# Self and Peer Rating in Multi-Disciplinary Group Projects

**J. Paul Hermon**
Queen's University Belfast (QUB), N.Ireland

**Stefan Hallström**
Royal Institute of Technology (KTH), Stockholm, Sweden

**Charles D. McCartan**
Queen's University Belfast (QUB), N.Ireland

**Jakob Kuttenkeuler**
Royal Institute of Technology (KTH), Stockholm, Sweden

## ABSTRACT

The ability to self and peer assess performance within a multi-disciplinary group is a valuable skill and one of many detailed in the CDIO syllabus as being desirable for graduate engineers. The design, build and test environment in which a student applies technical knowledge and develops teambuilding and other professional skills such as self and peer assessment is complex and challenging for both students and assessors. Operating such activities as multi-disciplinary and multi-national projects adds further layers of complexity. This paper seeks to examine the reliability of students' self and peer assessments in such complex team environments. The study examined peer ratings from 5 teams from Queen's University Belfast (QUB) and 6 teams from the Royal Institute of Technology (KTH) in Stockholm, and involved in total 105 students from 6 different disciplines. The analysis suggested that students tended to show bias in favour of those from the same discipline and against those from a minority grouping or from outside of their own university. Further analysis from many more institutions would be required to determine if such tendencies are generally typical.

## KEYWORDS

Peer rating, peer assessment, multi-disciplinary, group projects

## BACKGROUND

The need for graduate engineers to be able to work effectively in multi-disciplinary groups has been articulated by many accrediting bodies for several decades and is often cited as a desirable skill by employers. This requirement has been recognised by the CDIO Initiative from the outset and is included in the original CDIO Syllabus under section 3.1, and more specifically in version 2 of the syllabus in section 3.1.5 as "Technical and multi-disciplinary teaming". [1]

There are however a number of barriers to providing authentic experiences, facilitating the development of the required skills in undergraduate programmes. For example, timetabling can be difficult to coordinate between programmes. The allocation of suitable workspaces in a neutral location can often be problematic and preconceived negative stereotyping among students of differing disciplines can lead to groups becoming dysfunctional, generating a poor learning environment. Additionally the skills and experience required to effectively supervise such projects are often only found among a subset of the academic faculty, which can also inhibit widespread provision. The perception that adequate supervision is resource and time intensive is also often a barrier. The authors of this paper have a different standpoint in this matter, based on own experience, but that is not the scope of this paper. Nevertheless, perhaps because of these and other difficulties such multi-disciplinary group projects tend to be the exception rather than the rule at undergraduate level.

Another significant issue is the assessment of the individual in group projects. A range of strategies have been used in both mono-disciplinary and multi-disciplinary courses. Many examiners have adopted peer assessment methods to adjust marks to reflect the contributions of group members by applying multiplication factors or adding or subtracting marks about a group mean [2]. There are even a number of computer based systems which facilitate the automated collection of peer assessments and the feedback of collated scores to the students. Such procedures are not without potential pitfalls however and require careful implementation since the reliability of the students self and peer assessments cannot always be assured. The least able students have been shown to also be the least able at assessing both their own and others' abilities at a task [3]. The reliability of peer assessments has however been seen to improve with repeated use of a process, and most significantly if combined with several instances of formative feedback [4,5]. This would suggest that the use of such processes could be considered more reliable among senior students, provided they have previously had repeated opportunities to develop the required skills in self and peer assessment [6]. Friendships within a group can also lead to reluctance to award low scores to others [7], even if data is made anonymous, and especially if the peer assessment is linked to grades for the project. Often however the group project is a one time experience in the senior year of a programme, without any previous opportunity to develop either peer assessment skills or any of the other professional skills and attributes associated with working in teams. Where self and peer assessment has sought to maximise student learning the advocates recommend getting students involved in the development of the assessment criteria, engaging them in achieving targets related to their own personal learning outcomes [8], and suggest there is a focus on the peer learning aspects and not just the assessed outcomes of the project [9].

The study undertaken in this paper looks at one aspect of the layer of complexity that is added when projects are operated with multi-disciplinary teams in comparison to single discipline groups, specifically that of self and peer assessment.

## CONTEXT (QUB)

The degree programme in Product Design and Development (PDD) at QUB adopted a modified version of an existing peer assessment process already in operation in another programme within the same School. Significantly the link between peer evaluation and assessment was removed so that student assessments were not used to adjust marks for individuals in a group. The process was given a new name "Peer Rating for Feedback" to emphasise this separation from assessment. Students were engaged in a series of "one to one" feedback interviews at the conclusion of each of 5 projects over a 3 year period. Interviews focused on the results of the self and peer evaluations as part of a personal development process, aimed at developing the team working and professional skills aspects of the discipline [10]. This peer rating procedure had been operating successfully since 2008 with groups consisting entirely of PDD undergraduates but a fresh approach was adopted in 2011 to extend the challenge and learning opportunities of the major (capstone) project by establishing multi-disciplinary groups for the first time.

The PDD programme structure has a core of 5 Design-Build-Test group projects, supervised by the same 2 members of faculty, in the first 2 years of the programme. The small cohort size of 20 to 30 students enables project groups to be constructed so that by the end of the second year all students should have experienced working with all other students in their year group. Consequently the supervisors also get to know and understand the students very well. A previous study showed that the repeated experience of a “peer rating for feedback” process improved student skills in self and peer assessment [10].

## OBSERVATIONS (QUB)

The spreadsheet used at QUB for peer rating in group projects has 15 rows relating to skills, attributes and activities. As shown in Table 1, these are divided into 3 key areas of “technical contributions”, “contribution to deliverables” and “collaboration”, which are closely aligned with the learning outcomes for the project, and are set by the project coordinator.

Each student rates their own performance and that of all other group members on a zero mean basis, the total for each row adding to zero. Individual cells can be scored as a real number between -2 and +2 with students being required to supply justifying comments for any row with non zero cells. They are made aware before completing the spreadsheet that their comments will remain confidential and will not be communicated to other members of the group.

Table 1: Typical “Peer Rating for Feedback” spreadsheet criteria

| Technical Contributions | Contributions to Deliverables | Collaboration |
|---|---|---|
| Ability to apply technical knowledge from other modules (including stages 1 & 2) to project | market research | Effectively takes charge of tasks assigned |
| Contribute alternative design concepts | Preparation for interim group presentation | Is fair and even in the treatment of ideas/solutions put forward by other group members |
| Sourcing of relevant technical information | Writing of interim group report | Produces work on time |
| Demonstrate an ability to apply critical thinking | Construction of concept prototype | Willing to take on tasks |
| Effectively troubleshoot problems and find answers | Design (sketches, CAD etc.) | Communicates clearly with other members of the team |

While the peer rating procedures ask the students to score group members for relative contribution in each of 15 categories it is also possible to produce rankings within the group on the basis of the overall totals for each student. As can be seen in Figure 1, groups of PDD students provide rankings more similar to the rankings of the tutor as each year progresses. By the end of year 2 the average student rankings for the group are similar to the “expert” rankings of the supervisors. By the end of stage 3 there is a very close alignment of the group average, individual and tutor rankings. It is also worth noting an increased engagement in the process and a reduced tendency to rate individuals equally in senior years.

| | | | | | |
|---|---|---|---|---|---|
| tutor | 1 | 2 | 3 | 4 | 5 |
| group | 1 | 4 | 3 | 2 | 5 |
| # 1 | 1 | 5 | 1 | 1 | 4 |
| # 2 | 3 | 2 | 3 | 1 | 5 |
| # 3 | 1 | 1 | 3 | 4 | 5 |
| # 4 | 1 | 4 | 3 | 2 | 5 |
| # 5 | 1 | 1 | 1 | 1 | 1 |
| end of stage 1 - 2009 | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| tutor | 1 | 2 | 3 | 4 | 5 |
| group | 1 | 3 | 2 | 5 | 4 |
| # 1 | 1 | 4 | 2 | 5 | 3 |
| # 2 | 1 | 2 | 3 | 5 | 4 |
| # 3 | 1 | 1 | 1 | 1 | 5 |
| # 4 | 1 | 3 | 5 | 1 | 3 |
| # 5 | 1 | 4 | 2 | 5 | 3 |
| end of stage 2 - 2010 | | | | | |

| | | | | | |
|---|---|---|---|---|---|
| tutor | 1 | 2 | 3 | 4 | 5 |
| group | 1 | 2 | 3 | 4 | 5 |
| # 1 | 1 | 2 | 3 | 5 | 4 |
| # 2 | 1 | 2 | 3 | 4 | 5 |
| # 3 | 1 | 2 | 3 | 4 | 5 |
| # 4 | 1 | 1 | 3 | 4 | 5 |
| # 5 | 1 | 2 | 3 | 3 | 5 |
| end of stage 3 - 2011 | | | | | |

Figure 1. Comparison of typical PDD student and tutor rankings in group projects

While these results appear to show a positive improvement in self and peer assessment skills it is possible that this effect is simply due to familiarity with the cohort, or tutor, or rating methods used. If these skills have genuinely been developed then it would be reasonable to expect that the assessment of strangers or those from another discipline would be done with the same level of aptitude and reliability.

In 2011 a change in the structure of the major group projects in Mechanical Engineering and Product Design and Development was implemented with the objective of introducing a multi-disciplinary aspect to the group work, in line with CDIO syllabus item 3.1.5 "Technical and multi-disciplinary teaming". 5 groups were formed with 2 or 3 PDD students partnered with 2 Mechanical Engineering students (MEE). Additionally there were 4 Erasmus exchange students (EU), studying in Belfast for 1 semester, as part of a Masters in Industrial Design Engineering at Chalmers University of Technology in Gothenburg, who were also included adding a further dimension to the composition of 4 of the 5 teams. There were 2 academic supervisors assigned to each team and in total 7 different supervisors across the 5 teams, all of whom had a primary degree in Mechanical Engineering. At the end of 12 weeks of the project all students completed a "peer rating for feedback" spreadsheet.

| G1 | | | | | | |
|---|---|---|---|---|---|---|
| Tutor | 1 | 2 | 3 | 4 | 5 | 6 |
| Group | 3 | 2 | 1 | 4 | 5 | 6 |
| PDD#3 | 1 | 2 | 3 | 3 | 5 | 6 |
| PDD#4 | 1 | 2 | 2 | 4 | 5 | 6 |
| PDD#5 | 1 | 5 | 2 | 4 | 3 | 6 |
| MEE#3 | 1 | 2 | 3 | 4 | 5 | 6 |
| MEE#4 | 6 | 2 | 1 | 2 | 4 | 4 |
| EU#2 | 3 | 2 | 1 | 4 | 4 | 6 |

| G2 | | | | | | |
|---|---|---|---|---|---|---|
| Tutor | 1 | 2 | 2 | 4 | 5 | 6 |
| Group | 1 | 3 | 6 | 2 | 4 | 5 |
| PDD#6 | 1 | 5 | 4 | 2 | 3 | 5 |
| PDD#7 | 2 | 4 | 6 | 1 | 3 | 5 |
| PDD#8 | 1 | 2 | 3 | 6 | 3 | 5 |
| MEE#5 | 2 | 1 | 6 | 5 | 3 | 3 |
| MEE#6 | 1 | 3 | 6 | 2 | 5 | 3 |
| EU#3 | 1 | 5 | 2 | 3 | 6 | 4 |

| G3 | | | | | | |
|---|---|---|---|---|---|---|
| Tutor | 1 | 2 | 3 | 4 | 5 | 6 |
| Group | 2 | 1 | 3 | 3 | 5 | 4 |
| PDD#9 | 5 | 1 | 2 | 3 | 3 | 6 |
| PDD#10 | 5 | 2 | 2 | 1 | 6 | 4 |
| PDD#11 | 2 | 1 | 3 | 3 | 6 | 5 |
| MEE#7 | 1 | 4 | 3 | 2 | 6 | 5 |
| MEE#8 | 1 | 2 | 3 | 3 | 6 | 3 |
| EU#4 | 2 | 1 | 4 | 5 | 6 | 3 |

Figure 2. Tutor, group average and peer rankings from multidisciplinary groups (Dec 2011)

As can be seen in Figure 2, the peer rankings showed less alignment with the tutor rankings than previous cohorts of PDD students at the same stage (year) of study who had been working in single discipline groups. In some cases there was total disagreement with the tutors such as in groups G1 and G3 where students MEE#4, PDD#9 and PDD#10 ranked last the student whom the supervisors had rated most highly, while at the same time other students had agreed with the tutors' assessment and matched first choices in the group. The MEE and EU students where less familiar with the peer rating process and it would be expected that this lack of experience would make them less likely to agree with the tutor ratings. The PDD students' ratings were somewhat surprising, differing significantly from the trends of previous cohorts, and prompted further investigation.

Table 2: Normalised ratings for QUB multi-disciplinary groups (Dec 2011)

| | PDD#1 | PDD#2 | PDD#3 | PDD#4 | PDD#5 | PDD#6 | PDD#7 | PDD#8 | PDD#9 | PDD#10 | PDD#11 | PDD#12 | PDD#13 | PDD#14 | MEE#1 | MEE#2 | MEE#3 | MEE#4 | MEE#5 | MEE#6 | MEE#7 | MEE#8 | MEE#9 | MEE#10 | EU#1 | EU#2 | EU#3 | EU#4 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PDD#1 | 0.00 | 0.25 | | | | | | | | | | | | | 1.06 | -0.38 | | | | | | | | | -0.94 | | | |
| PDD#2 | 0.67 | 0.89 | | | | | | | | | | | | | -0.78 | 0.33 | | | | | | | | | -1.11 | | | |
| PDD#3 | | | 0.28 | 0.63 | -0.89 | | | | | | | | | | | | 0.28 | -1.40 | | | | | | | | 1.10 | | |
| PDD#4 | | | 0.25 | 0.42 | -0.28 | | | | | | | | | | | | 0.42 | -1.58 | | | | | | | | 0.77 | | |
| PDD#5 | | | 0.17 | 0.09 | 0.38 | | | | | | | | | | | | 0.43 | -1.56 | | | | | | | | 0.52 | | |
| PDD#6 | | | | | | 0.04 | 0.15 | 1.33 | | | | | | | | | | | -0.67 | -0.67 | | | | | | | -0.19 | |
| PDD#7 | | | | | | 0.40 | 1.20 | 0.80 | | | | | | | | | | | -0.40 | -0.80 | | | | | | | -1.20 | |
| PDD#8 | | | | | | 0.07 | -0.87 | 1.13 | | | | | | | | | | | 0.27 | -0.67 | | | | | | | 0.07 | |
| PDD#9 | | | | | | | | | 0.00 | 0.00 | 0.50 | | | | | | | | | | 1.00 | -1.00 | | | | | | -0.50 |
| PDD#10 | | | | | | | | | -1.35 | 0.65 | 0.48 | | | | | | | | | | 0.48 | -0.10 | | | | | | -0.16 |
| PDD#11 | | | | | | | | | -1.55 | 0.35 | 0.35 | | | | | | | | | | 0.45 | 0.00 | | | | | | 0.40 |
| PDD#12 | | | | | | | | | | | | 0.13 | 0.38 | 0.63 | | | | | | | | | -1.38 | 0.25 | | | | |
| PDD#13 | | | | | | | | | | | | -0.80 | 0.69 | 0.91 | | | | | | | | | -1.09 | 0.29 | | | | |
| PDD#14 | | | | | | | | | | | | 0.30 | 0.50 | 0.50 | | | | | | | | | -1.50 | 0.20 | | | | |
| MEE#1 | 0.00 | 0.00 | | | | | | | | | | | | | 0.00 | 0.00 | | | | | | | | | 0.00 | | | |
| MEE#2 | -1.00 | 1.00 | | | | | | | | | | | | | 1.00 | -1.00 | | | | | | | | | 0.00 | | | |
| MEE#3 | | | 0.32 | 0.42 | -0.05 | | | | | | | | | | | | 0.37 | -1.58 | | | | | | | | 0.53 | | |
| MEE#4 | | | 0.50 | 0.50 | -0.50 | | | | | | | | | | | | 1.00 | -0.50 | | | | | | | | -1.00 | | |
| MEE#5 | | | | | | 0.00 | -0.67 | 0.67 | | | | | | | | | | | 1.33 | 0.00 | | | | | | | -1.33 | |
| MEE#6 | | | | | | -0.67 | 1.00 | 1.33 | | | | | | | | | | | -0.33 | -0.33 | | | | | | | -1.00 | |
| MEE#7 | | | | | | | | | -1.39 | 0.61 | 0.17 | | | | | | | | | | 0.09 | -0.17 | | | | | | 0.69 |
| MEE#8 | | | | | | | | | -1.60 | 0.20 | 0.20 | | | | | | | | | | 0.40 | 0.20 | | | | | | 0.60 |
| MEE#9 | | | | | | | | | | | | -1.15 | -0.13 | 0.04 | | | | | | | | | 0.39 | 0.85 | | | | |
| MEE#10 | | | | | | | | | | | | -0.29 | -0.29 | -0.29 | | | | | | | | | -0.57 | 1.43 | | | | |
| EU#1 | 0.30 | -1.40 | | | | | | | | | | | | | 0.50 | 0.30 | | | | | | | | | 0.30 | | | |
| EU#2 | | | -0.33 | 0.21 | -0.33 | | | | | | | | | | | | 1.15 | -0.85 | | | | | | | | 0.15 | | |
| EU#3 | | | | | | -0.52 | -0.22 | 1.48 | | | | | | | | | | | -0.41 | -0.30 | | | | | | | -0.07 | |
| EU#4 | | | | | | | | | -0.89 | -0.67 | -0.56 | | | | | | | | | | 1.11 | 0.00 | | | | | | 1.00 |
| average | -0.01 | 0.15 | 0.20 | 0.38 | -0.28 | -0.11 | 0.10 | 1.12 | -1.13 | 0.19 | 0.19 | -0.36 | 0.23 | 0.36 | 0.36 | -0.15 | 0.61 | -1.25 | -0.04 | -0.46 | 0.59 | -0.18 | -0.83 | 0.60 | -0.35 | 0.35 | -0.62 | 0.34 |
| TUTOR | 0.06 | 0.30 | 0.04 | 0.30 | -0.15 | -0.36 | -0.27 | 1.45 | -0.69 | -0.62 | -0.08 | -0.71 | -0.71 | 1.29 | 0.19 | -0.97 | 0.15 | -1.20 | -0.09 | -0.55 | 0.92 | -0.85 | -0.43 | 0.43 | 1.03 | 0.79 | -0.09 | 1.15 |

Table 2 shows the self and peer rating values for each student in each of the 5 groups. The values have been normalised for each individual so that the total range of the marks allocated is 2 units. A negative value still indicates a below average contribution in the group. From this the average rating of students from each discipline for their own and each of the other 2 disciplines has been calculated as shown in Table 3 and Figure 3.

Table 3: Average ratings by discipline for QUB multi-disciplinary groups (Dec 2011)

| | PDD | MEE | EU |
|---|---|---|---|
| PDD | 0.24 | -0.30 | -0.11 |
| MEE | -0.04 | 0.13 | -0.19 |
| EU | -0.27 | 0.19 | 0.35 |
| Tutor | -0.01 | -0.24 | 0.72 |

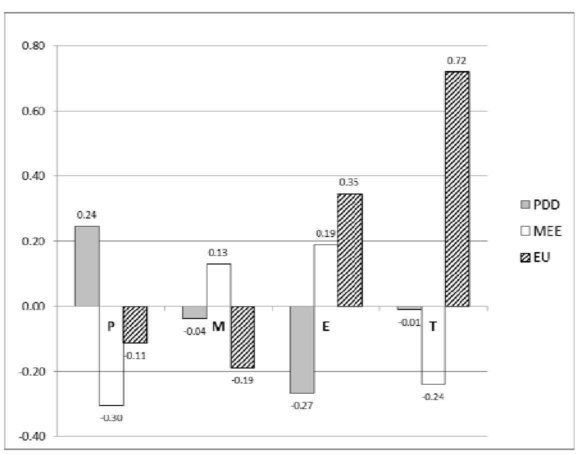


Figure 3. Average ratings by discipline for QUB multi-disciplinary groups (Dec 2011)

The results show that the students from all 3 disciplines tend to rate themselves or other students from the same discipline highest.

## CONTEXT (KTH)

In a final year cross-disciplinary design-implement course at KTH, peer assessment has been used for more than a decade as a means of providing both formative feedback during the course, and for summative feedback and input to the final assessment at the end of the course. Groups typically comprise students from Lightweight Structures (LS) and Naval Architecture (NA) programmes within Applied Mechanics but one group (2007-8) was also formed by LS students partnered with students specialising in wood technology (WT). The students are asked to assess the work and performance of their peers, as well as the quality of their own work and performance, twice during the course.

Students from two different specialisations are grouped together in larger project groups, typically 10-15 students in each. The two student categories are however often of dissimilar size, generating various levels of majority-minority constellations within the groups. By the end of the first semester the students write formative feedback to each other and are asked to suggest forecasts of grades to themselves as well as their peers, based on their achievements with respect to the course goals. By the end of the course they write feedback to each other again and suggest final grades for themselves and their peers. How the instructors finally set the grades in the course is another story, not addressed here.

Similarly as at QUB the learning outcomes and assessment criteria have been predefined by the course tutors. One difference, however, is that assessment is specifically addressed as a learning objective in the course at KTH. The students are supposed to demonstrate that they are able to assess their own work and work of others and, as a consequence, they are introduced to the challenges of assessment through class workshops on the matter. Since they have to perform assessment they need to consider how to interpret, relate to, and apply the assessment criteria, and they are trained to do that by means of exercises and discussions with their peers and the tutors. A main difference between the Swedish system and many others is that it strictly uses criterion-based grading. Obviously the interpretation of criteria can differ among tutors and institutions but the grade of one student is never depending on the relative success or failure of another student, as they might do in normative-based grading systems.

The groups at KTH are also significantly larger, being of between 10 and 15 students compared to the 5 or 6 students comprising a group at QUB. The data analysed represents 6 groups over 4 years between 2005 and 2011, and a total of 77 students.

## OBSERVATIONS (KTH)

In order to make a direct comparison with the QUB data the deviation between a specific student's grade and the average grade awarded by that student for the group has been divided by half the range the student used in his/her grading. This gives the same style of normalised values with a range of 2 about a mean of zero. The average ratings by discipline as shown in Table 4 exhibit similar tendencies for bias towards students from the same discipline as witnessed at QUB. The trend is however very much weaker with both the LS and NA students making broadly similar assessments of performance against the criteria.

The numbers of students in the groups is heavily weighted in favour of the Naval Architecture discipline (48 NA versus 17 LS). The learning objectives and the assessment criteria are common for the two groups – being stated at a general level, rather than topic-specific. Nevertheless, there seems to be a tendency towards students from the majority group ending up with better ratings than minority groups. The reason might be that the majority is more likely to set the agenda for the group as a whole, and thus might consider some of "their" activities more important than those of the other group. It is also clear that if there is tendency to rate students from the same discipline higher, such unbalance would benefit a majority group and be unfavourable for students in a minority group.

Table 4: Average ratings by discipline for KTH multi-disciplinary groups (2005 - 2011)

| | LS | NA |
|---|---|---|
| LS | -0.11 | 0.03 |
| NA | -0.14 | 0.07 |

While there appears to be little disciplinary bias when averages are calculated across all groups it was noted however that minorities tended to receive lower ratings when individual groups are analysed. The data in Table 5 shows the composition of each group (the number following the colon of the 2 letter discipline acronym e.g. group 1 in 2005-6 was made up from 3 Lightweight Structures students and 10 Naval Architects). In groups 1, 2, 3 & 5 where there was a significant majority the students from the majority discipline received a higher average rating. Group 6 is an exception but when the data is broken down into KTH and international students the 2 internationals can be seen to receive significantly lower ratings, which also skews the disciplinary ratings somewhat.

Table 5: KTH normalised group ratings by discipline

| 2005-6 (1) | LS: 3 | NA:10 |
|---|---|---|
| | | |
| | LS | NA |
| LS | -0.18 | 0.05 |
| NA | -0.38 | 0.11 |
| Average | -0.34 | 0.10 |

| 2007-8 (2) | LS: 4 | NA: 10 |
|---|---|---|
| | | |
| | LS | NA |
| LS | -0.52 | 0.14 |
| NA | -0.19 | 0.05 |
| Average | -0.26 | 0.07 |

| 2007-8 (3) | LS: 9 | WT: 3 |
|---|---|---|
| | | |
| | LS | WT |
| LS | 0.20 | -0.60 |
| WT | 0.15 | -0.44 |
| Average | 0.19 | -0.56 |

| 2008-9 (4) | LS: 5 | NA: 5 |
|---|---|---|
| | | |
| | LS | NA |
| LS | -0.20 | 0.08 |
| NA | -0.28 | 0.17 |
| Average | -0.24 | 0.13 |

| 2011-12 (5) | LS: 3 | NA: 12 |
|---|---|---|
| | | |
| | LS | NA |
| LS | 0.13 | -0.03 |
| NA | -0.11 | 0.03 |
| Average | -0.06 | 0.01 |

| 2011-12 (6) | LS: 2 | NA: 11 |
|---|---|---|
| | | |
| | LS | NA |
| LS | 0.50 | -0.09 |
| NA | 0.50 | -0.09 |
| Average | 0.50 | -0.09 |

| 2011-12 (6) | KTH: 11 | Internat.: 2 |
|---|---|---|
| | | |
| | KTH | International |
| KTH | 0.19 | -1.02 |
| International | 0.07 | -0.40 |
| Average | 0.17 | -0.93 |

It can also be seen from Table 5 that in 4 of the 6 groups (1, 3, 4 & 5) the average ratings for students from the same discipline was higher than that given by the students of the other discipline. In other words, their opinions were biased in favour of their own "species", as observed at QUB.

## DISCUSSION

The authors recognise that the data presented is not sufficient to draw conclusions but suggest that the issues raised merit further discussion.

The results suggest that bias exists by discipline and nationality at QUB and by majority rule, nationality and to a lesser extent by discipline at KTH. The possibility exists that such bias may occur in all such multi-disciplinary projects. The nature of such projects however makes it a lengthy process for one institution to collect sufficient longitudinal data. The retrospective analysis done on the data collected over a number of years at KTH could be carried out by other CDIO collaborators, or other institutions involved in similar multi-disciplinary group projects, to see if such biases are more widespread.

Irrespective of the typical nature of such biases it is incumbent on the coordinators of such projects to consider setting assessment criteria which are fair and equitable to all participants. Is it appropriate that all the same learning outcomes and assessment criteria are applied to students from different disciplines? How can the desirable richness and diversity brought by international students to such projects be best accommodated? If mixing disciplines and nationalities, should this be done only in equal subgroups? To fully evaluate such parameters would require parallel studies with control groups to measure the effects, but such an approach would in itself deliberately create inequality and potentially disadvantage some students.

There were also slightly different language barriers in the groups at QUB and KTH. At QUB native English speaking students were mixed with students having English as a second language. The course at KTH is given in English, including all communication in writing and at meetings. Nevertheless and needless to say a group of Swedish students would typiclly not discuss matters in English unless international students participated in the discussions. Since the international students did not speak Swedish at all, they could not easily overhear discussions between Swedish students and spontaneously join in. This is just one example of possibly many disadvantages minorities could face in a lingua franca course setting, and it would be difficult to avoid unless the students handled the language issues with great awareness.

The setting of the learning outcomes and the peer assessment / rating criteria are handled quite differently at QUB and KTH. The scoring done by the students is also different in that considering absolute rather than relative performance may produce a different distribution of grades. Both methods are also susceptible to the overestimation effect described by Kruger and Dunning [3], particularly by the least experienced and least able. As has been demonstrated at QUB the mixing of students who are experienced in the peer rating process with those inexperienced and with others new to the university moved the reliability of the peer ratings further away from the ratings of the tutors, who might be considered the experts. An interesting side note is that all 7 of the QUB tutors had primary degrees in Mechanical Engineering yet scored the MEE students lowest of the 3 disciplines across the groups.

The perceived progress of the PDD students at QUB in developing self and peer assessment skills did not prove to be sufficient preparation for them to effectively assess the contributions made by those from another discipline. While the small cohort and continuity of supervisors has the advantage of building a close community of practice the size of the change to a multi-disciplinary group is clearly a significant one from the student perspective.

## RECOMMENDATIONS

While involving students in discussing learning outcomes and interpreting assessment criteria has several positive effects on student motivation and engagement in assessment activities it does not prevent a minority of students becoming disadvantaged if not moderated in a multi-disciplinary or multi-national project context. At KTH, students working in such groups become engaged in the assessment process, and found out first-hand that even if there exist assessment criteria it is not easy to judge work that might be diverse in content, quantity and quality in a totally consistent way. The fact that the course objectives are the same for all allows students from diverse backgrounds and disciplines to work on a group project without compromise or disadvantage from that particular point of view. On the other hand it is essential that the objectives truly apply to all disciplines represented in the project.

If it is accepted that the ability to self and peer assess those from a different discipline is a skill which can be learned then it is reasonable to accept that students will better develop this skill if given repeated opportunities to practice. Thus the introduction of multi-disciplinary projects in the junior years of a degree would provide such opportunities. Such experiences may also prevent any negative disciplinary stereotyping that could develop due to lack of familiarity with students on a different programme. It is noted that working in a project group represents a very different learning experience from being enrolled in more traditional lecture style classes which are commonly co-taught across engineering disciplines in subjects such as the core engineering sciences.

The authors recommend that anyone involved in such projects consider retrospectively examining self and peer assessment scores to determine if similar biases exist in their specific contexts.

## REFERENCES


[1] Crawley E.F., Malmqvist J., Lucas W.A., Brodeur, D.R., "The CDIO Syllabus v2.0: An Updated Statement of Goals for Engineering Education", Proceedings of the 7th International CDIO Conference, Technical University of Denmark, June 20-23, 2011.

[2] Loddington S., "Peer assessment of group work: a review of the literature:, JISC WebPA Project, May 2008, http://webpa.ac.uk/files/Literature%20review%207_0.doc (accessed April 2011)

[3] Kruger, J., Dunning, D., "Unskilled and Unaware of It: Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments", Personality and Social Psychology, Vol 77, No 6, (1999) p1121-1134

[4] Brown-Parker, J., Thomas, I. & Wellington, P,, "Peer Assessment of Student Performance: Measuring Congruency of Perceptions in a Multidisciplinary Team", Proceedings of: Advancing International Perspectives, (1997), Adelaide, South Australia

[5] Oldfield, K.A. & Macalpine, J.M.K., "Peer and Self-assessment at Tertiary Level" Assessment and Evaluation in Higher Education, Vol 20, No 1 (1995), p125-132

[6] Norton, S.M., "Peer assessments of performance and ability: An explanatory meta-analysis of statistical artefacts and contextual moderators", Journal of Business and Psychology, Vol 6, No 3 (1992), p387-399

[7] Falchikov, N., "Peer Feedback Marking: Developing Peer Assessment", Innovations in Education and Teaching International, Vol 32, No 2 (1995), p175-187

[8] Boud, D., "Assessment and the promotion of academic values", Studies in Higher Education, Vol 15, No 1 (1990), p101-111

[9] Willey, K., Gardner, A., "Developing team skills with self and peer assessment", Campus-Wide Information Systems, Vol 26, No 5 (2009), p365-378

[10] Hermon J.P., McCartan C.D., "Peer Rating for Feedback in Group Projects", Proceedings of the 7th International CDIO Conference, Technical University of Denmark, Copenhagen, June 20-23, 2011

### *Biographical Information*

J. Paul Hermon is a Senior Teaching Fellow in the School of Mechanical and Aerospace Engineering at Queen's University Belfast. He holds a MEng Degree in Mechanical and Manufacturing Engineering (QUB 1987) and is Programme Director for the BEng and MEng Product Design and Development degrees. He is a Fellow of the Higher Education Academy and also a Fellow of the Royal Society for the encouragement of Arts, Manufactures and Commerce.

Dr. Charles D. McCartan is a Senior Teaching Fellow in the School of Mechanical and Aerospace Engineering at Queen's University Belfast. His current scholarly interests include developing applying and evaluating active and interactive learning methods, teaching mathematics to engineers, first year introductory courses and the assessment of group projects. In addition, he is a professional engineer with experience in industry, research and consultancy. He is a member of the Society of Automotive Engineers (SAE) and a Fellow of the Higher Education Academy.

Dr. Stefan Hallström is an associate professor in Lightweight Structures at the Royal Institute of Technology (KTH) in Stockholm. He has more than 15 years teaching and course development experience and has been active in several educational development projects at KTH. He and Dr Kuttenkeuler have both been active in the CDIO initiative since the start, and they have as part of the activities developed and run a design-build project course concept together for more than 10 years.

Dr. Jakob Kuttenkeuler is an assoiciate professor in Naval Architecture at the Royal Institute of Technology (KTH) in Stockholm. He has developed the KTH Master program in Naval architecture and teaches a range of topics spanning over propeller theory, hydromechanics, maneuvering, ship design and sailing mechanics. He received the KTH Education Prize in 2001.

Corresponding author:
J Paul Hermon
School of Mechanical and Aerospace Engineering
Queen's University Belfast
Ashby Building
Stranmillis Road
Belfast BT9 5AH
+44 (0)28 9097 4499
p.hermon@qub.ac.uk