A THREE-TIER EVALUATION RUBRIC FOR THE ASSESSMENT OF GROUP PROJECTS IN CHEMICAL ENGINEERING DESIGN COURSES

Andrew Sowinski1 and David Taylor2
1Faculty of Engineering, University of Ottawa, 2Chemical and Biological Engineering Department, University of Ottawa
sowinski@uottawa.ca, david.taylor@uottawa.ca

Abstract – Group projects are often key to engineering design courses since they simultaneously develop teamwork and communication skills in the context of solving difficult engineering problems. However, fair, consistent, and transparent grading of these projects are difficult to achieve, and the individual contribution of students can likewise be difficult to evaluate.

Standardized marking rubrics are often used to increase the consistency and fairness of project evaluations; however, these frequently lack a systematic means for evaluating individual effort within group work. Rubrics also are difficult to employ when there are numerous possible solutions and where some solutions are more elegant or challenging when compared to others.

To provide a consistent accounting of individual effort and the difficulty of a submitted group design solution, a three-tier marking rubric was developed. Comparing the project grades between two cohorts in the same course showed that there was a broader distribution of grades when using the three-tier marking scheme.

Keywords: Standardized rubric, Group project, Design course, individual grading

1. INTRODUCTION

Group projects are a valuable component to an engineering design courses since they can be used to develop teamwork and communication skills in the context of solving difficult engineering problems. Evaluating group projects, however, can pose several significant challenges:

a) project marking can appear subjective to students, and may differ between evaluators;

b) different groups that assume more difficult design projects or that provide a more sophisticated solution may not be compensated for the extra effort; and

c) an individual student’s contribution within a group report is difficult to identify and hence to evaluate.

Rubrics are often used to address item a) above and may take on many different forms, from general qualitative assessments to more discrete quantitative descriptions [1]. Rubrics can also be kept confidential from the students until after an assessment or can be provided up front with the assignment in question.

In addition, well-designed rubrics that have discrete, clearly defined criteria can help students understand the importance of different components of a project [2-4]. It has been demonstrated, for example, that students who were given a marking rubric in advance and subsequently used it as a guide showed improved scientific writing skills, compared to those who did not use the rubric [5]. A substantial improvement has also been shown in the quality of work when rubrics were used throughout a design project to provide useful feedback from the initial draft to the final submission [4].

Typical approaches to addressing individual contributions to group work rely on one or more of the following three assessment tools:

1. An evaluation of student performance through observation by the instructor or the teaching assistant.

2. An evaluation of the student work ethic, teamwork, and contribution by other team members.

3. A reflective or self-evaluation by the peer on their perceived work ethic, teamwork, and contribution to the group project.

From an assessment of one or more of these, instructors can then determine the individual contribution factor (ICF) of any student. Nepal [6] defines the ICF as:

\[ ICF = \frac{\text{Individual contribution} \%}{\text{Average contribution} \%} \]  

The combination of any number of evaluation methods produces an Individual Weight Factor (IWF) [7, 8] that can
be used to differentiate between the overall group project and individuals.

To address the assessment challenges posed by varying degrees of sophistication in student designs and the relative contributions of various team members to this solution, a three-tier evaluation rubric was developed and applied to project work for a chemical engineering undergraduate course in numerical methods. Following an initial project proposal by student groups, the corresponding overall marking scheme was provided to the students so that they were aware of how the project would be assessed, including compensations for using more complicated and sophisticated approaches.

### 2. DESIGN OF THE MARKING RUBRIC

The rubric was designed with three components: a clear and explicit marking rubric, a level of difficulty component, and a component to evaluate individual contributions to the group submission.

#### 2.1. The Marking Rubric.

The first tier was an overall rubric-based evaluation ($E_g$) that sought to minimize potential grading discrepancies between evaluators. The rubric was broken into five components:

1. Written Report: Content
2. Written Report: Style
3. Software Code
4. Group Meeting Assessment
5. Self & Peer Evaluation

Each section offered subcategories or bins for evaluation; within each bin was a series of questions for the evaluator to answer based on predetermined performance levels (e.g., substantially incomplete, poor, marginal, good, very good, excellent). Each bin was subsequently scaled (non-linearly) and multiplied by the weight of each question to determine the mark of that component. The bin names and their weighted scale are shown in Table 1, and an example of questions is shown in Figure 1.

#### 2.2. The Level of Difficulty

The second tier employed a level of difficulty assessment analogous to evaluation procedures applied in certain competitive sports such as gymnastics. In this case different levels of difficulty were assigned to the students’ proposed solution approach based on the identified level of challenge, with a maximum score of 100% for the most challenging of these. For example, if a group opted to carry out the minimum number and/or the least sophisticated methodologies to complete the assigned task, the work was placed in the bottom “difficulty bin” corresponding to a maximum achievable grade of 60%. As the sophistication of their proposed solution procedure increased, groups were assigned to difficulty bins with higher maximum achievable grades, with the top bin receiving a maximum possible score of 100%. The difficulty bins were predetermined before releasing the project and the students were made aware of these in advance. The difficulty bins are shown in Table 2.

As a minimum, a group was expected to submit functioning code to qualify for a baseline value of 60%. Additional design complexities would merit additional percentage points to a maximum of 100%. This percentage then defined the degree of difficulty factor for the group, $D_g$.

---

### Table 1: Performance Levels

<table>
<thead>
<tr>
<th>Performance</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Substantially Incomplete</td>
<td>0%</td>
</tr>
<tr>
<td>Poor</td>
<td>30%</td>
</tr>
<tr>
<td>Marginal</td>
<td>50%</td>
</tr>
<tr>
<td>Good</td>
<td>75%</td>
</tr>
<tr>
<td>Very Good</td>
<td>85%</td>
</tr>
<tr>
<td>Outstanding</td>
<td>100%</td>
</tr>
</tbody>
</table>

When the evaluator completed the assessment for a group, the total score was calculated and scaled to 100% to determine $E_g$. 

### Table 2: Difficulty Levels

<table>
<thead>
<tr>
<th>Difficulty Bin</th>
<th>Maximum Achievable Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bottom</td>
<td>60%</td>
</tr>
<tr>
<td>Middle</td>
<td>70%</td>
</tr>
<tr>
<td>Top</td>
<td>100%</td>
</tr>
</tbody>
</table>

---

### Fig. 1: An example of the part of the marking rubric for the Software Code section.
Table 2: Level of difficulty and additional marks based on the tasks completed.

<table>
<thead>
<tr>
<th>Level of Difficulty</th>
<th>Task</th>
<th>Additional Marks Obtained</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>The code compiles and executes.</td>
<td>60%</td>
</tr>
<tr>
<td>2</td>
<td>A solution is found using the optimal number of λ values.</td>
<td>+10%</td>
</tr>
<tr>
<td></td>
<td><strong>Only two of the following tasks</strong></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>A solution is found using bisection method.</td>
<td>+5%</td>
</tr>
<tr>
<td>1</td>
<td>A solution is found using linear interpolation.</td>
<td>+5%</td>
</tr>
<tr>
<td>3</td>
<td>A solution is found using modified linear interpolation.</td>
<td>+15%</td>
</tr>
<tr>
<td>3</td>
<td>A solution is found using Ridder’s method.</td>
<td>+15%</td>
</tr>
<tr>
<td>3</td>
<td>A solution is found using Newton-Raphson method.</td>
<td>+15%</td>
</tr>
</tbody>
</table>

2.3. The Individual Component

The third tier differentiated individual contributions to the group work, using a weighted average of the individual’s evaluation by the teaching assistants (graded through group meetings), a peer evaluation, and a self-evaluation. Each weighted individual evaluation was then divided by the group average to provide a normalized weighting factor for that member ($C_i$),

$$C_i = \frac{0.6E_{TA,i} + 0.3E_{peer,i} + 0.1E_{self,i}}{\sum (0.6E_{TA,j} + 0.3E_{peer,j} + 0.1E_{self,j})}$$

where $E_{TA,i}$ is the individual TA evaluation, $E_{peer,i}$ is the peer evaluation, $E_{self,i}$ is the self-evaluation, and $n$ is the number of group members. The denominator is the average of the TA, peer, and self-evaluations of the group. With Eq. 2, it is possible that a student would receive a value of $C_i$ that was greater than 1 while another student a $C_i$ value that was less than 1. This allowed the assessment to account for the fact that comparatively strong students who were part of a weaker group would likely have done better had they been a member in an academically stronger team. To ensure that the $C_i$ did not drastically affect the student mark, its range was capped to between 0.9 and 1.1.

The TA evaluation was calculated using a marking rubric to assess a mandatory group meeting midway through the semester. The TA evaluated each individual based on the following questions:

1. Did the student contribute reasonably to the group work?
2. Was the student actively participating in the meeting discussion?
3. How well did the student convey her/his understanding of the sections of the design work to which she/he was assigned?

The peer and self-assessment were completed anonymously and were mandatory. Failure to submit a self and peer assessment resulted in a grade of 0 for the project. Each student was required to assess themselves and their peers based on the following questions:

1. Participation: Is your teammate self-directed and does she/he consistently focus on the task and what needs to be done?
2. Dependability: Does your teammate turn all work in on time and not depend on others to do or complete work?
3. Listening and Discussing: Does your teammate respectfully listen, interact, cooperate, and help the group come up with positive ideas?
4. Research and Information Sharing: Does your teammate routinely complete research, gather information, and share with the group?

For each question, each student was required to distribute a limited number of points to each member (e.g., 400 points for 4 students in a group). If, for example, all of the work were equally shared then each student would receive 100 points.

All of the points for each student in a group were then combined and scaled to determine $E_{peer,i}$ and $E_{self,i}$ and used to determine $C_i$.

2.4. The Final Grade

The individual mark for each student ($I_i$) was then calculated by multiplying the three tiers together as follows:

$$I_i = E_{g}D_{g}C_i$$

3. RESULTS AND DISCUSSION

To determine the effectiveness of the three-tier marking scheme, the 2016 cohort project marks based on the three tiered marking rubric were compared to those of the 2015 cohort that had a project of similar difficulty and used the same marking criteria but was not tiered. The TA, self and peer evaluations were part of the marking rubric, therefore the $C_i$ was never larger than 1. Figure 2 shows a distribution of the marks for each of the cohorts. The individual marks
of each student were rounded to the nearest 5% and binned. The bins were then normalized to the number of students in each cohort. Although the overall average was similar (82% for the 2015 cohort, 84% for the 2016 cohort), there was a larger range of values when using the three-tier marking scheme. This was attributed to both the $D_g$ and $C_i$; however, the latter had a larger effect as most groups attempted the more difficult tasks, likely due to the fact that a thorough marking scheme was provided with the project description that showed the consequences of attempting differing degrees of complexity in the code design.

Scaling of the individual marks with $C_i$, which varied between 0.9 and 1.1, had an impact on the range of final marks where the mark distribution for the 2016 Cohort was broader than that for the 2015 Cohort as shown in Figure 2.

4. CONCLUSIONS

The three-tier marking scheme offered a transparent and fair evaluation tool that ultimately promoted maximum student achievement and that effectively accounted for individual performance within a group project setting. Using the three-tier marking scheme also resulted in a broader distribution of overall grades than that obtained using a single tier marking scheme.

5. ACKNOWLEDGEMENTS

The authors would like to acknowledge the support of the Faculty of Engineering at the University of Ottawa for the support on this project.

6. REFERENCES


