



Peer marking of talks in a small, second year biosciences course

Michelle Coulson, School of Molecular and Biomedical Science, The University of Adelaide,
Australia

michelle.coulson@adelaide.edu.au

Abstract: Peer assessment is one way to motivate students to take responsibility for their learning, by increasing engagement with the particular assessment task (Orsmond, 2004). Peer-marking extends the learning possibilities of a particular assessment task by facilitating “learning by teaching”; students learn not just from their own effort, but from the efforts of those whose work they are assessing (Topping, 1998). The current study reports the implementation and evaluation of peer marking of short talks by second year science students. Peer assessment is of particular relevance here. Firstly, the objectives of the assignment explicitly state that the target audience for the talk is “your colleagues in this class”, thus peer marking validates the assessment. Second, one aim of the assessment was to develop strong scientific oral communications skills, thus peer marking encouraged students to engage with multiple talks and hence better understand the components of a good talk. Previously, two staff members would mark each talk, and the average mark would be the final mark for the talk. Now, each talk is marked by a single staff member and the other students. A detailed rubric is used, tailor-designed for the assessment. The final mark for each talk is now comprised: 50% staff mark and 50% average peer mark for that talk. Thus, the staff mark balances any concerns about the peer mark, and likewise the peer mark contributes to fairness, by maintaining multiple markers for every talk. Students’ experience of the peer marking exercise was recorded by a short survey administered after the talk marks were returned. Students were happy enough with the process (certainly there were no complaints), and found the rubric useful. However, they were ambiguous about whether the peer marking exercise increased their engagement with the exercise; some reported that marking talks forced them to pay more attention, while others stated that having to mark a talk distracted them from learning anything by listening to the talk. However, some students reported that marking the work of others did help them understand what was required, and predicted the experience would be of benefit when they next had a similar assessment task. Comparison of peer marks with staff marks shows that generally staff and peers ranked talks in the same order, indicating that the assessment was reasonably reliable. However, it was observed that talks marked highly by the staff marker had an average peer mark lower than the staff mark, while the average peer mark was higher than the staff mark for talks marked lowly by the staff marker. This was because peer marks were generally more clustered than the staff mark, i.e. the difference between highest and lowest peer marks across talks is less than for the staff marks. Overall, implementation of peer marking of talks was quite smooth, marks are sufficiently reliable for a low-stakes assessment piece, and students were either positive or neutral; talks within these courses in the future will continue to be marked by peers plus a single staff member.

Introduction

Peer assessment occurs when students give feedback and/or a mark for the work of their fellow students (reviewed by Topping, 1998). Generally, all students have completed the assessment themselves, and also every student is a marker of at least one student’s work. Peer assessment is thought to promote learning by encouraging students to engage further with an assessment piece (Orsmond, 2004). For example, as a marker the student thinks critically about the work of others, and in doing so is more likely to critically reflect on their own performance. Furthermore, the student as marker experiences multiple exemplars by being engaged with the work of others. In this way, peer assessment aims to enhance the learning power of a single assessment piece (Topping, 1998). Published accounts of peer assessment represent a broad range of settings, including small class sizes and large class sizes, first year courses and advanced level courses. Peer assessment can be used for just about any assessment format, including written work, presentations such as talks or posters, group work skills and professional behaviours (see references within Falchikov & Goldfinch, 2000).

This study describes the implementation and evaluation of peer marking oral presentations in a small, second year biosciences course. I use the synonymous term ‘peer marking’ because the exercise described involves peers assigning marks that contribute (though in a very minor way) to the final mark for the course, rather than providing feedback only. Prior to the introduction of peer marking, each oral presentation was assessed by two staff markers. All students were expected to



attend all talks, but there was no specific motivation for them to be actively involved. Marking a talk should increase the involvement of the student as marker. Peer marking is of particular relevance to this assignment, as the objectives explicitly state that the target audience is “your colleagues in this class”. Knowing that their talk will be marked by peers (i.e. the audience) should better focus the student as presenter on engaging with the audience at an appropriate level. In this paper I describe the peer marking implementation, present results of evaluation of student perception of the exercise, and analyse the resultant marks.

Methods and Results

The peer marking exercise

The students involved in the peer marking exercise form a subgroup of around 30 students within a larger second year biosciences course. This subgroup completes a different practical program that includes literature research and scientific communication skills. Students research the literature on the genetic basis of a specific disease, with each student being allocated a different disease. Students are assessed by oral presentation (talk) and also written essay. The talk contributes 4% towards the final mark for the course. It also provides formative assessment for the later essay on the same topic (the essay is worth 12% of the final mark for the course). Previously, two staff members would mark each talk, and the final mark for the talk was the average staff mark. In the current study, each talk is now marked by a single staff member and other students. The final mark for each student’s talk is now the average of the staff mark and the average peer marks for that talk.

The oral presentations were presented over three weeks. It was agreed amongst students and the staff member that the peer markers in a particular week would be all students not presenting that week. This was done so that students as presenters could focus on their own presentation, and likewise students as markers were not distracted from the responsibilities of marking by stressing about their own upcoming talk.

All markers used a detailed, outcomes-based rubric, with criteria designed against the objectives of the assignment. This rubric was included in the course handbook, and explicitly discussed during a workshop on oral presentation skills, thus ensuring that all students were aware of the marking criteria before preparing their own talk and also before marking any talks. Presentation skills comprised 40% of the mark, with 60% relating to scientific content. Each of these was broken down into specific criteria. For each criterion, standards from unsatisfactory to excellent were described, with associated marks ranges. There was also space for written feedback. At the end of each talk, markers recorded a mark for each section, tallied the components, and recorded the total mark for the talk on the top of the rubric. Students as markers handed their completed rubrics to the staff member at the conclusion of the class. Completed rubrics contained no identifier as to the marker. The additional cost to staff time was entering all talk marks into a spreadsheet for calculation of the final mark for each talk. All rubrics were returned to the presenting student as feedback. The rubric was thus designed to provide detailed feedback as well as guidelines for markers.

Student experience of peer marking

Student experience of the peer marking exercise was evaluated by survey, administered once marks were received by all students as presenters. The survey consisted of 12 Likert-style questions on a five-point scale, plus two open-ended questions. Fifteen completed surveys were received. Students were asked about their engagement with the marking and the assignment, their perceptions of the fairness of the resulting marks, and desired involvement with peer-marking in the future. See Figure 1, Figure 2 and Figure 3 for text of the Likert questions.

Most students were positive about the peer-marking exercise. The lowest broad agreement (i.e. “strongly agree” and “agree” responses combined) in this section was for the question “I enjoyed



marking the talks” (67%, Figure 1). However, even those student who did not enjoy the experience overall, did feel that they benefited from marking the other talks (Figure 1). There was 73% broad agreement that marking other talks increased engagement with the assignment, which is a major aim of peer assessment. Specifically, students reported with very high broad agreement that marking their peers’ work meant they learnt more from the assignment (87%) and would better understand what was required from them in similar assignments in the future (93%). These results suggest that appropriate peer marking can help students achieve gain the most from each assignment. The written comment “Made me pay attention and learn something, helped me in preparing my own talk” support this conclusion. However, this feeling was not unanimous; one student wrote that marking distracted them from enjoying the talks: “When marking, listening to the talks was not as interesting as when I was not marking; I couldn’t engage as much when I was marking.”

Use of a detailed marking rubric by all markers maximised reliability of marks. There was 100% broad agreement that the rubric helped in assigning marks, and only 13% of students thought that marking the talks was difficult (Figure 3). However, one telling comment indicates the dilemma that all novice markers face: “It was hard to decide what mark to give and (I) often just reverted to 8/10 so my marks weren’t that reflective of the quality of the work, but I would like to develop this skill.”

Despite such individual reservations about their own marking skills, no student was negative about the marks they received (Figure 2). This was true for both the staff mark (87% broad agreement) and the average student mark (93% broad agreement). It is vital to validate peer marking in this way when the mark contributes to the student’s final mark for the course, as human nature means that disgruntlement over marks would mean possible learning benefits were not realised.

Only one student did not want to be involved in peer marking in the future, either as a marker or having their work marked by peers (Figure 3). This is an important indicator of the net benefit nearly all students placed on the exercise; other questions asked about specific benefits, but it is possible that these benefits could be outweighed by perceived negatives such as being ‘forced’ to mark.

Analysis of marks

Reliability, validity and fairness of peer marks have been reported as a recurring concern amongst students (e.g. Sittiworachart & Joy, 2008) and also staff (Orsmond, 2004). In the current study, peer marks were never intended to replace staff marks, but rather the question was whether multiple peer markers could act instead of a second staff marker. (It is standard practice in the academic unit to have all assessed oral work marked by at least two markers.) The current assessment design weights a single staff mark equally with the average peer mark to give the final mark for each talk.

Even a detailed rubric is not expected to result in multiple markers all giving a single piece of work the same mark. Although standard deviation amongst peer marks for a specific talk ranged from four to 14 percentage points, there was a very good correlation between average peer marks and the staff marks ($R^2=0.7$, Figure 4). However, the slope of the line of best fit was closer to 0.5 than to one, indicating that the range of marks given by the staff marker is approximately double that awarded by the average peer mark. The mean staff mark was 77%, with a standard deviation of 10 percentage units. In comparison, the mean average peer mark was 80%, but with a standard deviation of only 5 percentage units. The staff mark and average peer mark are the same at approximately 82%; final talk marks lower than this were boosted by the peer marks on average, while talk marks higher than 82% would have been higher on average without the peer mark contribution. The clustering of average peer marks relative to the staff marks can be seen in comparing the distributions (Figure 5).

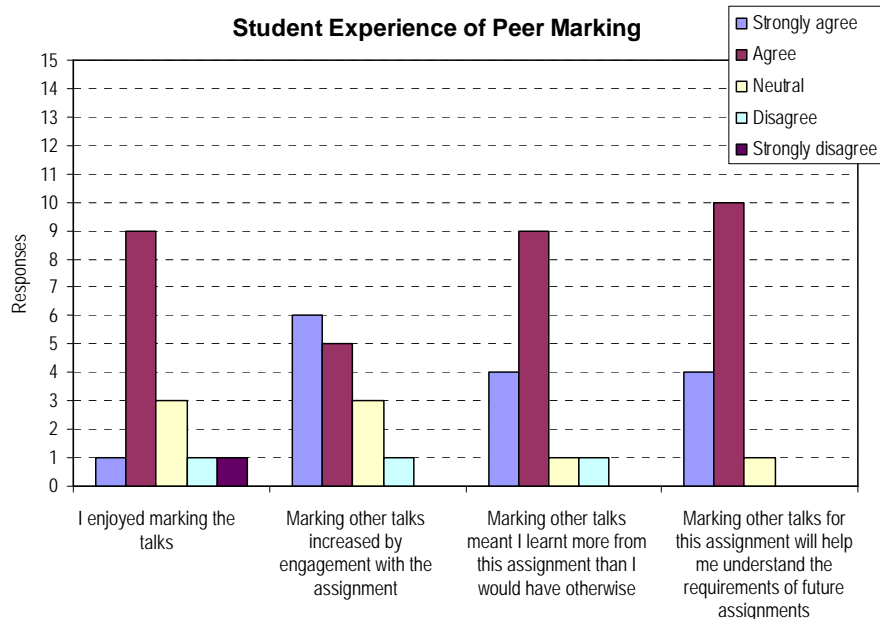


Figure 1: Students were generally favourable towards peer marking, as gauged by survey of the 2007 cohort.

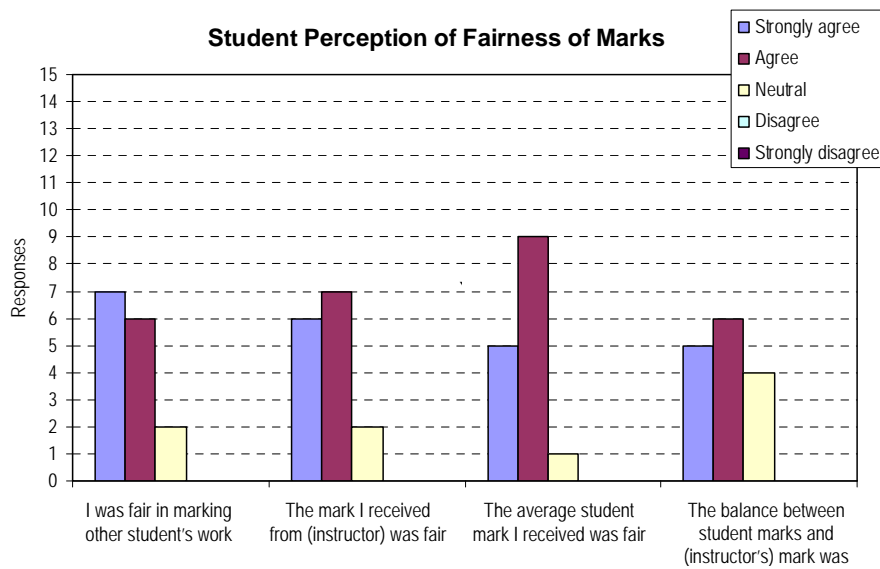


Figure 2: Students agreed that the marking was fair.

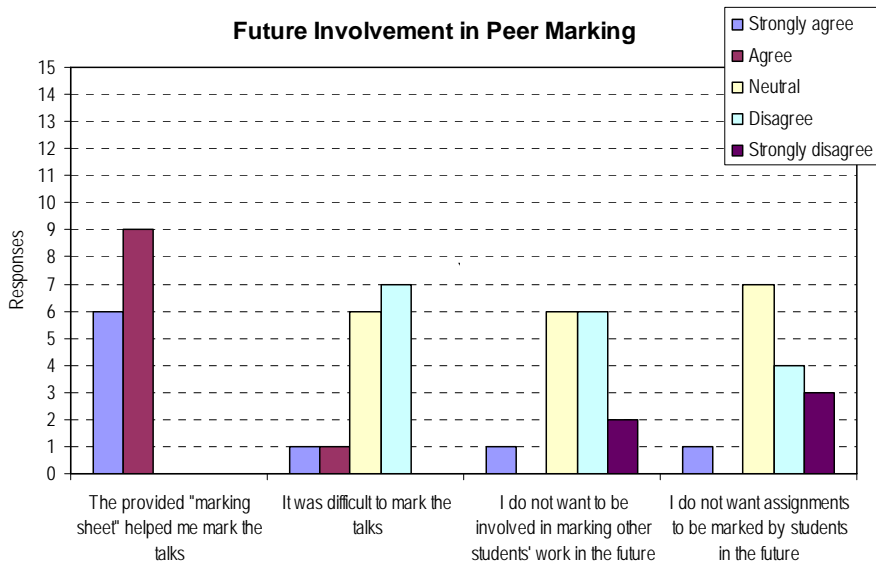


Figure 3: Most students would not mind further involvement in peer marking, and did not find the marking difficult.

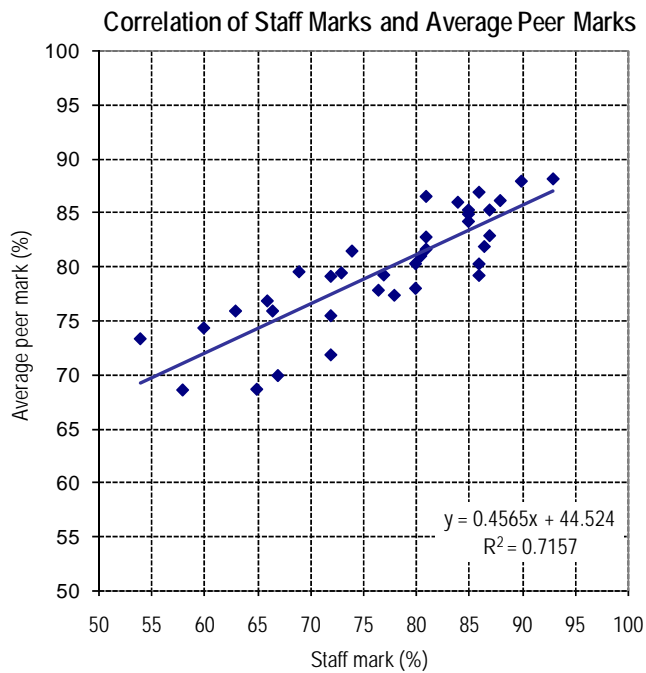


Figure 4: Average peer marks correlated well with the staff marks, but the shallowness of the slope reflects a narrower spread of peer marks compared with staff marks.

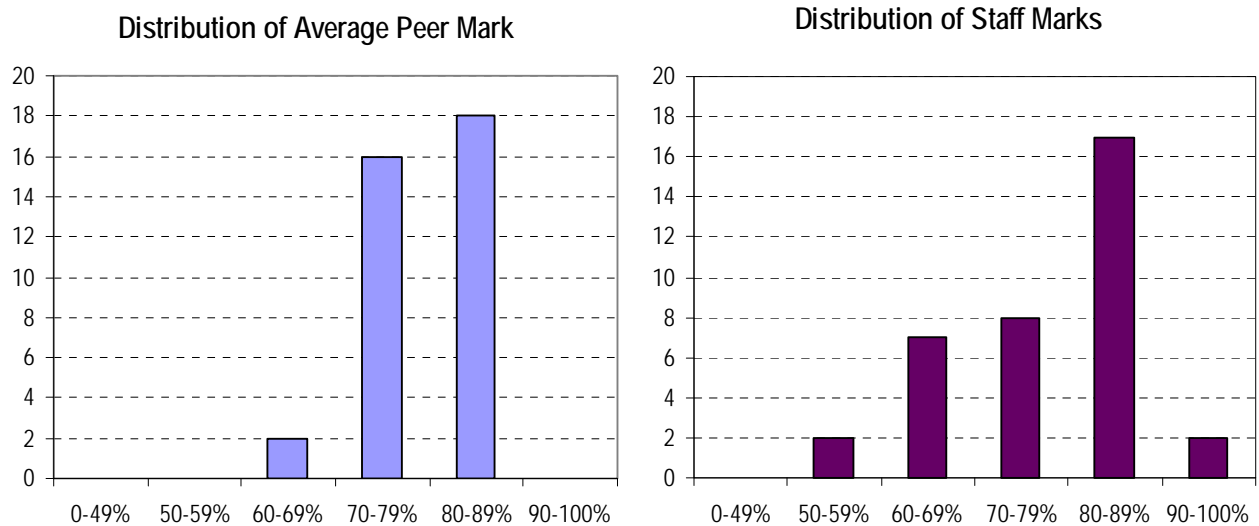


Figure 5: The average peer marks comprised a narrower distribution than marks from the staff member.

Discussion

A standard objective in oral presentation assignments is for students to develop good presentation skills. A main way this can be achieved is from feedback on their own effort. However, this implies multiple presentation opportunities (with incremental improvement), which rarely exist in real courses. Thus, the more ‘experience’ a student can gain from a single oral presentation assignment, the better. Actually giving a talk is direct experience. Additional, indirect experience comes from evaluating the effectiveness of talks experienced from the audience. While this is possible without peer marking, it is the author’s experience that students engage with student talks even less than they do lectures. Having to mark the talk they are listening to effectively forces each student to pay attention. Now that they are actively engaged, they are better placed to build a more sophisticated judgement on what constitutes good presentation skills, which will hopefully be applied to the next oral presentation that they give. Students reported that marking their peers’ work enhanced their experience of the current assignment, and predicted that as a consequence they would better understand what was required of them in future assignments.

Peer marking may be resisted by staff and students because of concerns about reliability of peer marks (Orsmond, 2004), even though there is little evidence to support this concern. For example, meta-analysis by Falchikov and Goldfinch (2000) found that most peer assessment studies reported high correlations between peer and staff marks. However, these authors did observe that some studies show only a low correlation, and thus it is good to determine the relationship between staff and peer marks for any new study, and continue to monitor this relationship over different cohorts of students. The current study showed a high correlation between staff and peer marks. Importantly, students perceived the marks as fair. However, one observation of the current study was that the student markers gave a narrower range of marks. This tendency was also observed by Paré and Joordens (2008), and is referred to in the comment by Sutton, Bartlett, Bellamy, Fincham, Mole and Perry (2004) “Students had appeared hesitant of awarding the full range of marks, not wishing to fail a peer or not confident enough to award top grades when they were deserved.” Using a staff mark as well as an average peer mark minimises any ramifications of clustering. However, if the assessment was worth a higher proportion of the final grade, then this could still be of concern. One would predict that involving the same students in multiple peer assessment exercises would broaden the range of peer marks, as the student markers became more discerning and also more confident in assigning marks.



In summary, peer marking of oral presentations was successfully introduced for a second year biosciences class of approximately 35 students. Peer marks correlated very well with staff marks (a single staff marker), but peer marks had a narrower distribution than marks from the staff member. Students were generally positive about the exercise, and would not mind being involved in peer marking in the future. The benefits of peer marking in the context discussed here are nicely summarised in the student comment: 'Learning more by listening to the talks, able to feel involved in the talks rather than just sitting and doing nothing. Enables me to improve on my talks in the future knowing what the requirements are.'

References

- Falchikov, N. and Goldfinch, D. (2000). Student peer assessment in higher education: a meta-analysis comparing peer and teacher marks. *Review of Educational Research* 70(3), 287-322.
- Paré, D. E. and Joordens, S. (2008). Peering into large lectures: examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning* 24(6), 526-540.
- Orsmond, P. (2004). *Self- and Peer-Assessment: Guidance on Practice in the Biosciences*. Teaching Bioscience: Enhancing Learning series. Leeds, UK: Centre for Bioscience, The Higher Education Academy.
- Sittiworachart, J. and Joy, M. (2008). Computer support of effective peer assessment in an undergraduate programming class. *Journal of Computer Assisted Learning* 24(3), 217-231.
- Sutton, R., Bartlett, T., Bellamy, M., Fincham, D., Mole, J. and Perry, C. (2004). *Evaluating peer assessment of practical write-ups using an explicit marking schedule – same approach, different institution* Retrieved June 20, 2009 from www.bioscience.heacademy.ac.uk/resources/guides/selfaddmat.aspx.
- Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276.

© 2009 Michelle Coulson

The author assigns to UniServe Science and educational non-profit institutions a non-exclusive licence to use this document for personal use and in courses of instruction provided that the article is used in full and this copyright statement is reproduced. The author also grants a non-exclusive licence to UniServe Science to publish this document on the Web (prime sites and mirrors) and in printed form within the UniServe Science 2009 Conference proceedings. Any other usage is prohibited without the express permission of the author. UniServe Science reserved the right to undertake editorial changes in regard to formatting, length of paper and consistency.

The author advises that Human Ethics approval has been sought and granted for this research from their institution.