

This item was submitted to [Loughborough's Research Repository](#) by the author.
Items in Figshare are protected by copyright, with all rights reserved, unless otherwise indicated.

Stakeholder perceptions of online peer mark moderation in university teamwork

PLEASE CITE THE PUBLISHED VERSION

PUBLISHER

Research Symposium on Engineering Education

VERSION

NA (Not Applicable or Unknown)

LICENCE

CC BY-NC-ND 4.0

REPOSITORY RECORD

Willmot, Peter, and Keith Pond. 2019. "Stakeholder Perceptions of Online Peer Mark Moderation in University Teamwork". figshare. <https://hdl.handle.net/2134/6489>.

This item was submitted to Loughborough's Institutional Repository (<https://dspace.lboro.ac.uk/>) by the author and is made available under the following Creative Commons Licence conditions.



CC creative commons
COMMONS DEED

Attribution-NonCommercial-NoDerivs 2.5

You are free:

- to copy, distribute, display, and perform the work

Under the following conditions:

 **Attribution.** You must attribute the work in the manner specified by the author or licensor.

 **Noncommercial.** You may not use this work for commercial purposes.

 **No Derivative Works.** You may not alter, transform, or build upon this work.

- For any reuse or distribution, you must make clear to others the license terms of this work.
- Any of these conditions can be waived if you get permission from the copyright holder.

Your fair use and other rights are in no way affected by the above.

This is a human-readable summary of the [Legal Code \(the full license\)](#).

[Disclaimer](#) 

For the full text of this licence, please go to:
<http://creativecommons.org/licenses/by-nc-nd/2.5/>

Stakeholder perceptions of online peer mark moderation in university teamwork

Peter Willmot

School of Mechanical & Manufacturing Engineering, Loughborough University, Loughborough, UK
P.Willmot@lboro.ac.uk

Keith Pond

Business School, Loughborough University, Loughborough, UK
K.Pond@lboro.ac.uk

***Abstract:** Peer assessment can provide a convenient solution to the problem of marking individual students fairly in group assignments. The developing methodology has numerous benefits for enhanced student learning and transferable skill development. Peer Assessment is not, however, universally embraced: critics cite potential drawbacks including collusion and vindictive marking and this paper briefly reviews the state of the art; it goes on to describe a new web-based peer mark moderation tool and outline the results of a quantitative research project based upon its use. Whilst much of the received data confirms and updates previous literature, important new insights are gained into the thoughts of students, who appear to recognise and value the fairness they believe peer mark moderation can offer. Statistical analysis verifies the lack of collusion associated with the web-based system and students comment positively on the system's anonymity and its ability to recognise different levels of achievement within teams.*

Introduction

Of all the problems associated with group or team-based education, the difficulties of precise and individual assessment are supreme. Academics who feel comfortable setting examinations and individual coursework assignments are often deterred from devising team assessments because the student-centred learning approach dictates that they have only a limited knowledge of the real contribution that each team member makes.

Awarding a group mark to everyone in a group has been a cause for concern for many educators due to the obvious lack of recognition of the different levels of effort and ability and a great deal of effort has been put into correcting the injustice. Different methods of resolving this dilemma have been tried and one of the most popular is 'peer' assessment.

Peer Assessment

In this context, the term '(self and) peer assessment' is used to describe the process undertaken by students to assess the performance/contribution of themselves and their peer group, in a group task. This has been more accurately described as 'peer mark-moderation' because the task itself is marked as an entity by the academic supervisor and the students' input just moderates or varies this mark. Though, not the same thing, there are some similarities with this and peer review in terms of the feedback and learning opportunities because they share the important element that students are actively engaged in the assessment process.

Falchikov(1995) identified two distinct types of peer assessment; of product and of performance (also referred to as process). Peer assessment of product is where students assess other students' work: either a finished product, in case of summative assessment, or a work in progress in the case of formative assessment. Peer assessment of performance is where students assess specific skills or level of participation in a group. Peer review is most commonly focussed on product while peer mark-moderation of group work could conceivably be of product and/or performance depending on how the assessment criteria are defined. Clearly, the tutor may choose to assess once or more than once at the

end of the work or at various stages along the road. There are a number of examples that show peer assessment can be used formatively or summatively, with the latter being the most reported. Some examples of summative peer assessment include case studies by Robinson(2006) and Loddington et al(2008), while Wheater et al.(2005) compare two case studies; one summative and one formative, to show success for both types.

An alternative strategy is to treat teams of students like teams in the sporting arena; where all team members benefit or suffer equally from the team's promotion or relegation from the league. Following this argument, in the case of team assignments, we require individuals to entrust their future to the collective outcome and it would be essential to agree this in advance: this argument is not one that is easily accepted, either by students or by teaching quality assessors. The fairness of allocating equal marks to all team members was questioned by Willmot & Crawford(2007) who concluded that this was not the correct approach and stated that the common belief is that "a lazy student might benefit from the efforts of team-mates or particularly diligent students may have their efforts diluted by weaker team members". Pond et al(2007) found that "bunched group marks often show a low standard deviation and the use of peer review {assessment} can help to spread this", which is generally a desirable feature in academia.

There is much concern, not least amongst the student body, over 'free riders' in group work. The term 'free rider' is frequently used to describe a student who relies on others in group work. Unfortunately tutors or project supervisors cannot be solely relied upon to identify and penalise free riders who may present very well but shy away from any real contribution. Moreover, it can be very difficult or near impossible for a tutor to assess students' individual effort of a group task when the majority of work necessarily takes place during non-contact periods. This gives a clear pointer to involving students in the assessment process.

Peer Assessment is not universally embraced: critics cite potential drawbacks including collusion and unfair or vindictive marking. Nevertheless, there are some powerful administrative drivers that continue to attract academics to both team assignments and peer assessment; these were identified by Hughes(2001). Langan & Wheater (2003) report a strong correlation between tutor marks and student marks whereas others(Cheng & Warren, 1997) argue that they have not found sufficient reliability of peer assessment. One commonly cited feature is the need to prepare students for peer assessment and to properly explain the assessment process. Discussions of the criteria beforehand might be helpful (Juwah, 2003) and students need to understand how to apply the assessment criteria (Cheng & Warren, 1997). Of course, this assumes that the methods employed actually have explicit criteria and this is not always the case: it is not uncommon for team members to be simply asked to rate each other at the end of a project through some simple metric but this mechanism clearly offers little pedagogic validity. A reliable and valid assessment must measure against specific targets that are aligned to the intended learning outcomes and course content.

The educational benefits of teamwork are well rehearsed and will not be exercised here and there can be benefits in including peer assessment of some sort. The most tangible benefit for tutors is that team reports save marking time and reduce the ever growing workload. If peer assessment can provide the vehicle to successfully combat complaints of unfair grading associated with simple team-mark allocation, then tutors will be more inclined to use team assignments. The literature suggests that students generally see well presented peer assessment as a fair way of assessing group work and can feel more involved in their learning.

Online Peer Mark Moderation

Web-PA is an online peer-moderated marking system designed for teams of students doing group-work, the outcomes of which earn an overall group mark. Each student in a group grades their team-mates (and their own) performance, and the resulting data impacts on the supervisor's group mark to provide each student with an individual grade reflecting their contribution to the effort.

Web-PA is currently in use in over half the departments across Loughborough university campus and is embedded into the university quality system as the recommended mechanism for group moderation. An open source variant has been developed and adopted in a total of 17 other UK and overseas

universities; in May 2008, the project won an IMS global learning impact award (Austin, Texas). The software incorporates enhancements that help to promote good practice and builds upon existing evaluation of assessment practices across a large range of subject disciplines. The easy to use system is flexible on team size and allows the tutor to define any number of assessment criteria. During setup, the tutor selects teams directly from the central university database and defines the timeframe within which the students must respond. Students enter data between dates specified by the tutor using simple clickable menus. Assessment may be applied at the end of a project or any time during it; more than once if required. Data-entry is confidential and only entry points for their own team appear on-screen.

Put simply, the system calculates a variation factor for each team member (Web-PA factor) based on the total scores received for an individual divided by the normalised average scores for the whole team. The supervisor's team mark, or part of it at the supervisor's discretion, is multiplied by the factor for each individual. If all team members score equally, the Web-PA factor is 1.0 so all members receive the unmodified team mark; moreover, if one student is advantaged by the process, another suffers a mark reduction. After the deadline, the supervisor retrieves a convenient set of data in a variety of customisable formats but retains the option of intervening if foul play is suspected.

This rapidly maturing online tool has been developed over a period of years. The design and development of the software has been supported by the Engineering Centre for Excellence in Teaching and Learning (engCETL) and by JISC in the last few years. The project site can be found at www.webpaproject.com where visitors can access a discussion forum and a demonstrator.

Reflective Research

We set out to discover if online peer mark moderation is considered by users as fair and if the inevitable 'free-riders' and their opposites, 'stars' are adequately dealt with. Do teams conspire together to rig results and are results affected by external influences such as how the teams were formed, their gender, subject discipline or year of study?

Quantitative research benefited from having unique access to copious good quality, consistent data captured by the Web-PA system itself and the same facility also gave email links to students so that surveys could be easily sent out to a large number of students studying a diverse range of degree subjects. Staff user interviews were used at the outset to help focus the survey questions.

The survey was sent out to 2209 students studying on 36 modules in 14 departments of the University in 2007 and resulted in 284 usable responses (response rate of 13%). It asked 27, 6-point Likert scale questions that interrogated the friendliness of the system, the benefits of peer assessment, the fairness of marking, the students own feelings and the extent of collusion at the point of data-input. There were also a number of static data questions describing each respondent in terms of gender, study-year etc.

The final part of the research was to separately analyse the *raw Web-PA data*. The large sample included group assignments taken by students in all years, both undergraduate and postgraduate: representing 730 student interactions. The analysis searched for evidence of fairness/unfairness and honesty/dishonesty of marking and, again, on collusion within teams.

Selected Results of the Student Survey

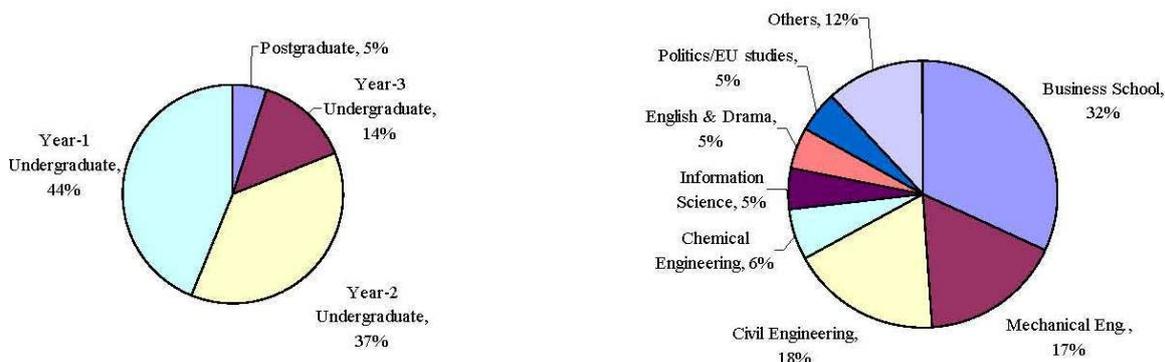


Figure 1: Survey Details

Loughborough has a particularly large engineering faculty and it was from here that Web-PA originated so it is perhaps not surprising that there was a numerical gender bias towards males in the survey (62%). Figure 1 demonstrates the breadth of the survey and names all departments that provided at least 5% of the total response. The majority of the groups were found to be formed by tutor-selection while a significant minority had been formed by the students themselves. It can be assumed that in these cases the students knew each other before the start. The residue (14%) used the method of team-formation known as ‘seeding’: formed by students around a seed-member predetermined by the tutor to have specific attributes.

Statistical Analysis

Standard statistical analysis tools have been used to analyse this significant survey. There follows a broad discussion of the key findings.

		FAIRNESS TESTS				ANONYMITY TESTS		
		Ability to identify 'free riders'	Ability to identify 'stars'	Personal mark perceived too Low	Personal mark perceived too High	Want to know others' marks	Happy to reveal marks	Believe anonymous is more accurate
Year-1 (u'grad)	N	159	159	153	153	156	156	156
	Mean	4.77	4.82	3.16	2.3	3.72	3.41	4.51
	Std. Dev.	1.41	1.18	1.46	1.08	1.88	1.81	1.59
Year-2 (u'grad)	N	56	56	55	55	55	55	54
	Mean	4.45	4.66	2.82	2.49	3.38	2.95	4.98
	Std. Dev.	1.57	1.47	1.28	1.07	1.66	1.71	1.44
Year-3 (u'grad)	N	23	23	23	23	23	23	23
	Mean	3.78	4.23	3.17	2.17	3.96	4.09	3.91
	Std. Dev.	1.59	1.42	1.72	1.27	1.87	1.67	1.86
Postgraduate	N	10	10	10	10	10	10	10
	Mean	4.70	4.70	3.20	3.70	4.40	4.30	4.10
	Std. Dev.	1.25	1.42	1.32	1.5	1.65	1.64	2.33
Total	N	248	248	241	241	244	244	243
	Mean	4.6	4.71	3.09	2.39	3.70	3.41	4.54
	Std. Dev.	1.48	1.29	1.44	1.14	1.83	1.8	1.64

Table 2: Mean responses to questions on perceptions of fairness & anonymity by year-group

Using an ANOVA analysis of mean scores we identified the most significant and statistically most reliable responses. Note: The specific questions analysed in each section are summarised in the tables and the results highlighted in bold are further discussed.

The double-entry table 2 shows how much postgraduate students differ from undergraduates in considering the own peer reviewed mark too generous ($F=5.39$, $p=0.001$). The system appears to be firmly accepted as a means of identifying weaker students ($F=3.31$, $p=0.021$), however, final year undergraduates show more reservations of the system's value in this respect and this could be related to their more 'individualistic' approach to team work because of their sharp focus on a final overall grade. Table 2 also suggests that postgraduate students differ from undergraduate students in regard to the disclosure of their marks to their peers ($F=3.22$, $p=0.23$). Postgraduate, in fact, would prefer that other group members could know what marks they had given them.

An interesting 'battle of sexes' is presented by the results in table 3. Women tend to more strongly value the anonymity that the present system offers; this difference is highlighted by a high level of significance ($p < .005$). They also tend to value the importance of Web-PA for understanding their role within the team. On the other hand, men seem more characterised by a sense of 'camaraderie': they reported to have found it more difficult to give to own friends in their group a low mark, even when this was deserved.

A further bi-variate analysis showed that little difference exists between the different discipline areas of the sample. The only significance emerged in the possibility of students gaining feedback on performance through the peer-mark data: for group members to know the mark that other students had

given them. This is not current practise but the data suggests students of the Business School/ Economics and Politics would appreciate such feedback while English & Drama and Engineering students would prefer not to disclose or receive this information.

GENDER TEST		Wanted to know others' marks	Happy to reveal own marks	Believe anonymous is more accurate	Difficult to award marks to friends
Male	N	153	153	152	153
	Mean	3.77	3.55	4.19	3.54
	Std. Deviation	1.91	1.86	1.79	1.66
Female	N	91	91	91	91
	Mean	3.57	3.18	5.13	3.03
	Std. Deviation	1.68	1.66	1.15	1.47

Table 3: Investigating gender differences

Assessing Honesty

We considered what 'honest' marking might look like from a data point of view. 'Honest' marking implies there is a willingness to discriminate between team members and we would expect there to be engagement with the process. So for 'honest' marking there will be a good chance of a student marking him/herself lower than others in the team: '**self-mark < peer-mark**'. The opposite would be where a student seriously overestimates his/her own scores. We would expect the **groups not to give out 100% of all available marks**. A null return would indicate lack of engagement. Furthermore, one would expect real variation in the performance of any group of individuals against a range of criteria, so a very low or zero standard deviation across the group would indicate possible collusion, laziness or a reluctance to take the review process seriously. This was numerically expressed as a **low percentage of zero standard deviations** by comparing the standard deviation for individuals within groups with the standard deviations of the whole cohort. There were hardly any instances of zero standard deviation or **null return**.

Interviews with staff and the survey of student users suggested that method of group selection; year group and group size all have an effect on 'honesty'. To test 'honesty' data was assembled for six modules; those with the largest population (from 3-departments) and we recorded variations with respect to method of team formation method and year of study. For all the groups we found a reasonable chance of a 'self less than peer' score, but there is variation in the other three measures (above).

Comparing the two year-2 modules that are otherwise similar. The self selected group had a much higher rate of non-submission (13%) and a substantial number (60%) of 'low standard deviations'. More than 40% of the teams had also allocated almost all of the available marks (>95%). In contrast, the 'seeded' groups fared much better with all three tests scoring under 5%.

When comparing two modules that were similar in every respect except 'year of study' the tests suggest much less honesty in the year-3 module compared with the one from year-1. Both modules were tutor-selected, however behaviour could be influenced by the fact that third year the students are likely to know each other well in any case; it is possible that the groups are behaving like a 'self selected' group.

Of particular interest were two modules from years-1 and 2; both seeded and of similar size. These groups both exhibited 'honest' marking behaviour with particularly low zero standard deviations. Students on the year-2 module had experienced the peer review process in their first year and appear to have confidence in and a commitment to the process. However, an alternative explanation might be the style of introduction to the process that this lecturer uses as both modules have the same Responsible Examiner.

Conclusions

There is considerable anecdotal evidence that the benefits of group work have sometimes been overshadowed by concerns of fair marking especially where students are allocated a single team-mark and the Web-PA online mark-moderation method, described here, has met with a very enthusiastic and rapidly growing following. Whilst much of the data it generated supports previous literature on this subject, important new insights are gained into the thoughts of the student participants. After experiencing Web-PA, there is much evidence that users perceive it as fair. More specifically, they comment positively on qualities of anonymity, recognition of 'stars' and 'free-riders' and point to a perhaps surprising lack of collusion associated with using web-based systems. Individual and group marking behaviours also suggest that most peer review marking is 'honest' but can be influenced by method of team formation and year of study. There appears to be evidence against allowing self-selection of groups and early years students show more marking 'honesty' than finalists.

The survey also delved into the area of feedback, i.e. attitudes to giving and receiving the assessment of peers: it is within this idea that there is basis for the frequently heard claim that Peer Review can develop key-skills. Clearly this is a sensitive area that needs careful treatment but, while there is agreement that anonymous marking is preferred, the research suggests women are more strongly opposed to full disclosure of marks and males encounter more difficulty marking their friends.

References

- Cheng, W. & Warren, M. "Having second thoughts: Student perceptions before and after a peer assessment exercise". *Studies in Higher Education*, (1997). 22(2), 233-239.
- Falchikov, N. "Peer feedback marking: Developing Peer Assessment." *Innovations in Education and Training International*, 1995, 32, pp.175-187.
- Hughes, I.E. "But isn't this what you're paid for?" The pros and cons of peer and self assessment. *Planet* (the LTSN Centre for Geography, Earth and Environmental Sciences Bulletin), pp. 20-23, June 2001.
- Juwah, C. "Using peer assessment to develop skills and capabilities". *The Journal of the United States Distance Learning Association*, (2003). 17(1) available online at http://www.usdla.org/html/journal/JAN03_Issue/article04.html
- Langan, A. M. & Wheeler, C. P. "Can students assess student effectively? Some insights into peer-assessment". *Learning and Teaching in Action*, 2(1), (2003). 02/01, 2007
- Loddington, S., Wilkinson, N., Glass, J & Willmot, P., "An Examination of Academic and Student Attitudes to the Peer Assessment of Group Work using WebPA in Engineering", *Proceedings of International Conference on Innovation, Good Practice and Research in Engineering Education; ee2008*, Loughborough, UK, July 2008
- Pond, K., Coates, D.S. and Palermo, O., "Student Experiences of Peer Review Marking of Team Projects", *International Journal of Management Education*, 6(2), 2007, pp 30-43, ISSN 1472-8117.
- Robinson, C. L. "Self and peer assessment in group work." *The European Society for Engineering Education: SEFI MWG Seminar*, Kongsberg, Norway, June 2006
- Wheeler, P., Langan, M. & Dunleavy, P.J." Students assessing student: case studies on peer assessment". *'Planet'; the Journal of the Higher Education Academy Subject Centre for Geography and Earth Sciences*, 2005. 15, p. 13-15. Available online at; <http://www.gees.ac.uk/planet/p15/p15.pdf>
- Willmot, P. and Crawford, A.R., "Peer Review of team marks using a web-based tool: an evaluation", *'Engineering Education: the Journal of the Higher Education Academy Engineering Subject Centre*, 2(1), 2007, PP 59-66, ISSN 1750-0044, Available online from: <http://www.engsc.ac.uk/journal/index.php/ee>

Acknowledgements

This research was made possible by Loughborough University Academic Practice Awards to the authors in 2006-7.

Copyright © 2009 P.Willmot, K.Pond: The authors assign to the REES organisers 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 authors also grant a non-exclusive licence to REES to publish this document in full on the World Wide Web (prime sites and mirrors) on CD-ROM and in printed form within the REES 2009 conference proceedings. Any other usage is prohibited without the express permission of the authors.