Peer Assessment in Physics

Sue Burrows and Michael Pounds

A trial of peer assessment took place early in the summer term, prior to the start of exams. This consisted of individual assessment, by 16 students, of group web pages designed by second year students as part of their Physics Skills module PX271, and was followed by a group discussion. A comparison was subsequently made with the marks and comments awarded by an academic earlier in the year.

All 2nd year students and 3rd year M.Phys. students were invited to take part, with 18 accepting but only 16 doing the assessment. As an incentive the students were offered £20 in the form of Amazon vouchers, to cover a predicted one hour for the assessment and a further one hour for the group discussion. Each student assessed two group pages, out of a total of 18, and was asked to mark the assessments in terms of both scientific content and presentation, using the University 17 point marking scale. The feedback form was the same one used by the academics. Weightings for scientific content and presentation were 70% and 30% respectively. One student marked six group pages.

The subsequent group discussion proved productive, with students giving constructive feedback on the process.

Benefits, problems and methods of Peer Assessment

Peer assessment and feedback involves students looking at each other's work (e.g. presentations, assignments, group work) and assessing it against pre-agreed criteria. The benefits of peer assessment are valuable to both students and tutors, if carried out properly and perceived to be fair.

Students can benefit from the process by taking responsibility for assessing the work of their peers using predetermined assessment criteria, with the potential to become more engaged, and to gain a greater understanding of their own skills. Assessment criteria are considered to be an important part of the process. These should be related to the learning outcomes of the course, and can either be developed in conjunction with the students, or else pre-existing criteria can be made known to the students so they can understand the framework against which their work will be assessed. Student-led discussion will promote self-evaluation skills, and also increases participation and confidence in group situations. Judging the individual contribution to group work can be described by two criteria – contribution to discussion, and contribution to group development. This second one displays more importance when there are multiple group activities and tasks rather than a single assessment.

Comparative judgement (CJ) is an alternative method which does not use assessment criteria, but instead uses a comparison between a pair of assessments or

scripts, on the basis that humans are better at making relative judgments than objective judgements, and therefore CJ can be a more effective learning strategy [1]. Previous studies have shown that there can be a gender bias against women when marking individual presentations [2], but this can be balanced by group studies which confirm a wide belief that women are more suited to teamwork. A recent large scale study [3] has shown that there is no overall gender bias in a variety of peer assessment case studies. Men were shown to be slightly more generous in marking, and there are occasions when a mixed gender cohort will give a lower rating than a single gender cohort, but this effect can be cancelled out if multiple assessments are used.

Feedback is an important part of the process, especially if this is given formatively at an interim stage, where marks are not counted towards a final grade, but instead allow the student to improve before submitting a final piece of work. An advantage for a tutor is of course a reduction in the burden of marking. Student perceptions of peer assessment are generally positive as a method of learning, although there is some distrust of the validity of other students' ability to mark [4]. Anonymity can reduce overmarking; this is beneficial where students are reluctant to be seen to award low marks to colleagues.

Results of Warwick Physics Case Study

The module used for this study was Physics Skills PX271, and the web pages assessed were produced in 2014/15. Two group web pages were marked by each student, in terms of both scientific content and presentation. The five marking categories for scientific content were: Scope of the scientific content; Clarity of the science presented; Scientific accuracy of the explanations; Useful links to relevant references; Appropriate scientific level for target audience. Six marking categories for presentation comprised: Word limit, no of pages and links; Logical presentation of material; Relevant figures and diagrams; Appropriate language for target audience; Clear explanations; Spelling and grammar. Each category was marked using the 17 point mark scheme and compared with the academic mark awarded. The difference between the overall mark awarded by the student and the mark awarded by the academic are recorded in Table 1 below.

It was found that when averaged, students were accurate in their marking, with just a slight tendency to overmark (+ 0.11 out of 17). In terms of validity, the standard deviation (of the difference between final marks awarded by student and academic) was 1.57, showing that individual students' marks were not precise. This implies that marking as a group should even out any imbalance in scores.

Group Feedback Session

A one hour feedback session was held after the students had submitted their marks. This was an informal session where students were encouraged to give their thoughts on the process, and to comment on what they perceived to be the benefits and

pitfalls of this peer assessment trial. The students were very engaged with the discussion, and overall were positive about their experience.

Summary of Feedback Discussion

- Some students felt they didn't have the expertise to give a valid mark. More specific guidelines and criteria would be needed for this level of assessment.
- Many students did background reading to enable them to complete the marking.
- There is an advantage in a student marker looking at a larger number of projects, to enable the student to better calibrate their marks.
- If each piece of work is assessed multiple times, biases amongst markers are ironed out.
- If the sample size is large enough, biases can be taken account of by an academic afterwards.
- It would have been viable to just give feedback rather than a mark as well.
- Marking accuracy may have been improved if the projects were more similar – the web pages were on different topics.
- There was a perceived problem that 'better' students may not appreciate being marked by 'mediocre' students.
- It was felt that enthusiastic 1st year students may be more generous markers than other students, although this should not be an issue for modules which are aimed at 2nd year students and above.
- A discussion on marker anonymity was unresolved – some students felt that it should not be anonymous and that feedback should be justified, while others felt that anonymity gave some protection for the marker.
- There was a benefit from looking at the work of others.
- Peer assessment should be done at an interim stage with feedback given, so they can learn from the process before submitting a final piece of work.

Table 1. Marks awarded by both students and academics for each assessment. Marks are given on the 17 point mark scale.

Web page	Student Mark 1	Student Mark 2	Academic Mark	Average Differences
Accretion In Astrophysics	15	15	12	3
Data Storage	13	14	12	1.5
Gravitational Waves	13	13	13	0
Seeing Atoms	12	14	12	1
Solid State Lighting	15	11	14	-1
High Magnetic Fields	14	14	12	2
Inertial Fusion Energy	12	12	13	-1
Medical Imaging	15	14	14	0.5
Muon Spectroscopy	12	14	13	0
Neutron stars and Pulsars	12	14	11	2
Quasicrystals	13.5	14	14	-0.25
Space Weather	12.5	14	14	-0.75
Spintronics	13	13	13	0
Superconductivity	12	14	12	1
Supermassive Black Holes	12	15	14	-0.5
Supernova Explosions	14	13	13	0.5
The First Stars	13	9	14	-3
The Space Elevator	12	12	15	-3
Average				**+ 0.11**
Standard Deviation				**1.58**

Summary

- Group work may lessen any potential bias and even out any imbalance in scores.
- If peer assessment is to go ahead, it should be a pillar of the module
- Credit should be given for doing the assessment
- The assessment should be done in a timetabled session
- Students should be involved with setting marking criteria
- For student to gain a benefit, peer assessment should be done at an interim stage with feedback given, so they can learn from the process before submitting a final piece of work.

References

[1] Peer assessment without assessment criteria. Jones I and Alcock L; Studies in Higher Education, 2014 Vol. 39, No. 10, 1774–1787.
[2] http://www.celt.mmu.ac.uk/ltia/issue4/langanwheater.shtml accessed 25 June 2015
[3] Sex does not matter: gender bias and gender differences in peer assessments of contributions to group work. Tucker R; Assessment & Evaluation in Higher Education, 2014 Vol. 39, No. 3, 293–309.
[4] Student perceptions of peer assessment: an interdisciplinary study. Llado AP et al; Assessment & Evaluation in Higher Education, 2014 Vol. 39, No. 5, 592–610.