perceptions of those managing the courses, the type of assessment, and the framework used to inform/support naïve assessors.

What works?

There is much advice about peer-assessment procedures, for example Race (1999) suggests that the following types of assessment lend themselves to peer-assessment; presentations, reports, essay plans, calculations, annotated bibliographies, practical work, poster displays, portfolios and exhibitions. There is good reason to use highly objective assessments, with straightforward answers (e.g. calculations) rather than assessments with low objectivity, such as essays. Even apparently 'obvious' answers can generate useful debate, particularly when results have to be interpreted. For example, a mean value from a data set

may be presented to four decimal places, yet the accuracy of the equipment may be only one decimal place, leading to debate about the both equipment and the use of means (in the context of the study involved).

The success of peer-assessment schemes depends greatly on how the process is set-up and subsequently managed. Several authors have provided guidelines for best practice for the management of peer-assessment (e.g. Race 1999; Magin & Helmore 2001; Stefani 1994). In brief, these authorities suggest that peer-assessment systems should include: keeping everybody in the picture (e.g. about how the marks are allocated and why); a simple assessment system (i.e. of high objectivity); negotiating assessment criteria with classes in advance (although this is not always possible); having a moderation system by tutors (for example 10% of the assessments being second marked by tutors); a complaints or review procedure so that peer awarded marks can be discussed/challenged; allowing plenty of time in peer-assessment sessions; and, some form of feedback to students to confirm that peer marks are valid and similar to that of their tutors. Perhaps these guidelines have contributed to the growing evidence that students are able to assess each other (e.g. Hughes 2001). Any tutor implementing such a scheme, needs to have confidence in the marks that are generated (as do the students) and how well student marks correspond to tutor marks. There is a paucity of information about the factors that influence how students assess each other. For example, greater understanding is needed about the effects on assessment by naïve markers, gender responses and the inclusion of learners from different backgrounds (e.g. different courses or even Universities). There is also little information about how inclusion in the development of marking criteria affects the final mark. Currently, we are investigating peer-assessment in two contrasting scenarios and present brief synopses below.

Case Study 1. Peer assessment of 2nd year student presentations.

This study evaluated peer-assessment of presentations by students ($n = 41$) from two universities during a residential field course in ecology / environmental science to southern Spain. A total of eleven tutors (from four Universities) were also involved in the marking process.



A selection of students meet out of the sun, with their tutors, to collectively devise assessment criteria for student presentations.

Students presented findings from their projects for a maximum of five minutes. Talks were grouped into thematic sessions with six or seven talks in each session, and lasted most of the day (with breaks between sessions). The marks for responses to questions were marked only by tutors and are not included in this analysis. One day before the presentations, a stratified random selection of students ($n = 12$; evenly divided between Universities and genders) met to develop the criteria to be used for assessment of the speakers. Therefore, the field course structure allowed several aspects of peer-assessment to be examined: speaker and marker gender; institutional affiliation; participation in development of the

assessment criteria; timing of presentation during each session and during the day. A preliminary examination of findings showed that:

≪≪ Student marks (i.e. students assessing their peers) correlated strongly with tutor marks (i.e. high precision). However, students were more generous, awarding 5% higher marks (i.e. only moderate accuracy).

- ✂✂ The requirement to assess meant that students were actively involved with all the presentations. We believe this increased their attentiveness compared to previous courses when they were more passively involved. Although this was not quantified, student feedback supports this view.
- ✂✂ There were no effects of University affiliation since student assessors were not biased towards or against speakers from a different university.
- ✂✂ There were gender effects. When assessors were male they marked other males more highly than females. Females did not exhibit significant gender biases.
- ✂✂ Involvement of the speaker in the development of marking criteria did not influence their grade as a speaker. However as assessors, participants did award lower grades overall (around 1.5% lower than the students who did not participate in creating the assessment criteria, and only 3.5% higher than tutor grades).

Case Study 2. Peer assessment of statistics assignments by first year undergraduates.

This is very much work in progress and we will identify the need for the current scheme and some very preliminary feedback. Statistics is a subject that consistently creates trepidation for many of our students (and from discussions with colleagues in similar disciplines we are certainly not alone!). The first year undergraduate unit ‘Data Collection and Analysis’ is a core unit to all students studying in our Undergraduate Departmental Network (around 250 annually from 11 courses). The diversity of learners is amplified by their range of interests (e.g. Human Geographers, Applied Ecologists, Physical Geographers, Environmental Scientists). Discussion of the problem of teaching this subject to such diverse learners, and outlines of approach we have taken to ameliorate this, can be found in Langan et al ([2001: on page 10 of the LTSN magazine for Geography, Earth and Environmental Sciences](#)). Despite the development of this unit, there was a perceived need to increase student reflection on some of their

Table 2. Simplified overview of the peer assessment process for Data Collection and Analysis

- ✂✂ Work is handed in to the coursework receiving office in the normal way. The office retains all scripts until the day they are peer assessed and complete a register of submission.
- ✂✂ Scripts are collected on the day and taken to the lecture room. A one hour session uses a Powerpoint presentation and the visualiser to remind students of the process and show examples of answers.
- ✂✂ Students mark work through discussion with tutor and the class. They are reminded that they must trust their judgement and to ask if in doubt.
- ✂✂ Non-attendance of the assessment sessions leads to a 50% reduction in the mark awarded to the work. Attendance by those who didn't complete the assessment gains marks (since the student learns about the assessment in some detail); in this case we award 50% of the class mean mark.
- ✂✂ Markers sign the assessment they are marking and include a general comment on the quality of the script.
- ✂✂ At the end of the lesson, scripts are collected and 25 are marked by a tutor before they are returned to the receiving office for collection.
- ✂✂ Student marks (not tutor marks) are used for summative purposes, unless there is an average of more than 5% difference between student and tutor marks (under which case scripts are reassessed by the tutor).

work since they took very little notice of markers comments other than the mark itself (this was not surprising since the statistics worksheets required mostly entering “the correct numbers in the correct boxes”). We also wished to maintain the rapid turnover of assessments. However, with over 250 students completing five assignments in this part of the unit with only a few weeks between them, the marking was an enormous burden on tutor time (and departmental resources as additional markers were employed). After many discussions, particularly about online assessment, it was decided that peer assessment would be one way forward to achieve the learning outcomes, increase reflection and decrease marking. This scheme focuses on five statistics worksheets, which comprised 50% of the unit (the value of assessments in order are; 5%, 10%, 10%, 15%, and 10%). The following is written during the implementation of this scheme and before we have any structured feedback from the class.

We capitalised on the lack of pre-conceptions of assessment of first year entrants. Students were introduced to the concepts supporting peer-assessment and the process they were involved with during an introductory lecture soon after induction into the University. The assessment procedure started with a relatively simple assignment of very high objectivity and low weighting (essentially an MCQ tick box answer to questions about data types). It was not possible to agree marking criteria with the students for several reasons; the large number of students; considerable time constraints; and, a lack of knowledge of the subject area (contrast this with the scheme developed to mark presentations in Case Study 1). Assignments were read through at the start of the class and, so far, students have indicated that they felt the mark allocation provided has been fair. There was also no time to rehearse the peer-assessment process with the class, but the simplistic/high objectivity nature of the first assessment meant that students settled quickly into their roles. We also apply a penalty system for non-attendance of these sessions, for details of this and an outline of the process see Table 2.

Early indications are that the assessment sessions have run very well. Students contributed to the class discussions (finding it easier to speak about other peoples work than their own) and some have commented that they enjoy the sessions. By chatting with students during the sessions, and listening to discussion between students when deciding on the marks that should be allocated, we have gained confidence in their ability to assess. From the sessions that we have completed, the student marks have been very similar to the selection (i.e.

10%) of scripts that are tutor marked. Classroom management has been of critical importance, not only since a suitable atmosphere is required (open, friendly and professional) but since student work is handed out it is of paramount importance that all scripts are collected at the end of the lecture (timing has to be good if the lecture theatre is required in the following hour). Tutors need support in larger classes, in this case the work is passed to the end of each row to be collected by a postgraduate student (who also handed

Table 3. Overview of assignment/assessment process of Data Collection and Analysis. This is cycled five times during the unit.	
Session 0	Lecture on peer-assessment (at start of unit only)
Session 1	Lecture on concepts behind the data collection for that day Laboratory class to collect data
Session 2	Lecture on concepts behind the statistics assessment Computer workshop with tutor support and completion of an exemplary exercise
Session 3	Follow-up lecture on the statistics and further examples Completion and submission of statistics assignment
Session 4	Peer-assessment session (mark colleagues script) Assignment collected to get feedback from marker Checked marking of own script against model answer on web.

out the work at the start of the class). Another interesting change we made was to abolish the anonymity codes that were introduced at the start of the session (incorrect usage created many problems and a few students began to use them for other courses). Students who wanted anonymity had the option of arranging their own code individually, although to date no one has done this. So far we have received some positive feedback about the assessment sessions and will be questioning the students further after the final session. We feel we have introduced many opportunities to reflect on their work (see Table 3).

Conclusions

There seems good reason to involve students with peer-assessment during their degrees. To avoid some of the problems associated with student empowerment of this type, schemes require openness in dialogue with students, planning, and close monitoring in the early stages. We are not suggesting that all courses should be dominated by peer-assessment, far from it. However, there seems to be a place for this practice at all levels in an integrative assessment strategy for degree courses. At least some types of assessment lend themselves to this type of procedure (e.g. fairly objective tests, presentations) and the process can generate interesting lessons and more reflection by, and involvement of, the students. There is no doubt that successful peer-assessment can reduce the burden of marking. Whether the marks generated are used summatively or formatively needs to be discussed with colleagues. Early indications from our projects indicate that students perceived benefits of peer assessing, enjoyed the sessions and gained a greater understanding of the assessments. Highly objective assessments (such as statistics worksheets, including the interpretation of findings) led to higher marking accuracy (and we are using these marks summatively) than less objective assessment such as evaluation of presentations. However, there was good precision in the marks generated by peer-assessing presentations and extensive literature has suggested that peer-assessment of student presentations for summative purposes is feasible. Some factors influenced marking objectivity (e.g. gender) whereas other potential sources of bias did not (i.e. learner background, in this case University affiliation). One of our considerations is what action should be taken if marks deviate from the tutor grades (from the 10% of scripts tutor marked). We currently have three scenarios in mind: if student marks are within 5% of tutor marks we accept the student marks; if student marks differ by >5% from tutor marks, and there is a predictable direction (e.g. always higher) then a correction could be applied; if there is a >5% deviation with no predictable direction then all the scripts may need to be tutor marked. Further interpretation of these results (leading to a review of the guidance given to future peer assessors) are currently being written up for publication.

Acknowledgements

Thanks to those staff and students from Manchester Metropolitan University and the Victoria University of Manchester who contributed so positively to our projects. Also thanks to the L&T unit for financial support for Case Study 2.

References

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

Freeman, M. (1995) Peer assessment by groups of group work. *Assessment and Evaluation in Higher Education* 20, 289-299.

Falchikov, N. (1995) Peer feedback marking: developing peer-assessment. *Innovations in Education and Training International* 32: 175-187.

[Hughes, I. \(2001\) But isn't this what you're paid for? The pros and cons of peer- and self-assessment. Planet Magazine, National Subject Centre for Geography, Earth and Environmental Sciences, Learning and Teaching Support Network, Issue 2, 20-23.](#)

[Langan, A.M., Wheater C.P., Dunleavy P.J. & Allman R.A. \(2001\) The 'statisticar': Driving data collection and analysis. Planet Magazine, National Subject Centre for Geography, Earth and Environmental Sciences, Learning and Teaching Support Network, Issue 2, 10-12.](#)

Magin, D. & Helmore, P. (2001) Peer and teacher assessments of oral presentations: how reliable are they? *Studies in Higher Education* 26: 287-298.

Race, P. (1999) 2000 Tips for lecturers. Kogan Page, London.

Stefani, A.J. (1994) Self, peer and group assessment procedures. In: *An enterprising curriculum: Teaching innovations in Higher Education*. Eds I. Sneddon and J. Kramer. Pp 24-46. HMSO, Belfast.

Swanson, D., Case, S. & van der Vlueten, C. (1991) Strategies for student assessment. In: *The Challenge of Problem Based Learning*. Eds. D. Boud & G. Feletti. Pp 260-273. Kogan Page, London.

Topping, K. (1998) Peer assessment between students in colleges and universities. *Review of Educational Research* 68: 249-276.

Mark Langan
0161 247 1583
m.langan@mmu.ac.uk

Phil Wheater
0161 247 1589
p.wheater@mmu.ac.uk