

Students assessing student: case studies on peer assessment

C Philip Wheater, A. Mark Langan and Peter J Dunleavy

Department of Environmental and Geographical Sciences, Manchester Metropolitan University

Abstract

Peer-assessment can be viewed in many ways: from a tool for reflection by students to a method of reducing staff marking loads. When deciding whether to use peer-assessment, several questions arise. Can naïve inexperienced markers evaluate their peers? What factors influence student assessors? Should you 'pass on' your marking to students? This article considers current issues about peer-assessment by discussing preliminary findings from two current projects currently being undertaken.

Introduction

A growing number of pedagogical and practical arguments support the use of peer-assessment in higher education (e.g. Falchikov 1995; Magin & Helmore 2001; and see Hughes 2001 for an overview of potential benefits). For example, one study showed many Australian graduates to consider evaluating other people's work to be an important graduate skill, but did not believe their university helped them to acquire such skills (Boud & Falchikov 1989). Classes where students mark a colleague's assignment may initiate skills of self-evaluation and reflection leading to a greater understanding of tutor requirements (Stefani 1994). Classes to support peer-assessment can be interactive sessions with detailed reflection on recently completed assignments, leading to improved understanding. Such assessment necessitates an open marking system (so each assessor knows what is required and how to improve the work in front of them) and provides an opportunity to see standards set, and mistakes made, by peers. Hopefully, student assessors gain an ability to 'stand back' from their own work to assess objectively. The process also provides a rapid way for tutors to assess large volumes of student work and provide detailed feedback. Therefore, peer-assessment can lead to interesting, interactive lessons and less marking for staff.

Many types of assessments lend themselves to peer-assessment: presentations, reports, essay plans, calculations, annotated bibliographies, practical work, poster displays, portfolios and exhibitions (Race 1999). Even so, there is good reason to use highly objective assessments with straightforward answers (e.g. calculations) rather than assessments with relatively low objectivity, such as essays. Even apparently 'obvious' answers can generate useful debate, particularly when results are interpreted. For example, mean values may be presented to a number of decimal places well beyond the accuracy of the equipment, leading to debate about the use of means in the context of the study involved.

However, there are pitfalls associated with peer assessment. The tutor must manage a group of inexperienced assessors and the process may use up valuable class time. There may be concerns about validity and reliability, including inaccuracy and low precision by naïve markers, and variability between 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), and many problems may be of minor importance where peer-assessment is integrated into a diverse portfolio of assessment. The success of peer-assessment depends greatly on how the process is set-up and managed. Several authors give guidelines for managing peer-assessment (e.g. Race 1999; Magin & Helmore 2001; Stefani 1994), including:

keeping everybody in the picture (e.g. about how marks are allocated and why); using 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; employing a complaints and review procedure so that peer awarded marks can be discussed and challenged; allowing plenty of time in peer-assessment sessions; and, giving feedback to students to confirm their marks are valid and similar to those from tutors. However, greater understanding is needed about the effects on assessment of inexperienced markers, gender responses differences, and the inclusion of learners from different backgrounds (e.g. from different courses or even universities). There is also little information about how involvement in the development of marking criteria by students affects the final mark. Currently, we are investigating peer-assessment in two contrasting scenarios and present brief synopses below.

Case Study 1. Formative peer-assessment of 2nd / 3rd year student presentations

This evaluated peer-assessment of presentations by undergraduate students ($n = 41$) from two universities¹ during an overseas residential field course in ecology and environmental science held in the summer vacation between their second and third years. Eleven tutors (from four universities²) were also involved in the marking. Students presented findings from individual projects for a maximum of five minutes. Talks were spread throughout a single day, grouped into six thematic sessions with six or seven talks per session. Each session ended with a general discussion of the theme and offered opportunities for the presenters to be asked questions. All the students and tutors marked all the presentations, but responses to questions were marked only by tutors. The 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. We thus were able to examine several aspects: speaker and marker gender; institutional affiliation; participation in development of the assessment criteria; timing of presentation during each session and during the day. Detailed analysis of some aspects of this work are reported in Langan et al. (2005); preliminary results showed that:

- Student marks correlated strongly with tutor marks (high precision), but students were more generous, awarding ~5% higher marks (moderate accuracy).
- Having to assess meant that students were actively involved with all presentations, increasing their attentiveness compared to previous courses when they were more passively involved.
- There were no effects of university affiliation: student assessors were not biased towards or against speakers from a different university.
- There were gender effects: male assessors marked other males more highly than females, while females did not exhibit significant gender biases.
- Involvement in the development of marking criteria did not influence students' grades as speakers, but participants awarded lower slightly lower grades (~1.5% lower than those who did not help to develop the criteria).

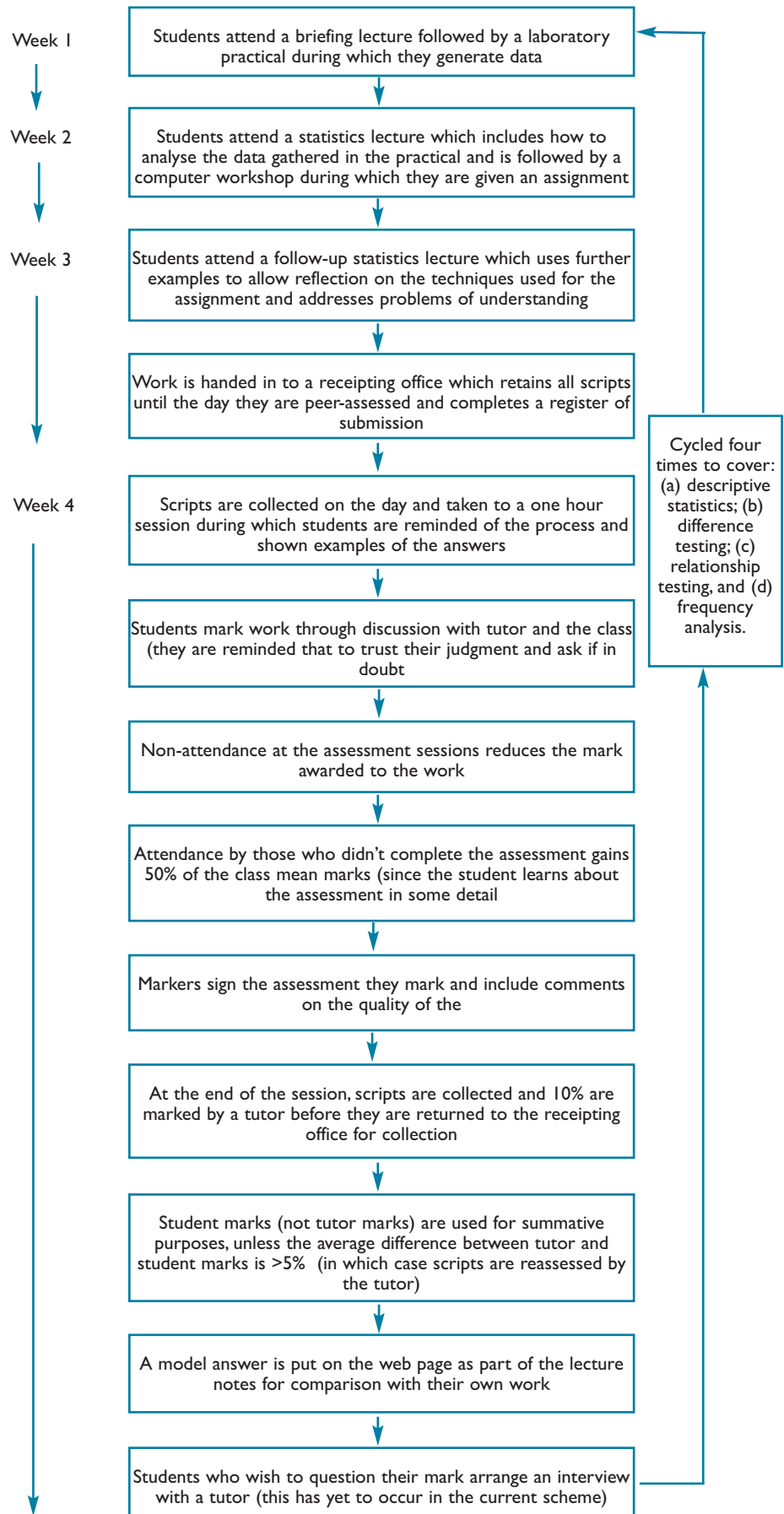
Case Study 2. Summative peer-assessment of statistics assignments by first year undergraduates

We run a first year undergraduate core unit 'Data Collection and Analysis' for all students studying on our Departmental Network (>200 annually from 11 courses). Statistics causes concern for many students and 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 the approach we have taken to ameliorate this, can be found in Langan et al. (2001). Following development of this unit, we perceived a need to increase student reflection on their work since many took little notice of feedback other than the mark itself. We also wished to maintain a rapid turnover of assessments. However, with >200 students completing five statistical assignments with only four weeks between each, marking became an enormous burden on tutors. After discussion, we decided that peer-assessment might enable students to achieve the learning outcomes, increase reflection and decrease staff marking loads. This scheme focuses on five statistics worksheets, comprising 50% of the unit.

Students were introduced to the process during an introductory initial lecture and the procedure started with a simple assignment of high objectivity and low weighting (MCQ tick box on data types). Later assignments, while based on statistical analyses (highly objective), also emphasized interpretation (more subjective). It was not possible to agree marking criteria with students in advance of the assessment because of the large class sizes, considerable time constraints, and students' lack of knowledge of the subject (cf the scheme developed in Case Study 1). So far, students indicate that they feel the mark allocation has been fair. There was no time to rehearse the peer-assessment process with the class, but the simplistic and high objectivity nature of the first assessment meant that students settled quickly into their roles. We also apply a penalty system for non-attendance at these sessions (for an outline of the process see Figure 1).

Early indications are that the assessment sessions ran very well. Students contributed to the class discussions (apparently finding it easier to speak about other's work rather than their own) and some commented that they enjoy the sessions. By chatting with students during the sessions, and listening to discussions between them when deciding on the marks to allocate, we have gained confidence in their ability to assess. Student

Figure 1 Simplified overview of the peer-assessment process for Data Collection and Analysis



marks have been very similar to the 10% of scripts that are tutor marked. Classroom management including efficient use of time is critical, not only because an open, friendly and professional atmosphere is required, but since student work is handed out, it is essential that all scripts are collected at the end of the session. During the year we abolished the anonymity codes that we introduced at the beginning, since incorrect usage created many problems. Students who want anonymity have the option of arranging their own code individually, although to date no one has done this. Two interesting results have been: that students who have not completed the assignment themselves are poorer assessors than those who have attempted the work, and the less objective the assignment, the more variability between student and tutor marks.

Conclusions

There are good reasons to involve students with assessment, as outlined in the introduction above. To avoid some problems associated with student empowerment of this type (e.g. inexperienced assessors, concerns about validity and reliability), schemes require openness in dialogue, good planning, and close monitoring in the early stages. We are not suggesting that courses should be dominated by peer-assessment, but 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 such procedures (e.g. 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. Early indications from our projects indicate that students perceived benefits of peer-assessment, enjoyed the sessions and gained a greater understanding of the assessments. Whether the marks generated are used summatively or formatively needs to be discussed with colleagues. 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 the evaluation of presentations. However, there was high precision in the marks generated by peer-assessing presentations and the extensive literature suggests 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 (e.g. university affiliation). One of our current considerations is what action should be taken if marks deviate from the tutor grades (from the 10% of scripts that are tutor marked). We have three scenarios: for student marks within 5% of tutor marks, we accept the student

marks; where student marks differ by >5% from tutor marks in a predictable direction (e.g. always higher) then a correction could be applied; where there is a >5% deviation with no predictable direction then all the scripts may need to be tutor marked. More detailed examination of these results is currently being undertaken.

1. The two universities were Manchester Metropolitan University and the Victoria University of Manchester (which was its name at the time prior to merging with UMIST to become the University of Manchester).
2. The four universities were the above two plus Sussex and Newcastle.

Acknowledgements

Thanks to those staff and students from Manchester Metropolitan University and the Victoria University of Manchester who contributed so positively to our projects, and to all involved in our Spanish field course and first year data analysis unit.

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.
- 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*, National Subject Centre for Geography, Earth and Environmental Sciences, Learning and Teaching Support Network, Issue 2, 20-23.
- Langan, A.M., Wheeler C.P., Dunleavy P.J. & Allman R.A. (2001) The 'statisticar': Driving data collection and analysis. *Planet*, National Subject Centre for Geography, Earth and Environmental Sciences, Learning and Teaching Support Network. Issue 2, 10-12.
- Langan A. M., Wheeler C. P., Shaw, E. M., Haines B. J., Cullen R. C., Boyle J. C., Penney D., Oldekop J. A., Ashcroft C., Lockey L. and Preziosi R. F. (2005) Peer assessment of oral presentations: effects of student gender, university affiliation and participation in the development of assessment criteria. *Assessment & Evaluation in Higher Education* 30(1): 21-34.
- 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. Felett. Pp 260-273. Kogan Page, London.
- Topping, K. (1998) Peer assessment between students in colleges and universities. *Review of Educational Research* 68: 249-276.

C Philip Wheeler, A Mark Langan and Peter J Dunleavy,
Department of Environmental and Geographical Sciences,
Manchester Metropolitan University, Manchester M1 5G
[P.Wheater@mmu.ac.uk]