Implementing peer review marking as part of group work activities

Cite as;

Brown, A. 2019. Implementing peer review marking as part of group work activities (Study 90) IN: Queen's Assessment Hub.

Available from:

https://go.qub.ac.uk/assesshub

Description | What was done?

A team project assessment has been introduced to a mathematical modelling module to:

1. Meet a university requirement for coursework-based (non-exam) assessment
2. Enhance the skillset of the participating students
3. Reduce the marking burden on lecturers

Over a period of six weeks, the students work in teams of three or four (chosen by the lecturer) to write a short piece of computer code to solve a mathematical problem, generate and analyse results therefrom, and finally to compose a written report on the project. The assessment of the project is worth 50% of the module total (the other 50% coming from a class test). This 50% is broken into three components: 30% for the written report, 10% for an oral exam and 10% for the peer review mark.

The team project component has been employed in a compulsory Level 2 module– Introduction to Partial Differential Equations– for two years beginning in the 2017-18 Academic year. A total of 249 students have participated across the two years. In the most recent academic year, a class of 140 students were split into 39 teams.

Students on the mathematics pathway are often peculiarly solitary in their academic study, and until the introduction of such team-assessments there would have been many students who did not engage with their classmates at all. Given that many of our students go on to careers in the financial sector, this solitary approach was seen as a gaping hole in the students' preparation for life beyond university. Thus, in concert with several other skills-based modules, the team project was introduced to address this shortcoming.

Peer marking implementation

At the conclusion of the project, students were provided with a spreadsheet to complete the peer marking. This spreadsheet was pre-populated with the names of the team members, and made available to each student for download from the VLE. Thus 39 unique spreadsheets (corresponding to the 39 teams) were prepared, and then duplicated for each member of a team, before being uploaded. This was accomplished using custom python scripts to generate and duplicate the files. The spreadsheets also contained several headings under which the students were to assign marks to their peers, a space for comments, and for a final mark for each team member. An excerpt from a sample spreadsheet is shown below. Students received instructions to give a mark out of 5 in each box.

TEAM:	A1			
NAME:	Harry Potter	Hermione Grainger	Ronald Weasley	Draco Malfoy
Attendance at group meetings	4	5	3	3
Contribution to meetings	4	5	3	1
Contribution outside meetings	3	5	2	0
Idea suggesting (e.g. getting started, solving difficulties)	2	5	2	3
Extracting something useful from the ideas (e.g. identifying trends, interpreting results)	3	5	2	2
Encouraging contributions from others	2	2	5	2
Comments		Hermione drove the whole process, without her we wouldn't have completed the project. However, she was a bit domineering, and wasn't very good at listening to others.	Ron was solid, but I feel he could have worked a little harder to prepare. He was good value in the meetings and did what we asked him to, but didn't show much initiative.	Draco seemed to be working against us rather than for us. When he did show up he was surly and uncooperative, and he didn't seem to want to do what he was asked.
final score:	3	5	4	1

In the first year, the spreadsheets were not prepopulated with the team names, but this made harvesting the resulting data from more than 100 spreadsheets a painstaking process, as there was no accounting for students misspelling their team-mates' names or using nicknames.

The spreadsheets were completed by the students, and then submitted online. Another custom python script was used to extract the data from the submitted spreadsheets and deposit a total mark for each student into a master spreadsheet.

The only mark which counted for the summative part of the assessment was the final score awarded in the last row. The other marks and comments were extracted, and each student then received a summary of all the marks and comments made by their team mates. A sample of this feedback is shown below.

AMA2008 project: peer evaluation feedback and scores

Peer evaluation scores for Lavender Brown

- Attendance at group meetings: 1.0
- Contribution to meetings: 2.0
- Contribution outside meetings:
- Idea suggesting:
- Extracting useful knowledge:
- Encouraging participation: 1.0

Overall score: 4.0

Peer evaluation scores for Lavender Brown

- Attendance at group meetings: 5.0
- Contribution to meetings: 4.0
- Contribution outside meetings: 5.0
- Idea suggesting: 4.0
- Extracting useful knowledge: 5.0
- Encouraging participation: 5.0

Comments: 'Very reliable and organised and wrote her part of the write up very well.'

Overall score: 3.0

Peer evaluation scores for Lavender Brown

- Attendance at group meetings: 5.0
- Contribution to meetings: 4.0
- Contribution outside meetings: 4.0
- Idea suggesting: 4.0
- Extracting useful knowledge: 3.0
- Encouraging participation: 4.0

Comments: 'Very organised; came to meetings prepared and with a clear idea of that we needed to do; worked on the theory and conclusion, as well as the other sections; a very good team member.'

Overall score: 3.0

Peer evaluation scores for Lavender Brown

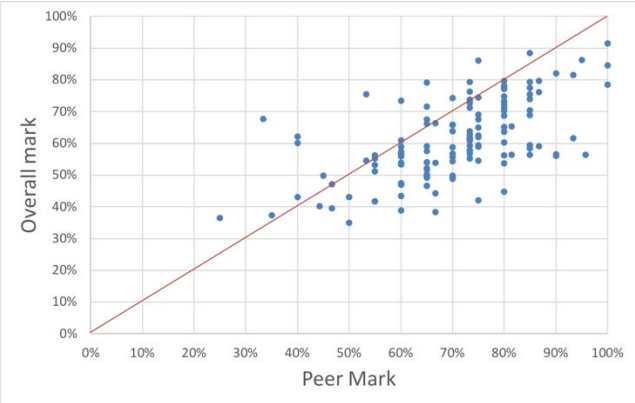
- Attendance at group meetings: 5.0
- Contribution to meetings: 3.0
- Contribution outside meetings: 4.0
- Idea suggesting: 3.0
- Extracting useful knowledge: 4.0
- Encouraging participation: 5.0

Overall score: 3.0

The total number of marks given by each team member was capped so that, if distributed equally, each member would obtain a mark of 3.25/5 (i.e. 65%). For a team of 4 this meant that the total number of marks available was 13. The way the spreadsheets were set up, students could see the number of marks remaining to be allocated, and were prevented from allocating more than the maximum. Additionally, the spreadsheet rules ensured no-one could award a mark higher than 5 or lower than zero. Finally, we note that the students were asked to assign a mark to themselves. The idea here was to force the students to think realistically about the contribution of every team member, including themselves, rather than simply awarding full marks to everyone in the team. The students' self-mark was counted towards their final mark, but as discussed below, it is debatable whether this is advisable or not.

Results

The average mark (as it must be by design) was 65% with a standard deviation of 15%. In general, there was good agreement between team members as to the distribution of marks among their peers, but more than 50% of the class awarded themselves a score 2 marks higher than their team awarded them. As can be seen from the plot below, there is a good correlation between the peer mark and the overall performance of the students but, as would be expected, there is a large number of students obtaining slightly higher for the peer mark.



Shortcomings of the marking scheme

It is arguable that there are two major flaws with the way the marking was implemented. The first is that by retaining the mark they award themselves students are encouraged award themselves full marks at the expense of a truthful reflection of the work done by the team. Although there was a noticeable trend for people to award themselves marks higher than their team mates awarded them (on average by two marks), it was never because they gave themselves full marks. Rather, it seems that students rated themselves as ‘about average’ even if their team mates thought their performance warranted a much lower score. Because the self-mark is worth only one quarter of the 10 percentage points available for the exercise, students over-estimating their contribution gained, on average, a single percentage point by doing so. At most, they could have gained 2.5 points (if their team mates awarded them 0 and they awarded themselves 5), so the potential pitfall is not catastrophic as far as the final marking is concerned. The advantage of counting the student’s self-mark is that it encourages a more realistic reflection on the performance of the team.

The second flaw is that there is no accounting for the quality of the work being done. It would be possible (and indeed quite likely) for a strong student in a strong team to obtain a lower mark than a relatively weak student who happened to find themselves in a weak team. However, this could rather be seen as a strength of the approach. It allows the students to view the ‘team-work’ aspect as a skill to be honed in its own right, separately from the technical aspects of the project. Furthermore, it is arguable that a student taking the lead in a weak team is perhaps worthier of reward than one doing so in a strong team.

It has been mooted that the peer-review mark could be used more creatively than simply as a flat mark added to the student’s project total. The argument is that if a student has contributed substantially more (or less) to the project, it is unfair that they receive the same mark for the final report as their less (or more) diligent team mates. One suggestion is that the peer-mark might be used to scale the report mark by some factor. Take, for example, two reports obtaining a mark of 20/30. In a team where the peer marks were awarded equally, all members would obtain 20/30, but in a team where there was a clear disparity of contribution, the person awarded the highest mark would obtain (say) 23/30, while the one

awarded the lowest mark would receive 17/30. These numbers are chosen somewhat arbitrarily, and having not applied it in earnest to a cohort I cannot comment on how well it might work. I include this information only as a point of interest for others who might wish to try it.

Student performance

In the first incarnation of this module, students completed an individual report on the work done by the team. In this latter instance, allowing the students to complete a joint report on the project improved the standard of writing improved substantially. It was also apparent that students were able to complete substantially more work as a team, than they were as individuals. In the oral exam, the whole team were examined together. The teams that performed best in this component demonstrated some degree of specialisation: i.e. one member took responsibility for the programming aspects, while another took the lead in interpreting the results etc. Interestingly, while there was a very slight trend for better performing teams to have a larger disparity in the distribution of peer marks, this effect was tiny. In other words, the students' assessment of their team mates' performance was largely independent of how successful the project was.

The response to the team project in the module evaluation questionnaire was largely positive, however several students commented that the allocation of the teams was unfair: *'Don't put us in groups with people who can't code!'* wrote one student. While this is an inevitable pitfall in an assessment of this type, we took two steps to address this in the second instance of the module. The first was to provide a one-day crash course in programming skills for students who had not taken the pre-requisite module in the previous year. The second step was to assign the teams so that there were at least two members in every team who had taken the programming course in the previous year.

While these particular steps are specific to this project, in designing your own team project it is worth thinking about the skills and knowledge required, whether everyone will have the same background and training, and if there are any pre-emptive steps which may balance the teams more fairly.

From the perspective of the lecturer the introduction of the team project and specifically the team report saw a substantial improvement in the quality of the work submitted. Indeed, the average mark in the project climbed from 62% in the first year to 68% in the second. Importantly, also, the amount of work was substantially reduced, both in terms of marking (39 projects to mark instead of 109), and in administration, as the use of the python scripts to generate and harvest the peer assessment marks is essentially automatic.

Conclusion

While team-based assessments may not be appropriate in all disciplines or for all modules, where they can be implemented they represent an opportunity for important skill development in students. There are two problems implicit in a team approach however: either unscrupulous students may parasitize their more diligent team-mates' work, or very strong students may dominate the team and discourage input from their peers. Including a peer-marking component can address both of these problems because students know they will be held to account for their contribution to the team dynamic. It also allows the assessment to differentiate the contributions of team members in a (hopefully) fair way,

In modules with a large number of students the practical overhead in collecting, collating the marks and redistributing feedback may be prohibitive. Automating the process with python scripts eliminates all of this overhead.

Resources

The python scripts used for generating the spreadsheets and harvesting the submitted data can be found in the PeerMark repository (**https://github.com/abrown41/PeerMark**). The scripts will almost certainly require customization, but the author is willing to help!