IDENTIFICATION OF INEFFECTIVE TEAM MEMBERS USING NORMALIZED PEER RATINGS

Michel F. Couturier and Guida Bendrich
Department of Chemical Engineering, University of New Brunswick
cout@unb.ca

Abstract – Peer assessments are used in the senior process design course at the University of New Brunswick to determine individual grades, reduce free riding and improve team dynamics. The results of the monthly surveys are shared with students as formative feedback on their performance and used to calculate the relative performance of team members by dividing the overall average score of each student by the team average. This study examined whether the resulting normalized peer ratings could be used to identify ineffective team members. We found that the normalized ratings of effective team members follow the normal distribution with a mean of unity and a standard deviation of about 0.04. The small monthly variations in the standard deviation of the distribution are not statistically significant indicating that the characteristics of the distribution are time-invariant in our course. Students who obtain normalized ratings less than 0.9 are generally ineffective team members and their ratings should not be used in the calculation of the team average to avoid inflating the normalized ratings of their teammates.

Keywords: peer evaluation, team project, assessment factor, free-riding, team member performance

1. INTRODUCTION

Many engineering courses today contain team projects to help develop teamwork, leadership and communication skills. Students are divided into groups of three or more individuals and assigned a task designed to generate a product which is graded by the instructor. The challenge of the instructor is then to derive individual grades from the product grade when all tasks are completed in a team setting outside of class [17]. The easiest approach is to give everyone on the team the same grade. This is however considered an unethical academic practice which demotivates individuals and encourages free-riding [12,17]. Free-riding or hitchhiking occurs when one or more team members do not do their fair share of the work and ride on the efforts of others [2,8]. Students who have had unpleasant experiences with team projects often cite free-riding as the main reason for their frustration and anger [7].

The grading and free-riding issues can be effectively overcome by using peer evaluations to derive individual grades and increase students’ accountability [2,13]. The peer assessment is used to develop a multiplying factor for each team member which is then used to compute individual grades as the product of the multiplier and the team product grade [3,8,9]. The multiplier is usually the student’s individual average rating divided by the team average [8,9,13], but can also be the square root of this number [6,8] or this number scaled by a weighting factor [6] if the instructor wishes to give less weight to the peer rating. This approach combines the assessment of the product by the instructor and the assessment of the process by the students in the computation of individual marks [9,17].

There are two main types of peer evaluation tools [10]: holistic and category-based. In the category-based assessment, students score teammates on several criteria whereas a single score that reflects a global judgement is assigned to each team member in the holistic approach. The criteria in the category-based approach can be task-related categories, team work functions or a combination of the two [10]. If a major goal of the group project is for students to work as a team and for each team member to contribute equally as far as possible, it is better to use criteria that stress teamwork skills over academic ability [15]. To discourage students from deliberately giving lower rates to their peers to score higher marks [3], it is important to keep students accountable to their peers by sharing average peer rating results with students. This practice has the added benefit of providing feedback to students on their performance as team members. In this respect, evaluation instruments that contain specific criteria not only provide team members with more effective feedback than global measures of performance but also provide individuals with clearer indications of desirable behavior [2].

Peer ratings also provide a mechanism for identifying poorly performing team members [3,8]. Students who are perceived as free riders normally receive lower ratings because they are a source of frustration within the team.
Free riding can be detected before it does much damage by performing student evaluations early in the course and at multiple points during the project [2]. The literature on the possible reasons for free-riding behavior suggests that it is not necessarily a deliberate attempt to do as little work as possible [7,15]. Numerous underlying issues (like domineering team members, team members with widely different goals or academic abilities) may prevent a student from completing work to a level considered acceptable by fellow team members [11]. By detecting ineffective team members early, it may be possible to get the team back together after the underlying issues have been identified [1].

One of the issues with the multiplier method for adjusting individual grades is that best-performing students from dysfunctional teams can receive highly inflated grades because one of their teammates received very low ratings [3,9]. The solution commonly proposed for this problem is to set a maximum deviation from the group mark of about 10% [3,8,17]. This is however an arbitrary and debatable practice.

Peer assessments are used in the senior process design course at the University of New Brunswick to determine individual grades, reduce free riding and improve team dynamics. Students evaluate their teammates at the end of each month by filling a survey containing questions on several team-member attributes. The results of the survey are used to calculate the relative performance of team members by dividing the overall average score of each student by the team average. This study examined whether the resulting normalized peer ratings could be used to identify ineffective team members such as free riders. The premise was that ineffective teammates likely get normalized ratings which fall outside the distribution of ratings obtained by effective teammates. The hypothesis was tested using the results of the peer assessment surveys performed during the last four offerings of the course. We found that the normalized peer ratings of effective students follow the normal distribution with a mean of unity and a standard deviation of 0.04. This finding provides a rigorous statistical method for identifying ineffective team members and for adjusting the individual grades of students in dysfunctional teams.

2. ASSESSMENT METHOD

2.1. Course Structure

A collaborative approach has been used for teaching the full-year capstone design course ChE 4225 Chemical Plant Design in the Chemical Engineering Department at the University of New Brunswick since 2010. The enrolment in this course is between 50 and 60 students per year. The students are divided into teams of four students and must complete the preliminary design of a chemical process. A few teams of five students are created when the total course enrolment is not an integer multiple of four. Students choose their teammates and the project on which they prefer to work. All design projects are sponsored by local companies and two teams are assigned per project. The two teams work independently and produce separate reports.

Each team is co-mentored by a faculty member and a practicing engineer. The role of the co-mentors is that of coaches and counsellors who must motivate and guide the students. The course coordinators (M. Couturier and G. Bendrich) ensure that all groups are progressing at the same rate and are evaluated against the same criteria by setting eight evenly-spaced milestones (four per term) that all teams must complete. The outline for each milestone defines the tasks that need to be accomplished and sets the marking scheme that will be used by the co-mentors to evaluate the deliverables.

Design review meetings take place with the clients in November and February. These meetings start with an oral presentation by the students to update their client on their progress and conclude with questions and feedback from the client. The design work of the students is also showcased at our annual Engineering Design Symposium held in early April.

More information about the course and the universal milestone framework used to coordinate the efforts of students and co-mentors is available in a previous publication [5].

2.2. Individual Grades

The breakdown of the final grade of the students is provided in Table 1. The final report has the largest weighting because it is the main product of the design exercise and the only document given to the client. Each milestone report is worth 5% of the overall mark (Table 1).

<table>
<thead>
<tr>
<th>Deliverable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Final report</td>
<td>45%</td>
</tr>
<tr>
<td>2. First 7 milestones</td>
<td>35%</td>
</tr>
<tr>
<td>3. 3 presentations</td>
<td>20%</td>
</tr>
</tbody>
</table>

The team grades on the milestones and final report are adjusted for individual performances using the following equation.

Individual Mark = CF₀ x Group Mark

The correction factor CF₀, which can be greater or smaller than unity, is derived from the teammate and co-mentor evaluations and the participation of the student in the course.
\[ CF_o = 0.2 \times CF_{\text{participation}} + 0.2 \times CF_{\text{co-mentor}} + 0.6 \times CF_{\text{peer}} \quad (2) \]

The peer evaluations collected at the end of each month are used to calculate the normalized peer assessment factor \( CF_{\text{peer}} \) as follows,

\[ CF_{\text{peer}} = \text{Student’s evaluation by peers in } \% \times (3) \]

Average of peer evaluations for the group in \( \% \)

The calculation of \( CF_{\text{co-mentor}} \) and \( CF_{\text{participation}} \) and the use of \( CF_o \) to adjust individual marks were previously described by Couturier and Bendrich [4].

### 2.3. Peer Evaluations

Each team member evaluated his/her teammates at the end of each month using a survey developed in Microsoft Excel. The survey used in November and March measured the sixteen team member dimensions listed in Table 2. The dimensions are divided into five general areas as suggested by Ohland et al. [16] and attempt to measure the quality and quantity of individual work as well as the teamwork skills of students. A shorter survey comprising ten dimensions was used at the end of September, October, January and February by omitting the second and third dimensions listed under areas 3, 4 and 5. Each dimension was evaluated using a pull-down rubric which contained four levels from poor to excellent. The levels were assigned scores of 0% (fails), 50% (marginal), 75% (meets expectations), and 100% (exceeds expectations) for averaging purposes. The survey was sent to the students by email and the completed surveys were returned to the instructors by the same method. The scores provided by peers for each dimension were averaged and the overall average of all dimensions for each student was used in Eq. 3 to calculate \( CF_{\text{peer}} \).

![Normalized peer rating CF](image)

### 3. RESULTS AND DISCUSSION

The results of the peer evaluation surveys conducted during the last four academic years are analyzed in this paper. Two surveys were performed in 2014–15 (November and March), six surveys in 2015–16 and in 2016–17 (September, October, November, January, February and March) and four surveys in 2017–18 (September, October, November and January), for a total of eighteen surveys. The frequency histogram of the 949 normalized peer ratings obtained in the surveys is shown in Fig. 1. As expected, the mean of the distribution is unity since the mean of the \( CF_{\text{peer}} \) values for the members of each team is 1.0. The pooled results indicate that about 87% of the students in our course get a normalized peer rating between 0.9 and 1.1, 6% of the students get a rating between 0.7 and 0.9, and 6% of the students get a rating between 1.1 and 1.3.

![Normalized peer ratings](image)

The normalized and absolute peer ratings are plotted against one another in Fig. 2. As can be observed in this figure, most students who get a normalized rating above 0.9 also obtain an average absolute rating which exceeds 75%. In other words, their performance is judged by their peers as meeting or exceeding expectations. Conversely, students who get normalized ratings less than 0.9 are considered by their peers as individuals who are not meeting expectations (i.e. their peer rating is less than 75%). According to our experience, most students with normalized ratings between 0.7 and 0.9 are ineffective team members who negatively impact team dynamics. It is important to identify these individuals early to help them improve their team skills. Those who repeatedly get normalized ratings below 0.7 are typically highly disruptive individuals who should be removed from their teams to prevent team collapse. Although only about 7% of the students in our course are ineffective members,
they impact about 30% of the students because they are each a member of a four-member team. It is rare that a team contains more than one ineffective member.

When the students who are members of a team containing one or more ineffective members are removed from the data presented in Fig. 1, the distribution shown in Fig. 3 is obtained. As shown in Fig. 3, most of the members of effective teams (about 98.5%) get normalized ratings between 0.9 and 1.1. The ratings also follow the normal distribution

\[ f(x) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \]  

with mean \( \mu \) of unity and standard deviation \( \sigma \) of about 0.04. The observed peak frequency of the distribution (i.e. at CF\(_{\text{peer}}\) = 1.0) is greater than predicted by the normal distribution because some students, especially in March when the project is nearly over, agree to give one another identical ratings. As pointed out by Kaufman et al. [8], this is no cause for concern as it is likely an indication that the team is working well and members feel that they should all earn the team project grade.

The standard deviation of the normalized ratings of students in effective teams was determined for each of the eighteen surveys. The results are shown in Table 3. The number \( n \) of students in effective teams tend to decrease as the project progresses because the number of ineffective members tend to increase. The standard deviations presented in Table 3 do not appear to vary with time and the hypothesis that they are all estimates of the standard deviation \( \sigma_0 \) of the same population can be tested using the statistic

\[ z = \frac{s-\sigma_0}{\sigma_0/\sqrt{n}} \]  

Table 3. Standard deviation of ratings in effective teams.

<table>
<thead>
<tr>
<th>Survey</th>
<th>No. n of students in effective teams/ % of total</th>
<th>Standard deviation of normalized ratings, s</th>
<th>z</th>
</tr>
</thead>
<tbody>
<tr>
<td>2014-15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>November</td>
<td>41/80.4%</td>
<td>0.0597</td>
<td>4.79</td>
</tr>
<tr>
<td>March</td>
<td>28/54.9%</td>
<td>0.0421</td>
<td>0.588</td>
</tr>
<tr>
<td>2015-16</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September</td>
<td>49/92.5%</td>
<td>0.0391</td>
<td>0.037</td>
</tr>
<tr>
<td>October</td>
<td>41/77.4%</td>
<td>0.0328</td>
<td>-1.44</td>
</tr>
<tr>
<td>November</td>
<td>45/84.9%</td>
<td>0.0489</td>
<td>2.41</td>
</tr>
<tr>
<td>January</td>
<td>40/75.5%</td>
<td>0.0425</td>
<td>0.803</td>
</tr>
<tr>
<td>February</td>
<td>32/60.4%</td>
<td>0.0406</td>
<td>0.328</td>
</tr>
<tr>
<td>March</td>
<td>32/60.4%</td>
<td>0.0365</td>
<td>-0.513</td>
</tr>
<tr>
<td>2016-17</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September</td>
<td>38/68.6%</td>
<td>0.0398</td>
<td>0.179</td>
</tr>
<tr>
<td>October</td>
<td>47/92.2%</td>
<td>0.0387</td>
<td>-0.076</td>
</tr>
<tr>
<td>November</td>
<td>37/72.5%</td>
<td>0.0369</td>
<td>-0.463</td>
</tr>
<tr>
<td>January</td>
<td>37/72.5%</td>
<td>0.0273</td>
<td>-2.58</td>
</tr>
<tr>
<td>February</td>
<td>33/64.7%</td>
<td>0.0311</td>
<td>-1.65</td>
</tr>
<tr>
<td>March</td>
<td>33/64.7%</td>
<td>0.0392</td>
<td>0.042</td>
</tr>
<tr>
<td>2017-18</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>September</td>
<td>52/92.9%</td>
<td>0.0406</td>
<td>-0.601</td>
</tr>
<tr>
<td>October</td>
<td>44/78.6%</td>
<td>0.0365</td>
<td>0.418</td>
</tr>
<tr>
<td>November</td>
<td>32/57.1%</td>
<td>0.0413</td>
<td>0.472</td>
</tr>
<tr>
<td>January</td>
<td>40/71.4%</td>
<td>0.0420</td>
<td>0.688</td>
</tr>
</tbody>
</table>

whose sampling distribution is approximately the standard normal distribution [14]. The hypothesis can be accepted with a level of confidence of 95% if \(|z|\) is less than 1.96. It can be seen in Table 3 that the \( z \) value associated with each standard deviation meets this criterion in all cases except for three surveys (Nov. 2014, Nov. 2015 and Jan. 2017). The value of \( \sigma_0 \) used in these calculations was obtained by pooling the CF values used to produce Fig. 3. The small monthly variations in the standard deviation are thus not statistically significant indicating that the distribution of normalized ratings for students in performing teams is time invariant in our course. This is
an important finding which indicates that the probability that an effective student gets a normalized rating less than 0.9 (i.e. $\mu - 2.5\sigma_o$) is less than 0.6% in our course. This confirms that students who get normalized ratings less than 0.9 are outliers and likely not effective teammates.

The rating distribution for the students who are members of a team containing one or more ineffective students (i.e. students with a normalized rating less than 0.9) is shown in Fig. 4. This distribution also has a mean of unity but is much broader than the distribution of Fig. 3. This is because ineffective students with low ratings decrease the team average and thereby inflate the normalized ratings of their teammates. The ratings of Fig. 4 were therefore re-calculated after excluding the ratings of the ineffective students from the calculation of the team average (i.e. denominator of Eq. 3). The resulting distribution is shown in Fig. 5. Interestingly, the normalized ratings of the effective members are now between 0.9 and 1.1 and have a distribution very similar to that of Fig. 3. This demonstrates that all effective members, irrespective of the quality of their team, are part of the same distribution if the ineffective members are excluded from the calculation of the team average rating. It is statistically appropriate to exclude ineffective members from the team average calculation since they are outliers as demonstrated earlier. The exclusion however increases the team average and thereby decreases not only the normalized peer ratings of the effective members of the team but also those of the ineffective members.

When the corrected peer ratings of students in teams with ineffective members are pooled with the ratings of the remaining students, the distribution shown in Fig. 6 is obtained. By comparing Fig. 6 with the original global distribution of Fig. 1, we see that the exclusion of the ineffective members from the team average calculation has essentially eliminated the presence of normalized ratings with a value above 1.1. Except for the ratings of ineffective students, the remaining normalized peer ratings are all between 0.9 and 1.1 with a mean of 1.0.

Many authors who use the multiplier method to adjust individual grades have argued for maximum deviations within ±10% of the group mark. To achieve this goal, some authors have proposed to impose an arbitrary maximum multiplier value of 1.1 [8] and others have developed complex normalization techniques [17]. The approach proposed in this paper is statistically sound, less subjective and simpler.

Because the frequency distribution of normalized peer ratings is invariant in our course, we can reliably identify problematic students after each peer assessment survey from the normalized peer rating values. Any student with a normalized peer rating less than 0.9 is deemed an ineffective team member and must meet with the instructors of the course to discuss remedial measures. No similar attempt to use the distribution of normalized peer ratings to identify ineffective team members has been found in the literature. It is therefore unknown whether the invariant nature of the distribution is specific to our course.

In addition to peer evaluations, we have implemented in our course a variety of strategies suggested in the literature to minimize free-riding such as allowing
students to select their own teammates to maximize group cohesion [7,15,18], keeping group size between 3 and 5 individuals to facilitate team coordination [15,17,18], and instilling accountability for individual contributions by asking students to use time sheets to record time spent on various project tasks [1]. As a result, most individuals identified as ineffective members in our course are not free riders trying to take an unfair advantage of the work of others. They are instead individuals who are either rejected by their teammates because they cannot complete allocated tasks for which they feel ill-equipped, or are not willing to let the project consume all their time to the detriment of their other courses, or work with academically stronger students who feel that they are carrying the group. Such issues can be resolved if identified early.

During the 2015-16 and 2016-17 academic year, about 50% of the students who got normalized peer evaluations less than 0.9 were able to resolve the issues troubling their teammates within a month and did not get further poor ratings as can be seen in Fig. 7. Students who got ratings less than 0.9 three or more times during the year were a source of severe frustration within their teams. We met with these students and their teammates to seek solutions because it is desirable that all students have a positive experience in the course. In the few cases where the underlying issues could not be resolved, we decided to remove the ineffective team members from their team to alleviate frustration and anger. The ineffective members were informed to drop the course or undertake a sub-project on their own. Work performed before the split became common property, but subsequent work had to be done independently.

![Data from 2015-16 and 2016-17 academic years](image)

**Fig. 7.** Percentage of students who got one or more poor monthly ratings during the year versus no. of such ratings.

### 4. CONCLUSIONS

The original hypothesis has been verified. Ineffective team members obtain normalized peer ratings which fall outside the distribution of ratings for effective team members. Most effective team members get normalized peer ratings between 0.9 and 1.1 whereas ineffective members obtain normalized ratings below 0.9. The ratings of ineffective members should not enter in the calculation of the team average to avoid inflating the normalized ratings of their teammates.

**Acknowledgements**

Funding received from the Natural Sciences and Engineering Council of Canada for the NSERC-UNB Chair for Collaborative Engineering Design Education is gratefully acknowledged. The authors are also grateful to the faculty and industrial co-mentors who have contributed to ChE 4225 between 2010 and 2018.

**References**


