

Effects of Behavioral Peer Evaluations on “Free Riders” in Engineering Group Projects

Francesca Lopez, Cristian A. Dionisi, Matthew J. Traum, and Sean R. Niemi

Department of Mechanical and Aerospace Engineering, University of Florida, Gainesville, FL 32611

Abstract

In 2009, Kenneth Reid developed a multi-dimensional survey instrument for freshman to self-evaluate their own performance in team-based engineering projects. Building upon Reid’s instrument, a multiple-choice peer- and self-evaluation assessment was created for Capstone senior design; it is deployable via Learning Management Software. When using this instrument, peer grades were previously collapsed into a single score and shared with students after each major course assignment. Capstone design is fundamental both to a mechanical engineering student’s education and a department’s accreditation. Students’ ability to work effectively in teams informs ABET Criterion 3 Learning Outcome 5. Numerous studies have investigated the reliability and effectiveness of various peer evaluation techniques and instruments. Commercial engineering team peer evaluation software informed by this work has been developed. However, contrasting the easily deployed peer survey instrument developed following Reid, most commercial peer evaluation software is pay-to-play and too complex for Capstone faculty to deploy. In this study, we investigated the influence of providing detailed peer feedback on individual and Capstone team performance versus providing no peer evaluation feedback to students.

Keywords

Capstone Design, Peer Evaluation, Online Learning Environment

Introduction & Background

The mechanical engineering design course sequence is a fundamental component of both a mechanical engineering student’s education and departmental accreditation¹. A crucial part of an effective education is ensuring that students are involved with their academic progress, take charge of their own learning, and develop critical thinking skills that can be applicable to solving problems². The best practices to encourage this type of learning atmosphere are to engage the student population, while placing an emphasis on developing beneficially cooperative and reliant relationships with other peers². A key element of the mechanical engineering sequenced courses and how they are assessed by the accreditation board is the ability of students to work effectively within an engineering team³. One consistent point of research is the effective evaluation of student effort and efficacy in group environments⁴. This has been a growing topic of discussion, particularly surrounding assessment of “free riders,” i.e., students who do not contribute to the project yet receive credit nonetheless⁵.

Many studies have been conducted investigating the reliability and effectiveness of peer evaluation styles⁶; commercial peer evaluation software has been developed from some of this

research⁷. Particularly, work by Ohland, et al, found that dependability in a group member resulted in higher peer evaluation scores for that person, while obnoxious behavior and habits resulted in lower average scores⁸. Recently, Tenenberg discussed free-riding in teams as a multi-fold societal issue framed by two facets: monitoring and sanctioning³. Peer evaluations do monitor and sanction student team performance. However, giving students a single score or no score at all does not provide enough comprehensive feedback for an under-performing group member to effectively adjust behavior toward success. By contrast, providing students full (anonymized) feedback empowers them to correct weaknesses in their teamwork implementation and further develop skills and strengths over a two-semester Capstone sequence. The course sequence at the University of Florida has previously used a peer evaluation component of grading to monitor and sanction students accordingly. However, the previously established system did not include a feedback component beyond a singular numeric score provided at the semester's conclusion.

The study described herein looks at providing direct feedback to students in various behavioral categories deemed critical to collaborative environments. By providing anonymized feedback in these areas, students will be able to correct weaknesses and further develop strengths over the course of multiple semesters⁹. At the University of Florida, the mechanical engineering design sequence is split into two courses:

- 1) A “paper-design” course, henceforth known as COURSE1, wherein students are given a group design project lasting 10-11 weeks, and
- 2) A “realization” course, henceforth known as COURSE2, where students must prototype and validate the winning design from the previous semester.

Because this course sequence is a) required, and b) typically taken sequentially, peer and self-evaluations can eventually be tracked over a full academic year to study the effects evaluations and lack thereof have on the development of students' “soft skills.”

For this study, anonymized student data was tracked throughout the duration of COURSE1. Specifically, we monitored fluctuations in feedback, focusing on top performers and those that show signs of free riding to quantitatively discern the following:

- 1) What is the overall response in student behavior from peer evaluations in a course, particularly those students who receive evaluations indicative of free riding?
- 2) Do persistent “excellent” reviews of members in a group that also has substantial free-rider behavior lead to a change in student behavior?
- 3) Do peer evaluations affect long term student development or are the effects limited to a singular environment?

Methods

This study tests a new peer evaluation tool and studies its effects on student professional development and ability to mitigate free riders in group project environments throughout a two-semester course sequence within the mechanical engineering curriculum at the University of Florida. Peer evaluations are distributed three times in COURSE1, in tandem with each major report submission. The three reports in COURSE 1 are the Concept Design Report (CDR), Preliminary Design Report (PDR), and Final Design Report (FDR). For COURSE2, the peer evaluations are sent to students at intervals aligning with project waypoints and deadlines.

The peer evaluation survey was developed following the work of Kenneth J Reid who created a multi-dimensional survey instrument for freshman to self-evaluate their own performance in team-based engineering projects¹⁰. The derivative survey used in University of Florida capstone courses assesses students on 12 points related to both their academic and professional performance: the frequency of contributions, quality of contributions, perceived effort, accountability, communication frequency, communication relevance, professionalism, courtesy and respect, ability to maintain focus, technical competence, contribution to group synergy, and ability to self-start. All evaluation points are assessed on a standard 4-bin Likert scale to prevent submission of an “average” score. This scale ranges from values of one to four, with the former representing the lowest possible score for each individual question (e.g., “demonstrated technical incompetence”), and the latter the highest (e.g., “demonstrated high technical competence”).

Impacts of the study are measured through three metrics:

1. Students’ initial peer- and self-evaluations,
2. Students’ response to both positive and negative feedback, or lack thereof, through additional evaluation checkpoints at regular intervals during the semester, and
3. Students’ ability to maintain professional growth in future semesters through similar peer evaluation benchmarking.

Student performance was tracked during four total semesters (Spring 2020, Fall 2020, Spring 2021, and Fall 2021). During the Spring 2020 semester, students were provided with a mono-numerical peer and self-feedback for every major design report in the semester. In the Fall 2020 semester, students were provided with no peer feedback until after the final report was submitted. In Spring 2021 and Fall 2021 semesters, students were provided detailed reports on their peer-evaluated performance.

Results and Discussion

As mentioned above, the peer evaluation survey developed at University of Florida was distributed to students in COURSE1 after each of three major design reports. Data collected from survey submissions were analyzed and the results are in Table 1.

Table 1: Tabulated results from four semesters of COURSE1 peer evaluation data.

Report	Spring '20 ⁺				Fall '20			
	Average	Std. Dev.	Minimum	Maximum	Average	Std. Dev.	Minimum	Maximum
CDR	3.626	0.480	0.933	3.867	3.870	0.308	1.371	4.000
PDR	3.644	0.440	1.022	3.867	3.803	0.378	1.000	4.000
FDR	3.685	0.472	0.933	3.867	3.781	0.387	1.000	4.000
Report	Spring '21				Fall '21			
	Average	Std. Dev.	Minimum	Maximum	Average	Std. Dev.	Minimum	Maximum
CDR	3.700	0.472	1.108	4.000	3.621	0.364	1.667	4.000
PDR	3.587	0.554	1.052	4.000	3.608	0.402	1.555	4.000
FDR	3.552	0.619	1.042	4.000	3.606	0.646	1.333	4.000

Table 1 displays the key results obtained from data collection from Spring 2020 through Fall 2021. Although data were also collected in Summer semesters, these data were not utilized for this research due to their low sample size. Data from Spring 2020 has been highlighted (+) due to the sudden transition to online instruction mid semester at the beginning of the COVID-19 pandemic. The Fall 2020 and Spring 2021 semesters were also conducted in an online learning environment; hence, the Fall 2021 semester was the first primarily in-person semester since March 2020. The average values reported in Table 1 were obtained by averaging each student's score for the 12 survey questions. The standard deviation (std. dev.) was calculated from the students' individual average scores, rather than the average score for each question. The minimum value reported was the lowest individual average score received by a student that semester, and the maximum reported was the highest individual average score.

As described above, Spring 2020 semester's students were provided with a mono-numerical score resulting from their peer and self-evaluation surveys. Fall 2020 received no feedback until after the final report was submitted. Both Spring and Fall 2021 semesters were provided with a detailed breakdown of their score for each of the 12 survey questions, as well as a personal average and standard deviation.

It should be noted that most of these data was collected during semesters where instruction was primarily online due to the COVID-19 pandemic. The Spring 2020 CDR was the last report submitted in an in-person learning environment prior to the Fall 2021 semester. It is suspected that the combination of online instruction and the pandemic setting led to the scores being higher than during an in-person semester. This phenomenon is occurring even though the quality of work submitted by students would be equivalent to that produced in a live teaching environment¹¹. Other studies show similar impacts on the perception of student work resulting from mandated switch to all-online instruction due to COVID-19¹².

As shown in Figure 1, the average peer evaluation score tended to decrease as each semester progressed, with a notable exception in Spring '20. This trend may be attributed to a few factors. Typically, as semesters progress, course difficulty and workload increases (i.e., in COURSE1, the PDR is more rigorous than the CDR, and naturally the FDR, being a *final* report, is more rigorous yet). In combination, an online learning environment isolates students and strains communication; having a more challenging task at hand with a simultaneous reduction in communication effectiveness may lead to more frustration, and consequently, lower peer evaluation scores.

Another factor to peer evaluations decreasing with each report submission is fatigue experienced due to multiple courses increasing in difficulty simultaneously affecting student performance in all of them. This also implies that providing feedback to the students in COURSE1 did not affect the results collected from each report's peer evaluation survey, thus suggesting students make no effort to improve the areas in which they might have underperformed previously. As noted above, the only semester that differs from this trend is that of Spring 2020, where the class was transitioned from in person to online due to the COVID-19 in early March. This increasing trend is hypothesized to have occurred due to a combination of grade inflation and students being more lenient in their peer evaluations due to the natural turbulence of the online transition.

2022 ASEE Southeastern Section Conference

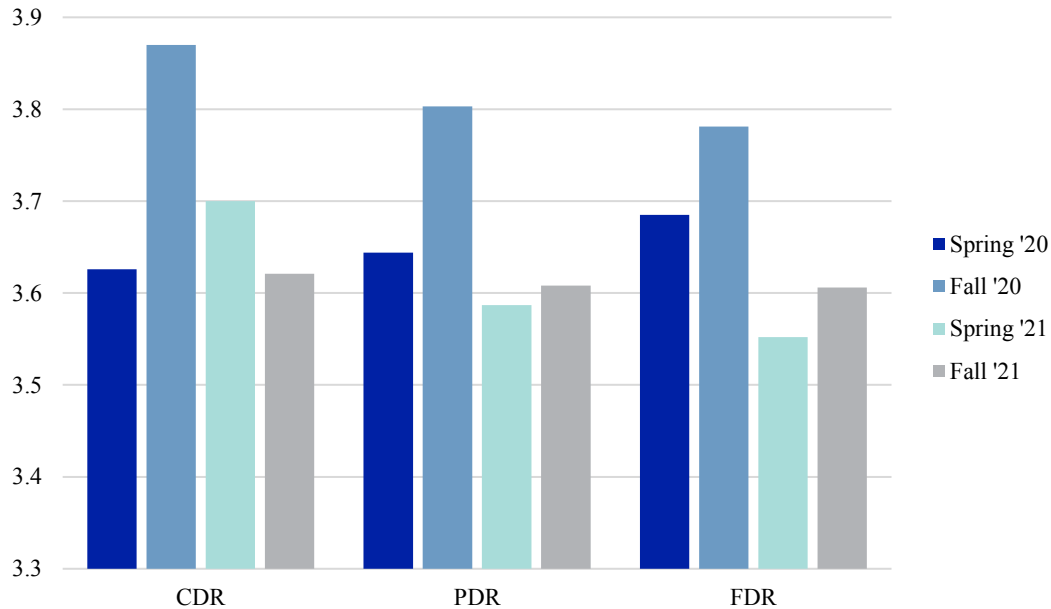


Fig. 1: Comparison of four semesters' average peer evaluation score for each of the three report submissions. Peer evaluations are grouped based on the corresponding report submission. A general downward trend of peer evaluations throughout the semester can be noted for all terms other than Spring 2020, which is considered an outlier due to mid-semester onset of COVID-19.

Although Fall 2020 still portrays a negative linear trend, the average score for each report is much higher than those of any other semester. This is hypothesized to be due to Fall 2020 being the first semester conducted in a fully online learning environment. Because of this, students tended to be more forgiving in their peer evaluations. Resources were limited during online semesters, and communication more infrequent when compared to in person environments. The compounded result of this was increased student satisfaction from fewer individual group member contributions.

By comparison, the Spring 2021 peer evaluation average scores were closer in value to those of Spring 2020, thought to be a result of students becoming familiar with the online environment and thus less tolerant of "free riding" behavior in their groups. Furthermore, although COURSE1 lectures were strictly online, most students moved back to campus as facilities were reopening. This resulted in many groups meeting in person for group reports, amplifying the impact of both underperforming and high-performing group members.

The Fall 2021 semester was conducted primarily in person, with one COURSE1 section remaining online, but with full access to course laboratory facilities. Consequently, the resulting peer evaluation average scores were rather consistent across all reports. Behavior was more monitorable, individual absences and/or participation was more noticeable, and students' standards for the work quantity and quality largely returned to pre-pandemic benchmarks, resulting in more consistent peer evaluation scores.

A two-part trend was noticed through interactions with students and discussions regarding peer evaluations. Students would tend towards harshly assessing any group members showing signs

of free riding; however, they would trend towards awarding the highest scores for everyone else, even if a group member was not particularly amazing. This can be seen through each individual report having an average score greater than 3.5 on a scale of 4; however, there were still multiple individual average scores more than two standard deviations below the mean. It has been studied and proven that student evaluations for professors tend to be higher the easier that professor is with grading, hence implying that the higher course evaluation results are “bought” by the lack of harshness by the professor¹³. Thus, a similar statement can be inferred about students whose performance is barely passable but completes their obligations as a team member. It is as if the high peer evaluation scores are “bought” by the lack of causing imbalances of workload within the group, resulting in the inflated scores reported above.

Conclusion & Future Work

As the study progresses, the goal is to implement the peer evaluation survey and feedback analyses in COURSE2 and follow the students from COURSE1 to assess their professional growth while evaluating the efficacy of the evaluation tool. The study also aims to highlight and track individual students that perform in the upper and lower percentiles of their peer evaluations to determine if there are individual trends that diverge from the course averages. There is also consideration to implement the peer evaluation tool in a sophomore level design class within the curriculum so that the period of data collection can be increased, and students can be assessed across multiple years. Although the trends discussed indicate that students’ performance did not significantly adapt to detailed feedback, more individual student tracking needs to be performed before this statement can be validated.

The trend of decreasing average performance as the semester progresses poses a question of thought: are the students with habits coinciding with “free rider” behavior exclusively underperforming in a singular group-project course, or are these indicators consistent across other senior level labs? Implementing the peer evaluation tool in other lab classes that incorporate group work may help determine if these behavioral markers are appearing due to the subject matter versus a student population’s mindset.

References

- 1 ABET, Criteria for Accrediting Engineering Programs, 2018.
- 2 S. Odom, B. Glenn, S. Sanner, K. A. Cannella, “Group Peer Review as an Active Learning Strategy in a Research Course,” *International Journal of Teaching and Learning in Higher Education*, Vol. 21, No. 11, pp. 108-117, 2009.
- 3 J. Tenenber, “Factors affecting free riding on teams: Implications for engineering education,” *International Journal of Engineering Education*, Vol. 35, pp. 1703-1724, 2019.
- 4 K. S. Double, J. A. McGrane, T. N. Hopfenbeck, “The Impact of Peer Assessment on Academic Performance: A Meta-analysis of Control Group Studies,” *Educational Psychology Review*, Vol. 32, pp. 481-509, 2020.
- 5 R. Hardin, G. Cullity, “The Free Rider Problem,” *The Stanford Encyclopedia of Philosophy*, 2020.
- 6 M. W. Ohland, R. A. Layton, M. L. Loughry, A. G. Yuhasz, “Effects of behavioral anchors on peer evaluation reliability,” *Journal of Engineering Education*, Vol. 94, pp. 319-326, 2005.
- 7 M. W. Ohland, M. L. Loughry, D. J. Woehr, L. G. Bullard, R. M. Felder, C. J. Finelli, R. A. Layton, H. R. Pomeranz, D. G. Schmucker, “The comprehensive assessment of team member effectiveness: Development of a behaviorally anchored rating scale for self- and peer evaluation,” *Academy of Management Learning & Education*, Vol. 11, No. 4, pp. 609-630, 2012. <https://doi.org/10.5465/amle.2010.0177>
- 8 M. L. Loughry, M. W. Ohland, D. D. Moore, “Development of a Theory-Based Assessment of Team Member Effectiveness,” *Educational and Psychological Measurement*, Vol. 67, pp. 505-524, pp. 2007.

2022 ASEE Southeastern Section Conference

- 9 S. Şahin, "An Application of Peer Assessment in Higher Education," *The Turkish Online Journal of Educational Technology*, Vol. 7, No. 2, pp. 5-10, 2008.
- 10 K. J. Reid, "Development of the Student Attitudinal Success Instrument: Assessment of first-year engineering students including differences by gender," *Purdue University Doctoral Dissertation*, 2009.
- 11 E. Karadag, "Effect of COVID-19 pandemic on grade inflation in higher education in Turkey," *PLoS ONE*, Vol. 16, No. 8, 2021. <https://doi.org/10.1371/journal.pone.0256688>
- 12 L. E. Rogers, K. J. Stubbs, N. A. Thomas, S. R. Niemi, A. Rubiano, M. J. Traum, "Transitioning Oral Presentations Online in Large-Enrollment Capstone Design Courses Increases Panelist Participation," *Advances in Engineering Education*, Vol. 8, No. 4, Fall 2020.
- 13 H. Rosovsky, M. Hartley, "Evaluation and the academy: Are we doing the right thing," *American Academy of Arts and Sciences*, ISBN#: 0-87724-030-2, 2002.