Online peer assessed marking of team projects

Authors:

Peter Willmot, School of Mechanical and Manufacturing Engineering, Loughborough University, UK.
Tel 44 1509 227555 p.willmot@lboro.ac.uk
Adam Crawford, Engineering Education Centre, Loughborough University, UK. Tel 44 1509 227198
a.r.crawford@lboro.ac.uk

Abstract *— Group project work plays an increasingly important part in today's university courses as it enhances employability through developing team -working skills. It also solves some of the su pervision problems encountered when coping with large student numbers. When assessing such projects, however, there is a common difficulty: how to determine the appropriate mark for each individual. Much of the recent work in this field has focused on vari ous forms of self and peer assessment as the fundamental mechanism by which an individual mark may be generated. In this paper, we review a number of existing peer marking systems employed by staff at a number of British Universities. Research shows that t here are almost as many variations upon the basic method as there are projects. However, we identify three generic types in common use and describe how we compared the outcomes of applying the different peer assessment systems to the same project and concl ude that the final numerical outcome depends heavily upon type of peer assessment used. We compare the results from the three models and speculate which is most effective. The paper then describes how a flexible web based version of this particular system was developed at Loughborough University and has been put to good use across a wide variety of disciplines and project types. The prime considerations in the development of the software were that students should perceive the process to be fair and that t he resource should be very flexible and easy to use by both staff and students. We use real examples of student's peer marking data to demonstrate the system's effectiveness and compare how students rate themselves with how they are rated by the rest of th eir project group .*

Index Terms *— Peer review, self assessment, team projects, web based, peer moderated.*

INTRODUCTION

Group project work plays an increasingly important part in today's university courses. Well developed team-working and associated professional skills such as good time management, self reliance and effective communications are attributes that are highly sought by today's employers of graduates. Experience in team projects can therefore enhance employability. Staff also perceive teamwork assignments as a potential solution to some of the supervision and assessment problems associated with today's large student numbers so it is not surprising that they have become popular in universities.

There is a common difficulty when assessing team projects: how to determine the appropriate mark for each individual. Failure to address this very real problem not only causes resentment amongst students but is also quite unacceptable from the Quality Assurance viewpoint. Studies by working groups of academics on behalf of the Loughborough based Learning and Teaching Support Network (LTSN) for Engineering have distilled current practice in a number of British Universities [1],[2]. Most acknowledge the problem and that many of the mechanisms in place to address it are, in some cases, less than ideal. Much of the recent work in this field has focused on various forms of self and peer assessment as the fundamental mechanism by which an individual mark is generated and it is clear that improvements are gradually being made.

CAUSE AND EFFECT

A number of pedagogic benefits to self and peer assessment have been advanced:

Potential Student Benefits

- Increased student ownership of learning, encourages team working and provides additional motivation.
- Encourages active participation in the learning process by making assessment a shared activity.
- Encourages improved understanding of the assessment requirements.
- Promotes reflection and hence, potentially deeper understanding.

Potential Staff Benefits

- Can increase opportunities to monitor student progress and identify potential problems.
- Strengthens reflection on assessment criteria and how achievements are to be evidenced and judged.
- Promotes student-centred-learning as students are required to be active participants
- Can avoid disputes if the system chosen is seen to be open and fair.

These laudable features are not always, however, the primary drivers for many staff when they first considered self and peer assessment, often they are by-products of the system. Where peer marking systems have been adopted, it has generally been out of a need to demonstrate fairness coupled with a lack of good information about the contribution of individuals involved in teams. Few alternatives present themselves. Closer scrutiny shows that peer marking has more to offer to the learning process than just being a convenient marks generator, provided that the system can be relied upon as accurate. The primary purpose of self and peer assessment is clearly to generate an appropriate mark for each individual and the question is whether or not such systems can be relied upon to do so.

Furthermore, self and peer assessment challenges well-established beliefs about who should rightfully assess students' work and the respective roles/responsibilities of lecturers and students. The methods can be contentious and invite debate about the maintenance of standards. Peer assessment can be found in many disciplines, Allen and Lloyd-Jones[3] working in the humanities make reference to various peer mark distribution systems and warn that students can be naturally opposed to peer assessment. Nevertheless they acknowledge the need to address the problem of the notorious free-rider that fails to actively engage in group work. It is vital, therefore that any methodology adopted must be perceived as fair and gain a wide acceptance from the students.

CHANGING THE PEER ASSESSMENT ALGORITHM – AN EXPERIMENTAL WORKSHOP

Our earlier research shows that there are almost as many variations upon the basic peer assessment method as there are team projects going on in universities. Staff responsible for project modules commonly hand out a bespoke pro-forma to their teams soliciting a numerical response. This can be done in secrecy but more commonly a collective response is required where teams are asked to agree and sign up to a proportionate mark distribution. The data can be compiled at the end of he project or more than once, at specific intervals. Typically the tutor will award a mark or grade for the project as a whole and students will generate some modifying factor which is applied to the total mark to provide a mark for each individual.

The calculations involved in peer assessment are normally very simple but the detail varies substantially. Each staff member tends to become wedded to his or her particular system and will defend it rigorously.

At a workshop for engineering lecturers in 2003 we demonstrated that three generic types of peer assessment schemes typical of systems we have seen in use give different numerical outcomes when applied to the same mini-project [4]. There were eight teams of between 4 and 5 engineering academics involved in this brief study. Delegates tackled a short project team-task and were then invited to rate each other using the three systems. Steps were taken to provide anonymity and avoid embarrassment.

The three algorithms are described. In each case, the tutor's TEAM mark is varied by student input according to the algorithm. The example data used in each case is that from experimental team 2 who were allocated a team mark of 50%.

Deleted: according to the algorithm being applied

Scheme A – Shared Assessment

Individual team members' marks are the sum of a TEAM mark, determined by the tutor and an INDIVIDUAL mark determined and agreed by the team. A predetermined weighting factor, say 90/10, applies for the two marks.

Student	Team mark(90)		Peer mark(10)	Total
A	45	+	5	=50%
B	45	+	7	=52%
C	45	+	8	=53%
D	45	+	7	=52%

TABLE 1,
EXAMPLE OF SCHEME A: TUTOR MARK = 50% (45/90)

Scheme B - Peer mark distribution

Individual marks are generated by multiplying the TEAM mark by a weighting factor which is distributed amongst the team by the students, who decide on the individual weightings as a group. The maximum weighting factor for distribution is specified, in this case at 25% and the sum of the factors must always equal zero.

Student	Team mark	Factor (max 25%)	Total
A	50	-10%	=45%
B	50	0%	=50%
C	50	+10%	=55%
D	50	0%	=50%

TABLE 2,
EXAMPLE OF SCHEME B: TUTOR MARK =50%

Scheme C - Confidential peer weightings

Each Student completes a separate form in secret.
Marks (out of 5) are solicited against a number of stated criteria e.g. leadership, creativity, work rate etc. Each student allocates marks for themselves and for each of their colleagues.

On analysis, the sum of marks awarded each individual is compared with (divided by) the average for all team members and this reveals a weighting factor. I.e. a student who's marks equal to the average for the team generates a factor of 1 while a high scoring student might have a factor of 1.2 and a low score would reveal, say 0.8. TEAM marks (or an agreed proportion of the team mark) is then multiplied by this factor.

	Criteria 1	Criteria 2	Criteria 3	Calculations:		
Student A	**4**	**5**	**2**			
Student B	4	5	2	Total score for each student:		Team average:
Student C	4	5	5	Sum(A) =	33	Sum(ABC)/4 =41.3
Student D	4	5	2	Sum(B) =	38	
				Sum(C) =	52	
Student A	2	3	2	Sum(D) =	42	
Student B	**3**	**4**	**2**			
Student C	4	5	5	Weighting factors (f):		
Student D	4	5	2	Student A =	33/41.3 = 0.80	
				Student B =	38/41.3 = 0.92	
Student A	3	3	2	Student C =	52/41.3 = 1.26	
Student B	4	4	2	Student D =	42/41.3 = 1.02	
Student C	**4**	**4**	**5**			
Student D	4	5	2	Weighted Mark (factor applied to 50% of team marks)		
				Student A =	45%	
Student A	1	4	2	Student B =	48%	
Student B	2	4	2	Student C =	57%	
Student C	3	4	4	Student D =	51%	
Student D	**2**	**4**	**3**			

TABLE 3,
Example of scheme C: Tutor mark (T) =50%

Analysing the full set of results showed some interesting trends. Scheme A, the shared assessment system give the highest set of individual marks in all but one team. Most players quickly caught on to the fact that it is in everyone's interest to award maximum marks to each student within the peer award. Strong candidates are not disadvantaged even if a free-rider is awarded top marks. The tutor's intended average team mark is effectively inflated and the standard deviation across the team is kept low. Had we allowed a higher proportion of marks to be generated from the peer mark (we used 10%) this effect would have been even more pronounced.

The peer mark distribution technique (scheme B) resulted in a four teams allocating equal marks to all individuals. Experience with this scheme in practice suggests also students tend to take the easy way out rather than enter into conflict. There appears to be a natural instinct to act kindly towards weaker students when faced with a decision form. Nevertheless it is not uncommon for students who have signed a declaration for an equal distribution to subsequently expressed their dissatisfaction with the outcome. Even where unequal efforts are acknowledged, the tendency is to underplay the differences. It is also notoriously difficult to close the circle and ensure all students sign up to this sort of agreement.

Our test results from teams of academics confirms that system B gives the smallest deviations from the norm. Team 8 were the exception who appeared keen to punish backsliders. One must bear in mind, however, that our academics had a particular interested in this topic and they were well aware that, unlike undergraduates, the results held no future consequences for them. One would expect these teams would be more inclined to differentiate between team members.

Scheme C introduced two new features: confidentiality and criteria based marking. The obvious drawback in practice is the increased size of the data set and the complexity of implementation with large student groups. Faced with transcription

and analysis of up to two hundred forms for a large module, any academic might hesitate. Like system B, the tutor's average mark was maintained.

It is reasonable to expect a fairly wide range of abilities and work rate in randomly selected teams which suggests that the larger standard deviation is the more accurate assessment. If the team mark is multiplied directly by the peer assessment factor this system produces a consistently high standard deviation. While trialling system C however, we have exercised caution and it has become the norm to factor only 50% of the team marks which delivers a standard deviation only slightly higher than the mark distribution system (B). Others may choose to apply the effect more dogmatically and arrive at a higher standard deviation. The results presented below for scheme C give the team marks factored 50% where the individual score is given by (0.5 + 0.5f)T.

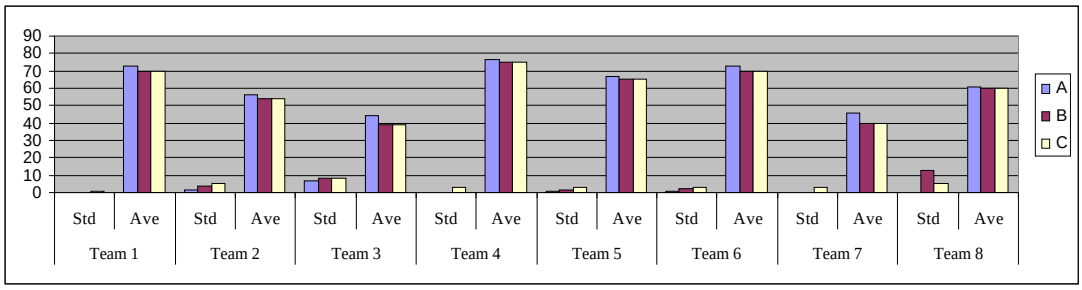

FIGURE 1,
COMPARISON OF AVERAGE MARKS AND STANDARD DEVIATION FOR EACH TEAM.

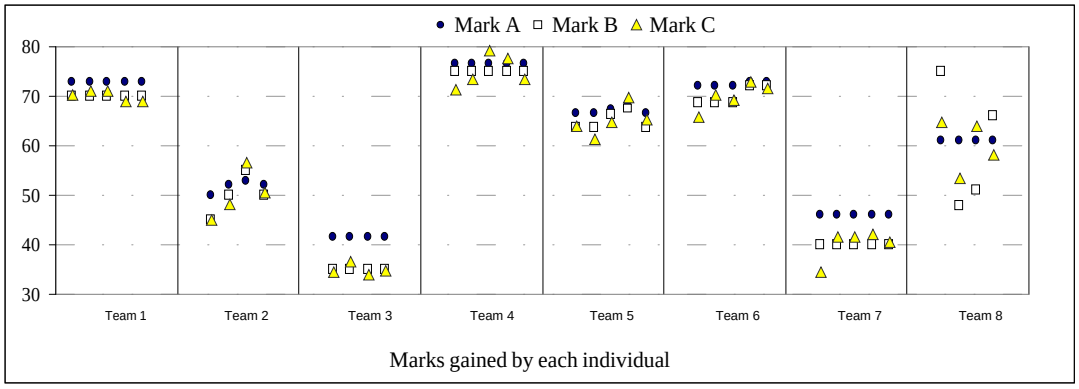

FIGURE 2,
MARKS FOR EACH INDIVIDUAL GENERATED BY THE THREE ALGORITHMS.

Main Workshop Findings

It is clearly desirable to assess against a number of specified criteria in line with the intended learning outcomes of the module and while it might be possible to implement all three assessment systems against multiple criteria, it is likely that for teams to discuss and agree weightings using complicated criteria could well prove frustrating and time consuming. System C is implemented on an individual basis in secret and can be completed very quickly.

The plenary discussion expressed some surprise and concern at the relatively large changes in marks obtained by the same assessors using different algorithms. Most delegates expressed a degree of confidence in system C and liked the less confrontational anonymous submission. There was unanimous distaste for the shared assessment system A.

PEER MARKING EXPERIENCE WITH UNDERGRADUATES

Staff in the school of Mechanical and Manufacturing Engineering have used variants of all three systems with undergraduates and variants of schemes B and C are still in current use. A flexible online version has also been developed with the help of the University's Engineering Education Centre and is now available for use throughout the campus. The software, called 'Web-PA' is based upon an original paper based system similar to scheme C and was developed with a view to making data entry and analysis more convenient and providing flexibility for many types of group assessments. The resource has been used with some success for a variety of projects and assignments from engineering to European Studies and remains under constant development. Web-PA allows the tutor to define any number of assessment criteria or 'form elements' and invites objective marking statements which guide students what performance should be associated with a given mark (see ② in figure 3). The tutor also selects teams online③ and a timeframe for students to enter their data①. Students are required to visit a terminal between specified dates and complete a simple form using pull down menus. Only entry points for their own team members appear on screen and they rate each member in turn, including themselves, against the stated criteria. After the deadline, the tutor can retrieve a complete data set including a peer assessment factor for each student④.

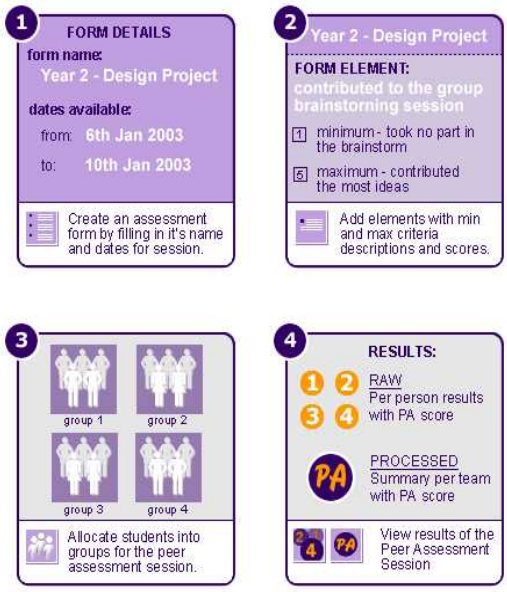

FIGURE 3,
THE FOUR ELEMENTS OF WEB-PA SETUP.

Feedback

Prime considerations in the development of the software was that students should perceive the process to be fair and that the resource should be very flexible and easy to use by both staff and students. The system has been well accepted by those who have used it. Students universally welcome the anonymity and react positively to the openness of the assessment criteria and the fact that they are personally included in the process. Students appear more reassured with data entry to a computer rather than on paper. Rather than positive feedback we have found that the feedback is completely lacking: complaints of unfairness having virtually disappeared where this system is implemented. We plan to build a feedback section into the software in the near future in order to solicit real opinions but the indicators are certainly good.

The operational benefit from the staff point of view is that it makes what would otherwise be a complicated process very simple to implement. Setting up the forms takes only a few minutes and the results are available instantly no matter what the group size. Judging the accuracy of the assessment is not easy but the level of confidence is at least as good as with peer mark distribution systems.

Deleted: operationsl

Deleted: operate.

Analysis of Trends in Self and Peer assessment

It is interesting to separate and compare the self assessments with the peer assessment results. Data was analysed from two separate student cohorts; 61 first year students in 13 groups with 4 to 5 students per group and 102 second year students in 31 groups with 3 to 4 students per group. A comparison of the normalised individual's mark allocated to themselves with the normalised marks averaged from the rest of the group for the same individual produced remarkably similar results across the two cohorts. The results are illustrated for the both years in figure 4.

Deleted: results.¶

Deleted: ¶
****
****¶
¶
¶
THIS IS YOUR BIT ADAM!¶
¶
¶
¶
****
******

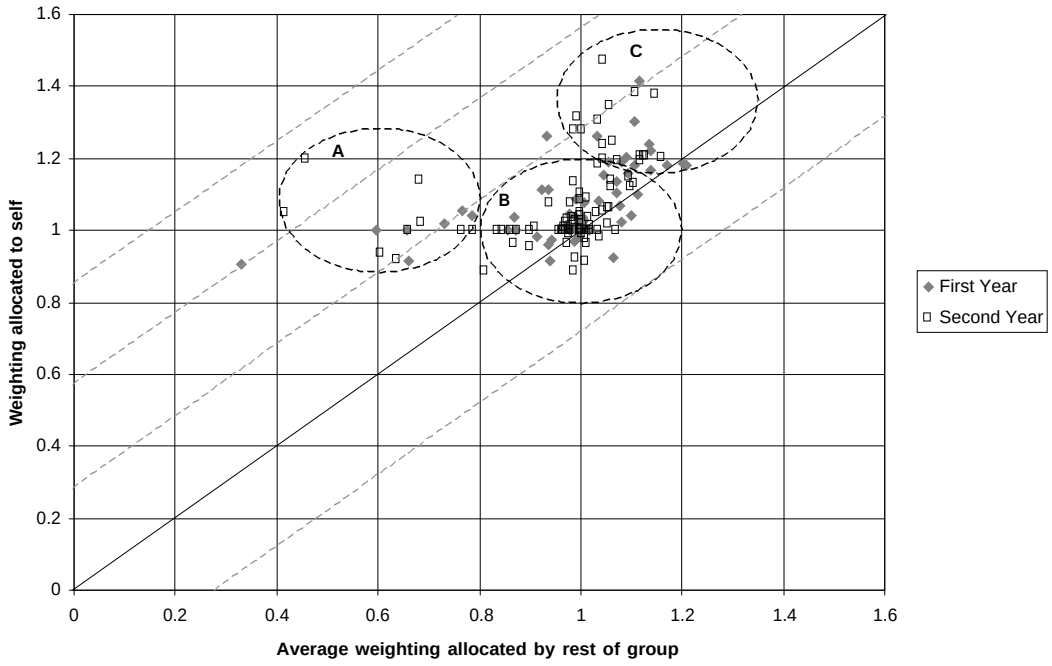

FIGURE 4,
COMPARISON OF PEER ASSESSED WEIGHTINGS ALLOCATED TO INDIVIDUALS BY THEMSELVES AND BY THEIR PEERS FOR 163 INDIVIDUALS IN 44 GROUPS

In figure 4 the solid diagonal line represents y=x, points above this line indicate that the students perceptions of their own input to the group are greater than the perceptions of their input by the other team members. Conversely points below the line indicate where students have a lower perception of their relative input than the rest of the group do. A weighting of 1.0 indicates that the group member undertook a fair share of the group work.

Overall the majority of individuals over-estimate their own input when compared with the assessment of the rest of their group. The plot can be approximately split into three groupings as indicated by A, B and C in figure 4. Group A appear to be the weakest team members, many of whom are poor participants and have failed to engage in the team activity. It is noticeable that they assess their own contribution significantly higher than do the rest of their group. This mechanism boosts a mark that might otherwise be justifiably poor and tends to reduce the standard deviation. Of course there may be the unfortunate student amongst them who has become disenfranchised though not at fault and the effect of the elevated self assessment is to cushion what would otherwise be a harsh assessment. Group B contains the majority of the cohort scoring around 1.0 and this indicated that the individual performed as might have been expected and these students gain a mark close to the tutor's team mark. For these 'average' students there is evidently little difference between the marks allocated by these individuals and by their peers. Group C contains many of the strongest members of the cohort and it is noticeable that although others in the team have marked their contribution above average, the individual has marked their own contribution higher still. Although students recognise the strength of a dominant team member they are naturally keen to protect their own relative position in the team and moderate the extremes.

Comment [a1]: Don't fully understand this : does the safety net mean that the discrepancy acts as a flag to the tutor to look in more detail at this group and talk to them? Or is it that their mark is increased by their own high weighting – this is not very significant in a group of five students.

CONCLUSIONS

- Self and peer assessment can help students reflect on skills, knowledge and the learning process.
- Online Peer assessment allows rapid feedback even with a large number of students.
- Criteria based assessments are desirable.
- Individual private submissions lead to greater harmony.
- Individuals perceptions of their contribution is usually higher than the groups perception of their contribution
- The biggest difference in self and group marks occurs for the weakest group members.
- The strongest group members mark their contribution even higher than the group do.
- A number of groups appear to work harmoniously with equal marks allocated across the group and with similar marks allocated to an individual by them self and the group.

REFERENCES

[1] Edited work by various authors, "Assessment of Individuals in Teams September 2002", *LTSN Working Group Report* . ISBN 1-904190-21-9, available online at http://www.ltsneng.ac.uk

[2] Crawford, A,R, Tennant, J, Choo, B, S, Wilson A (Ed). "A Guide to Learning Engineering through Projects", *FDTL(3) report 43/99 PBLE: Project Based Learning in Engineering* edited works, Sept 2003. http://www.pble.ac.uk

[3] Allen, J, and Lloyd-Jones, R, "The Assessment of Group Work and Presentations in the Humanities", *FDTL(1) report Sheffield Hallam University* http://www.shu.ac.uk/schools/cs/fdtl/grpwk

[4] Willmot, P and Crawford A, R, "Group Assessment Systems", *from LTSN -Eng Workshop " Project and Group Work in Eng ineering", September 2003, Loughborough University* , http://www.ltsneng.ac.uk/nef/events/project_groupwork2.asp