Peer Assessment: Preparing for Professional Practice

Denard Lynch
University of Saskatchewan
denard.lynch@usask.ca

Abstract As members of a learned profession, engineers are often required to assess and critique the work of others. Preparation for this professional responsibility should be developed during their academic training, alongside other required skills. This author proposes that there are generic skills and training methodologies that can be applied to both technical and “soft skill” situations to prepare students for this task. This paper discusses results of a peer assessment exercise applied to a “soft-skills” situation.

The main objectives of this experiment were to i) develop peer assessment skills in students, ii) maintain or improve the accuracy of assessments of subjective material, iii) improve students’ skills in the subject area, and iv) potentially reduce marking effort for instructors.

The experiment described in this paper involved peer assessment of a short report (3 – 5 pages) required as a term assignment in a senior course on ethics and professionalism. The reports were prepared and submitted by groups of two students. Each student was then randomly assigned two other reports to assess in a double-blind fashion, except that no student reviewer received their own report. For reference and analysis, each report was also assessed by both the instructor and a Teaching Assistant resulting in approximately six separate assessments per report. The results were used to determine a grade for the assignment. The original assignment rubric was used for all assessments. In addition, formative feedback was provided by the reviewers and returned to the authors.

The quality of the numerical results was analyzed by comparing the marks determined by the student assessors to the reference (instructor, TA) assessments. An average difference of 8.5% was observed, and was considered generally acceptable given the subjective nature of the material. Student “generosity bias” was also considered, but found to be virtually non-existent with a difference in student versus reference averages of less than 0.2%. “Outliers” were anticipated, and student assessment showed approximately twice the standard deviation of the reference marks. A weighted average was used to determine the assignment mark, and any marks outside a 20.0% band were de-weighted. Approximately 25% of cases were weight-adjusted, resulting in a maximum mark adjustment of 4.1% and an average adjustment of only 1.6%.

Feedback was solicited from students prior to the peer review period and at the end of term. Informal feedback was solicited prior to the review period regarding instructions and logistics, and was used to refine the setup for the peer review phase. Questions on the value of both the exercise and the feedback provided were included in an end-of-term survey of students about the course, with 83% finding the exercise “a bit” or “quite” educational and 74% finding the peer feedback “a bit” or “quite” helpful.

Involving students in this peer evaluation exercise had generally positive outcomes and provided experience from which to improve future implementation of peer assessments to achieve the objectives of this experiment. Recommendations regarding future application include: importance of instructions and setup, student training and rehearsal, and mark determination considerations.

Keywords: Student assessment, Assessment tools, summative, formative, grading systems

1. INTRODUCTION

Preparing students for the engineering profession includes developing technical skills and soft skills required to communicate, understand and analyzes their role in, and effect on, society as well as fulfill their other professional obligations. Most professional licensing bodies in Canada expect their members to follow a code of ethics similar to that included in the Engineering and Geoscience Professions Act of Saskatchewan which expects members to: “… give credit where credit is due and accept, as well as give, honest and fair professional criticism” [1]. This means that in professional practice, engineers will be expected to both review the work of colleagues and to have their own work peer reviewed. It would seem appropriate to develop skills for doing so by providing instruction and experience in this subject during an engineer’s education.
Most instructors have undoubtedly observed that there is a learning opportunity in reviewing even students’ work, and that experience at doing so improves their ability to provide “honest and professional criticism”. These same benefits should also be made available to students through their courses. [2] Further, students may benefit more from peer feedback than instructor feedback, being openly honest and using terms with which students are familiar. [3] These benefits are consistent with those cited by Orjuela-Laverde and Chen. [4] Providing guidance and opportunity for students to develop skills in this area should, logically, be part of an engineer’s education and training.

Finally, the environment and instruction must be setup appropriately to facilitate these benefits.

2. THE RESEARCH QUESTION

The two main questions addressed through this experiment were: 1) “Did students learn from doing a peer assessment or from receiving peer feedback?” and 2) “Was the peer assigned grade accurate and acceptable to students?”

A voluntary survey of participants was used to estimate educational value, and a comparison between instructor-determined grades and peer determined grades was used to assess accuracy.

3. EXPERIMENTAL DESIGN

The course in which this experiment was conducted was a senior course in ethics and professionalism which has been instructed by the author extensively over the past eighteen years. A report assignment was chosen as the subject of the peer assessment because of its suitable characteristics:

- Reasonable length (nominally four pages) to minimize the extra course load on students;
- Involved only general expertise that all students should equally possess (communication skills);
- Would be a beneficial skill to improve for all disciplines;
- Had common objectives familiar to all students
- Instructor and teaching assistant have expert knowledge in order to reliably provide a “control” mark for comparison.

Each submitted report received three to four student reviews and two expert reviews nominally resulting in five or six sets of feedback and marks to be combined and compared. Also utilizing the expert reviews helped lower the risk of error and gain student acceptance. The course enrollment was 44, which also lowered the risk of unrecoverable, unforeseen complications.

3.1 Peer Assessment Instructions

An excerpt from the assignment instructions relevant to the peer assessment is shown in Appendix A. Note the inclusion of the peer assessment as a required part of the assignment, the details on mark determination, a brief list of benefits to the student, and the opportunity to appeal. The same rubric given for the preparation of the report was to be used to do the peer assessment.

Appendix B shows additional benefits, instructions and guidelines for providing an assessment and feedback in a professional manner. This review was assigned in a “double-blind” fashion; neither the submitter or reviewer had access to the other’s identity (although the instructor could, necessarily, decode the assignment).

4. RESULTS & ANALYSIS

This section will discuss the results of the student survey to determine the learning benefits to students as well as statistical analysis and comparison of grade results to determine the acceptability of numerical results.

4.1 Student Benefits of Peer Assessment

As previously stated, enrollment in this course was 44 students. A brief survey was administered at the end of term after completion of the peer assessment exercise and following all related mark submissions. Top Hat, a classroom response system, was the vehicle employed to gather student impressions of usefulness of the peer review exercise and the resulting report feedback.

Thirty-two (73%) of students provided input to the survey although not all students answered all questions. Figure 1 shows the responses to the first part of the question: “How would you rate the Peer review exercise and feedback on your Interview reports? (select one each for the review, and the feedback)”. 

CEEA18; Paper 116
University of British Columbia; June 3 – 6, 2018  – 2 of 8 –
Twenty out of twenty-four (83%) of respondents indicated they found the exercise “at least a bit helpful” or “quite helpful”, while 17% indicated they found no value in the exercise. (Note: in retrospect, there is a possible source of error in interpreting this data as respondents were free to respond for the review exercise and the feedback in either order and to select any of the six possible responses. While the term “feedback” was used specifically in two responses and “reviewing someone else’s…” was explicit in another, “quite helpful” and “quite educational” could have been applied to either the review or the feedback. On closer examination, this did not appear to affect general interpretation of the overall response.)

In order to minimize the extra effort required for this exercise and focus the feedback, students were asked to provide only one suggested improvement and one compliment on the reports they reviewed. This was gathered and returned to the authors for their consideration. Figure 2 summarizes the 17 responses attributed to feedback. 76% responded that they found the feedback “a bit helpful” or “quite helpful”, while 24% found it of no value.

Students also have the opportunity to offer free-form comments on the peer exercise. Relevant suggestions included: more instruction on reviewing, a more detailed rubric, and more detailed information on the grade composition. As a criticism, a few respondents indicated a lack of confidence in their peers ability to perform a credible review or determine an accurate grade (also consistent with criticisms noted by Orjuela-Laverde and Chen. [4]), although the statistical results in the next section don’t generally support this view.

Part of the rationale for formally developing peer review skills and that as practicing professionals, these students will be expected to do so in a responsible and professional manner. Although not specifically targeted or associated with peer reviews, one question related to professionalism is included here for its potential relevance to this topic. As summarized in Figure 3 in response to the question: “Do you feel this course increased your appreciation of the profession’s impact, and helped prepare you for life as a professional?”. 94% responded that it had at least some positive effect. Admittedly, peer evaluation was only one part, and probably a small part, of the material and concepts covered in this course.

4.2 Comparison of Grade Results

130 assessments were completed on 23 assignment reports. Two reviews of each report were performed by the instructor and an experienced teaching assistant (TA). These were intended to provide a “control” against which student assessments would be considered.

Table 1 summarizes basic statistics for all reviews and compares those for student reviewers as a group to the “control” reviewers.
Previous work by the author has suggested that the average of a number of peer reviews tends to be very close to the assumed “accurate” reference. This can be seen again in this case where although the standard deviation is approximately twice that of the control, the averages are extremely close. This may also be evidence of Dylan Wiliam’s thesis that peers can be more honest than instructors. [3]

In an effort to validate the control’s accuracy, the two independent reviews by the instructor and experienced TA were examined. Table 2 clearly shows a close agreement between the experienced assessments. Calculating Pearson’s Correlation Coefficient corroborates this observation with an r of 0.66 and a p-value of 0.006

| Table 1: Peer Mark Statistics (Student vs Instructor) |
|-----------------|-------|-----------------|-------|
|                 | Student | Control |         |
| Average         | 11.9   | 79.5%       | 11.9  | 79.6% |
| Min             | 7.0    | 46.7%       | 10.5  | 70.0% |
| Max             | 15.0   | 100.0%      | 14.0  | 93.3% |
| StDev           | 1.6    | 10.5%       | 0.9   | 5.8%  |

Assuming the control average as a reference, Figure 4 shows that while there is a close agreement between the student and control overall averages, variation between the individual student marks and the control of up to almost 15%.

This variation is illustrated in a different way in Figure 5 by plotting cumulative number of report marks against an increasing variance with the control marks. While 83% of the marks are within 10% of the control, there are still 17% that vary by up to 15%.

The provision in the mark specification to accommodate such cases was to lower the weighting of any marks that were outside a band of ±10% of the overall average. Of the six cases (out of 23 reports) that triggered this adjustment, the change in mark ranged from zero to 0.62 of a mark (0 to 4%). Individually, the adjustments were: 0.00, 0.00, 0.03, 0.32, 0.48, 0.62 out of a maximum of 15 marks for this assignment. Another potential strategy to minimize the potential inaccuracies of a broad variation in marks would be to eliminate outliers as used previously by Schmid. [2] In this case, the integration of the control marks help minimize the error, resulting in an acceptable result.

| Table 2: Instructor vs TA Statistics |
|-----------------|-------|-----------------|-------|
|                 | Instructor | TA |         |
| Average         | 11.96 | 79.7%       | 11.93 | 79.6% |
| Min             | 10.50 | 70.0%       | 10.50 | 70.0% |
| Max             | 14.00 | 93.3%       | 13.30 | 88.7% |
| StDev           | 0.98  | 6.5%        | 0.76  | 5.0%  |

CONCLUSIONS AND OBSERVATIONS

The first objective of this work was to assess the benefit to students of formally introducing peer evaluation in an undergraduate course. The subjective results of a survey suggest that the concept was reasonably well received by students with the majority (>75%) indicating they found such an exercise at least somewhat beneficial. While this is encouraging, it does not provide any objective assurance that the exercise actually improved their skill in this area.

The analysis of mark results indicates that broad averages are in close agreement with a control, but individual results may have more than an acceptable error and could not be used reliably by themselves for grading without further refinement or development of student skill in this area. The close agreement on a broad basis may suggest that an increased number of peer reviews,
especially with less experienced practitioners, could result in an acceptably accurate assessment.

Future Work

The feedback and experience suggest that increased preparation, via instruction and practice, could both improve the result and the student experience. Repeated trials while monitoring the quality of feedback may verify the objective of developing skill to a level appropriate for a pre-professional-practice environment.

An investigation into an acceptable error rate for mark determination would be a helpful target not only to determine whether students are learning to recognize an acceptable level of performance in their peers, but also to help determine an accurate grade for student work and potentially reduce instructor marking effort. The subject course on ethics and professionalism is arguably an ideal opportunity to exploit external instructors “real world” experience and wisdom, but the typically heavy marking burden is a barrier to recruitment.

This peer assessment exercise was to be implemented again, with modifications, in the same course in the following term involving approximately 250 students. Planned changes include additional detail in instructions and rubrics, and including a short “practice” peer exercise prior to the main peer assessment to familiarize students with the procedures. There was to be no “control” marking, but the number of peer reviews was to be increased. The methodology for dealing with “outliers” will also be reviewed and modified.

Acknowledgements

The author wishes to acknowledge the support of the University of Saskatchewan for providing the means and opportunity for its instructors to explore and enhance their pedagogical skill and knowledge and contribute to the body of knowledge in this area. The authors would also like to thank Dr. Sean Maw and the other members of the Innovative Teaching and Research in Engineering Education Group at the University of Saskatchewan for their support and suggestions. Finally, the author thanks the students in the subject courses for their cooperation, understanding, and suggestions for the setup related to this research.

Works Cited


APPENDIX A: Review Instructions and Rubric

Grading

The report will contribute 15% towards your final grade. Grading will be based on the rubric below, and will be a combination of peer and instructor evaluations.

As part of this assignment, in addition to submitting a partner-group report, each student will be individually assigned, via BBLearn, two other reports to evaluate using the same rubric given below (with the omission of the criteria for partner group formation in the first category). Assignment of reports will be done within approximately one week following the original submission due date. *The peer evaluation will be due two weeks following its assignment.* Author groups may also evaluate the value of the feedback provided by their peers to provide reverse feedback, although this is not a strict requirement of this assignment. The reverse feedback, consisting of similarly brief comments, should be emailed to your instructor.

Deliverables: An evaluation for each report consisting of:

1. A grade out of 15 marks. There are 5 rubric categories. It would be helpful to break the mark down for each category. E.g.: “2/1/2/3/4 = 12/15” would mean 2 for “Selection of Subject”; 1 for “Questions” etc.
2. Also, make the following two comments on each of the reports:
   i. One thing I liked about this reports was…..
   ii. One thing I felt would help improve this report was….

Please also consult the guidelines and suggestions on peer evaluation given in Appendix B at the end of this document.

Show the result (grade) and your comments right in the assignment submission area on BBLearn by typing it in as a text submission in the appropriate area. *Reference source not found. for detailed peer submission instructions.* You are expected to do a serious, professional evaluation and make meaningful, constructive comments. Your formative feedback will be shared, anonymously, with the authors for their benefit.

Mark Determination: The grade for this assignment will be determined by combining the peer evaluations and the instructor evaluation as follows:

i. if all grades are within 3 marks, the average shall be used.

ii. if the difference between the highest and lowest mark is > 3 marks, any outlying mark’s weighting will be discounted by the amount it deviates from the average of the remaining marks (e.g. 14.4, 11, 13.8, 12, 14; two groups of 4 that are within 3 marks, so the 11 will be weighted 1.615, where 1.615 is the deviation from the average of the remaining 4. Mark = (9.9*12+12.7+14.4+14.9)/3.98 = 13.2)

Note1: the peer evaluation must be completed to receive credit for this assignment.
Note2: students may appeal to the instructor to review the final mark if they feel it does not accurately reflect their report’s merits.

Benefits to the student:

1. Peer-assessment provides the student with the opportunity to critically examine other’s work, reflect on their own performance and use the opportunity to improve. When students enter the work force they will be required to have their work reviewed by peers and review the work of others.
2. Providing formative (helpful) feedback helps the author improve the quality of their reports in the future.
3. Providing ‘reverse’ peer assessment provides feedback to the reviewer about the quality of the review provided.

Report Rubric (total 15 marks):

- Selection of subject
  - 2 Clear justification of relevant subject selection related to assignment objectives and formation of a partner group in a timely fashion
  - 1 Some suggestion that relevance was considered in subject selection or late formation of team/group
  - 0 No indication that suitability of subject was considered or no effective group formation in a reasonable timeframe (e.g. instructor intervention required to form group)

- Questions
  - 2 Interview questions were designed to stimulate conversation (i.e. open-ended) and focus on assignment
objectives; adaptable to flow of conversation
  o 1  Some questions were open, evidence of some adaptation (if required)
  o 0  Questioning did not draw out relevant conversation or adapt to responses.
• Analysis and interpretation
  o 3  Clear evidence of analysis or interpretation beyond direct response of subject
  o 2  Some instances of evidence of analysis or interpretation beyond direct response of subject
  o 0  Very limited or no evidence of analysis or interpretation beyond direct response of subject
• Lessons learned
  o 3  3 or more interpretations of results that could be applied to improve engineering practice, or corroboration of principles considered in class
  o 1.5  1 or 2 interpretation of results applied to engineering
  o 0  No clear interpretation of results applied to engineering
• Formatting, grammar and spelling
  o 5  Well formatted/organized to accommodate reader; only very minor or no grammatical or spelling errors; very readable.
  o 3  Some grammatical or spelling errors, but did not substantially detract from readability
  o 0  Noticeable formatting, grammatical or spelling errors that significantly affected readability

APPENDIX B: Self-Peer Submission and Assessment Guidelines

General:
These guidelines pertain to documents submitted through Blackboard Learn which are intended for “blind” (either the author or the assessor are unknown to each other) or “double-blind” assessment (both the author and the assessor are unknown to each other).
The Engineering and Geoscience Professions Act (EGP Act) in Saskatchewan states in part that “…members and licensees shall:…(e) conduct themselves with fairness, courtesy and good faith towards clients, colleagues, employers and others; give credit where credit is due and accept, as well as give, honest and fair professional criticism;” (Regulatory Bylaws; 20 (2) (e) ). In some circumstances, anonymity helps provide the objectivity that can be the basis of an honest, unbiased review. That anonymity should not be misconstrued as an excuse to do any less that one’s best to provide an honest and fair assessment of other’s work, or as an excuse to ignore criticism given in this spirit.
There is clear evidence that peer review of work provides benefits for both the giver and the receiver. As future professionals, you are expected to recognize and respect your obligation, and to benefit from exercising the associated skills.

Submissions:
In order to ensure anonymity, when this is desired, any submitted documents (e.g. uploaded to Blackboard Learn in response to an assignment) should not contain any information that could identify the author(s). This could include:
• Text that contains names, NSIDS, Student numbers, Group names or numbers (especially those assigned in Blackboard Learn). Be sure to check cover pages, appendices, headers, footers etc. for any identifying information;
• File names that contain any of the above (e.g. do not submit a file named “Denards_assignment.pdf”, “Group73.docx”, “dj210.docx”, “1010011.pdf” etc.);
• Personal information included from your word processing (or other) software. For example, Microsoft Word typically asks for and stores personal information about the author during setup (name, initials, email address…), and embeds some of that information in the “Properties” of each document produced. You can edit the “properties” of an individual document to remove this embedded information, or remove it from Word altogether.
The level of “sanitation” required will depend on the setup and process for your assignment or exercise. Consult your assignment description for details, or consult with your instructor.

Teamwork and Constructive Feedback
There are a variety of guidelines or “rules” for developing an effective team environment. An example of a simple set of such rules (Neil Mercer, Cambridge) is:
1. all students must contribute – no one says too much or too little
2. every contributor is treated with respect – must listen thoughtfully to each other
3. groups must reach consensus – work at resolving differences
4. every suggestion or assertion has to be justified – arguments must include reasons
While these guidelines target group interaction, the underlying principles can also be applied to feedback situations.

When giving feedback or a “peer review”, some other points to keep in mind include:

- The ultimate objective of peer reviews is to provide helpful feedback so the recipient can improve their skills.
- Reviewing and critiquing other’s work can improve your own skills by identifying both effective and ineffective practices.
- Provide constructive feedback by considering both positive items and suggestions for improvement. For example:
  - One thing I learned from your work is…, and
  - One way I think it might be improved is….
- Certainly most of the value in peer assessment is derived from the formative feedback, but a mark or grade (based on a rubric or exemplar) can also help the recipient track their improvement and progress.
- Use “I” statements to frame suggestions or observations in terms that illustrate how something affects the reviewer, e.g.:
  - “I found it difficult to understand the first concept you discussed because you only explained an underlying concept later on.”
  - “I found it easy to remember your key findings because of the way you summarized them in point form at the end.”
  - “I really like the way you used a graphic to describe the big band theory; I’ll try to use more of that technique in my future writing.”
  - “It was difficult for me to tell when you were moving to your next point as there was always the same space between all your paragraphs.”
- When reviewing work, focus on the work, not the person.