Effect of Peer Evaluations in Influencing Behavior in a Laboratory Course

Amy Eastburg

This paper was completed and submitted in partial fulfillment of the Master Teacher Program, a 2-year faculty professional development program conducted by the Center for Faculty Excellence, United States Military Academy, West Point, NY, 2012

Abstract

This paper looks into whether or not peer evaluations can motivate members in a group to perform better and also discussing other advantages on conducting peer evaluations. In many classes students are forced to work together in groups, and receive the same grade regardless of each individual’s effort in the group. Some group members contribute a large amount of work in an effort to receive a better grade or to prevent letting others in their group down. Other students may put forth little to no effort and receive the same outcome. Peer evaluations are a method to evaluate how members of a group perform together. In addition to influencing behavior, peer evaluations can also provide another perspective to the instructor.

Introduction. Peer evaluations are a method to distinguish the level of effort and amount of work from individuals in a group from the perspective of a student. At many colleges, students take a laboratory class as part of their major. As the students are generally interested in their major, they each tend to contribute a fair amount because they are interested in the subject and want to do well. At USMA, all cadets are required to take laboratory classes, including physics, as a requirement to earn a bachelor of science. Since these classes are required but are not in the students’ major, students may not be as motivated to do well and work hard. In the lab portion of the introductory physics class, cadets are forced to work together and each receives same grade, regardless of the effort that each individual puts into the work. While some cadets are motivated by grades and have an interest in learning, other cadets are only taking the class since it is a requirement. The latter group may put in little effort and time. The issue is how to motivate these students, and how to identify students that don’t contribute an equal share of the work.
One advantage of peer evaluations is they offer another perspective of individual performances. Instructors are limited to observations including exam performances and class participation. Even in a group or lab setting, students may appear to be performing well with others in the group, however the same students may not assist others outside of the classroom as part of a lab write up or project[1]. I have experienced some cases where cadets have performed well during class and are very approachable and interactive with the instructor during lab and in class, but receive poor ratings from their peers due to their lack of work conducted outside of the classroom.

Research has indicated that the peer evaluations reflect the perspective of peers who are in close contact and are familiar with the behavior and group characteristics of others in the group that are not readily apparent to the instructor [2,3,4]. Peer evaluations can also be limited by the lack of participation or resistance of students, as their active and honest participation is required for peer evaluations to be effective[1]. Peer evaluations can also be affected by friendship bias where group members may rate each other higher based on preexisting relations instead of based purely off their performance on the group work[2]. Group members may also reluctant to differentiate the performance of peers when the ratings may disturb a positive working relationship[3]. Often during peer evaluation, students will rank everyone evenly, even though the workload was not evenly distributed. In addition there are also cases where there is a minimal distinction between different members of a group, even when one or two individuals contribute the majority of the work. There have been cases where cadets have complained about others in their group, however they still rated each other as being equally, or significantly higher than they deserve.

Another drawback of peer evaluations is that poor performers are susceptible to rating inflation when they provide self ratings. Below average performers may lower the ratings of others in their group in an effort to maintain their own self-esteem [3]. These performers may have a false perception about their ability and contributions, or may be unwilling to inform others of their shortcomings of the group and unwilling to take responsibility for their lack of actions.
At Quinnipiac University, a study was conducted on using peer assessments to reduce free riding in group projects [4]. In this study, peer evaluations were conducted at multiple times during a course. After the first evaluation was conducted, there was a significant reduction in the number of free riders. However there was no change in the number of free riders between the second and third iteration of the evaluation. Group members received a grade on a project based on the feedback from peers and also the instructor’s grade on the project. Using this formula, students who conducted the majority of the work can potentially received a grade above 100 percent[4]. When there is a large portion of a grade that is affected by peer evaluations, students are more likely to perform at a higher capacity. However when peer evaluations are only responsible for a fraction of the number of points a project is worth, there is less motivation to perform and contribute, limiting the ability to reduce the number of free riders.

In addition to academic settings, research on the effectiveness of peer evaluations has also been conducted in the military. Peer evaluations were shown to be effective in predicting the performance of senior level officers [5]. In another study of Special Forces training, peer rankings predicted final training outcomes better than the staff ratings. In training, the presence of the supervisor will influence the person to perform at a higher level compared to when they are not being evaluated by supervisors. The peer evaluations related to characteristics such as personality, effort and interpersonal skills, instead of non personality characteristics such as knowledge which are commonly evaluated by supervisors. Motivation and interpersonal skills are important characteristics both in the military and other organizations. In the study, peer ratings were able to be a better predicted of future performance, as peers placed more importance on motivation [6].

As an instructor in the physics department, one is allotted with a minimal number of discretionary instructor marks to allocate, roughly ten percent of the total marks available. The marks can be used by assigning homework, quizzes, and other events. Another use of the marks is to distribute them based on peer assessments of their lab partners. The marks assigned for the peer evaluations were worth between 0.6% and 1.0% of the total marks for the course. As the amount of marks available is a fraction of that
compared to the study conducted at Quinnipiac University, the marks alone may not be enough of a motivation for many. Even with the minimal amount of marks, peer evaluations can provide a minimal amount of satisfaction to the cadets who contributed more than their peers.

**Methodology.** Lab peer evaluations were conducted over a period of two semesters, evaluating approximately 65 cadets per semester during an introductory physics course at the United States Military Academy. The first iteration was conducted at the end of the semester and cadets hid little advanced warning. Cadets chose their lab groups at the beginning of the semester and were informed that they had the option of switching lab groups, however only a couple of cadets switched lab groups, which was due to a lab partner leaving the academy. Cadets filled out the form below evaluating themselves and other cadets in the laboratory group. For the second semester, peer evaluations were conducted midway through the semester and again at the end of the semester. Cadets were notified at the beginning of the semester that they would be evaluating their lab partners for instructor marks. Once again cadets chose their own lab groups at the beginning of the semester. Lab groups evaluated their peers using the Lab Program Peer Evaluation Sheet (Figure 1). Cadets were also given the option to identify any cadets they would specifically want to work with in the future and not want to work with and an explanation for the choice. For the evaluations conducted during the second semester, cadets were also asked if their effort in lab changed knowing they would be evaluated by their peers.
LAB PROGRAM PEER EVALUATION SHEET

Rate all members of your lab group in the following categories including yourself. Please provide me candid and frank feedback. These forms are for my eyes only.

<table>
<thead>
<tr>
<th>Group Members:</th>
<th>CDT</th>
<th>CDT</th>
<th>CDT</th>
<th>CDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate average effort on any given lab report. (1 = no effort, 10 = max effort)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Estimate percentage of work done on an average lab by each group member. (must add to 100%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate each member on “ease to work with” (1 = impossible to work with, 10 = easy to work with)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assign a letter grade to all members of your lab group based on your semester’s experience.</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1. Lab Program Peer Evaluation Sheet.

Results.

Conducted at end of semester with no advance warning. The first semester the peer evaluations were conducted, cadets were not notified until the end of the semester that they would be conducting peer evaluations. Cadets chose their own lab group during the first lab and generally maintained the same lab group throughout the semester. There were 65 cadets evaluated in four sections.

One of the four sections evaluated appeared to have lab groups that worked well together and rated each other relatively well. In this section, cadets gave each other an effort between a 7-10 for effort, split the level of work relatively fairly rated group member as easy to work with, and generally gave each other a grade of an “A” for lab.

In the other three sections, there averaged at least one person in each lab group who gave significantly less effort and completed minimal work on the lab compared to their peers, and could be considered a “free rider”. On the peer evaluations, there were 13 cadets (out of 67 instructed) who were rated significantly lower than their peers. These cadets received a rating between 1 and 5 from their peers for effort, but rated themselves on par with the other members of their lab. Lab partners also stated these
lab partners were only responsible for between 1-10% of the work on the lab, as opposed to their share of at least 25% depending on the size of the lab group, and were commonly given a rating below 5 on the ease to work with. Additionally, there were multiple cases of cadets feeling that these underperformers deserved a “D” or an “F” in lab.

Many cadets described these underperforming cadets with the following comments “lazy”, “distracting”, “not a team player”, “doesn’t work at all”, “never did anything”, “doesn’t pay attention”, “doesn’t like to work”, “easily distracted and unreliable”, and “did little work and complained”.

Many cadets were frustrated working with the same groups for the entire semester, but did not make an effort to change the lab groups. Providing the peer evaluations allowed these cadets an opportunity to vent frustration in working closely with others during the semester, however as the marks were only worth approximately 0.6% of the total course, the peer evaluations had minimal impact on their grade. Based on the feedback received from the peer evaluations especially with the relatively large number of free riders during the first semester, I decided to conduct peer evaluations midway through the semester with the intent of switching lab groups as needed. The goal of switching the lab groups would be to improve cohesion with groups by allowing cadets that worked well together to be grouped together, but also forcing free riders to be grouped together to force these cadets to put forth an effort during lab.

**Conducted midway with advance warning.** The second semester I conducted lab peer evaluations, they were conducted midway through the semester and again at the end of the semester. Cadets were notified at the beginning of the semester that they would be evaluating their lab partners for instructor marks. Cadets chose their own lab groups, and were also told that they may move lab groups, as long as all groups contained between three and four cadets. However cadets may find it challenging to switch lab groups or move unproductive people out of their lab group, as I saw minimal movement with the exception of cadets who were absent and making up labs.
For the peer evaluations conducted midway through the semester, most cadets tended to rate the effort of those in their lab group between a 6 and a 10. Based on reading and grading the labs, there appeared to be very few groups that put in the maximum effort. Out of 67 cadets evaluated midway through the semester, there were three in one section who had received more than one rating between 1 and 3, indicating limited effort exerted on the lab. However these three cadets had all rated their effort between a 7 and 9. Peers had also reported that there three cadets completed between 0-7% of the lab on average for a four person lab group, and received ratings of 1-3 on the ease to work with. The underperformers also were assigned a lab letter grade ranging from a “C” to a “D”.

Cadets also provided general comments about these lab partners and some of the comments included “doesn’t do much work or attempt to learn”, “right now he is free riding”, “contributed very little”, and “didn’t answer emails for help”.

Among the other cadets, there were a few that received lower ratings from their peers on effort. The ratings appeared to be based on more of their limited understanding of the material, as opposed to evaluating their motivation and interpersonal skills.

The main objectives of conducting peer evaluations were to reward positive behavior, and to influence behavior and identify and reduce free riders in the lab. The feedback from the peer evaluations could be used to rearrange lab groups as needed.

Three of the sections were comfortable with the labs groups and did not want to switch, so I maintained the same lab groups. For the fourth section, I placed the three cadets who scored the lowest in the peer evaluations in the same group. This would result in more equity, while also forcing cadets who were “free riding” to do some work. In addition I requested feedback from others about who they wanted to have in their lab group. I used this feedback and comments of the peer evaluation to rearrange the lab groups so there were no more than two people from each original lab group working together.
Follow up conducted at the end of semester with advanced warning. During the third iteration of the peer evaluations, ratings were similar to those conducted midway through the semester. Three of the responses indicated that they had increased their effort in lab knowing that they would be evaluated by their peers.

For the lab group put together based on low peer evaluation during mid-semester, two of the three cadets appeared to contribute heavily to the lab, since no one else was going to do their work for them as during the first half of the semester. One of these two cadets admitted that he changed his effort knowing that he was once again going to be evaluated by his peers. The cadet who still did not contribute much received comments like “never knew what was going on” and “comes down to the lack of understanding”. By rearranging the lab groups, two of the three free riders in the class conducted work.

Other instructors also conducted peer evaluations, using different techniques. Instructors who conducted peer evaluations at the end of the semester did not feel that the peer evaluations influenced behavior at all, there were no apparent changes in sections where cadets were informed at the beginning of the semester compared with the sections who did not know they would be evaluating members of their lab group until the end of the semester [7].

Some instructors conducted peer evaluations midway through the semester and again at the end of the semester. After the first peer evaluations were conducted, instructors either switched the groups at their own discretion based on the peer evaluations, or informed the lab groups that they needed to switch up lab groups on their own. When the lab groups were changed, sometimes there would be no one in the new group who knew how to conduct part of the lab particularly involving calculations. These newly formed lab groups became more of a burden on the instructor who had to provide additional assistance[8]. Although it only impacted a few people, rearranging the lab groups caused cadets who were previously free riders to work. However in a group with cadets who had all previously been free riders, often there will still be at least one person who fails to contribute to the group. In some cases when lab groups were
told they needed to change up on their own, some cadets would essentially plead with other cadets stating that they would be willing to work harder if they could be in their lab group [9].

In general, instructors feel that the peer evaluations have limited effectiveness in influencing behavior as the marks associated with the evaluations are generally worth less than 1% of the grade in the course. Instructors feel that one of the major benefits of the peer evaluations is that they give cadets a chance to voice their opinion and frustrations.

**Observed Limitations.** One of the limitations of the peer evaluations is that it requires honesty of the students. There have been times when cadets have complained about other members in their lab group not contributing at all. However, during the peer evaluation they gave the other lab partners higher scores compared to what they should have received. Cadets were given the opportunity to rate the effort of others in their lab group on a scale of 1 to 10. There have been a couple of instances where cadets will inform an instructor that another cadet in their lab group did not contribute anything to the lab. However at the end of the semester when the peer evaluations were conducted, they may still give those members a 5 or a 6, whereas other lab groups have given that individual a rating of a 1.

It is also quite common to see all members in a group rated evenly, when the work wasn’t likely split evenly. Based on this, there is likely to be a higher number of cadets who contribute minimally, but are not identified by their peers. Cadets may feel that by rating everyone evenly, everyone will receive the maximum number of points. Cadets may also be unwilling to cause any kind of disturbance or upset anyone in their group.

It is also quite common to see rating inflation when poor performers rate themselves. These individuals receive low ratings from their peer with comments about their lack of effort and understanding. When these individuals rate themselves, they may give themselves a rating of 7-10, and there may not be any significant difference in the ratings of themselves compared to their peers who contributed the majority of the work.
Overall, the ratings tended to be inflated regarding the average effort of all individuals in each lab group. On more than half of the peer evaluation sheets, at least one person received a maximum rating of 10 for their effort. Based on the effort observed while conducting lab, as well as identifying numerous errors while grading the labs, there have been very few lab groups that actually put in a maximum effort.

**Discussion.** There are several ways that marks can be assigned for peer evaluations. Marks can be distributed such that the person who receives the highest score will receive the maximum number of marks. This method can be used to distinguish between those who received high and low rating on the evaluations. One of the drawbacks of this method is that cadets will often rank each other equally in an attempt to receive the most marks, and nothing is done to reward those who contributed the most and penalize those who contributed the least.

By combining the average effort from the peer rating with the lab grade, it is possible to provide a minimal amount of equity. The majority of marks are based off of each individual lab report, the peer evaluation are only worth a fraction of the points for the lab and course grade. However this method can be used to penalize someone with a small amount of points based on their lack of contributions which will lower their grade in the course. As a result, those in the same lab group who had significant contributions are able to receive a score greater than 100% which would increase their grade and overall standing in the course. When this was conducted at Quinnipiac University, students tended to increase performance after the first iteration of peer evaluations.

Based on the evaluations conducted, there was a minimal effect of students performing better in lab when they knew they would be evaluated by their peers. Although only a few cadets admitted the peer evaluations affected their performance, cadets appeared to work better and distribute more of the work when they knew that they would be evaluated by their peers. There were 13 cadets during the first semester who were rated extremely low compared to three cadets during the second semester when they were told of the first day of class that they would be conducting peer evaluations. Since the peer
evaluations were taken at different times during the semester, cadets could also have had more frustration working over the entire semester and could have been grading each other harsher, compared to only working with the lab group for half of the semester. Additionally, during the second iteration of the peer evaluations, cadets have already completed a semester of physics labs which may influence behavior.

In addition to the minimal impact of affecting cadet performance, peer evaluations also provide several benefits. Although the peer evaluations are worth minimal marks as part of the course, they are able to provide some sense of fairness and equality, particularly for those cadets who conducted the majority of the work in the lab. In addition it gave them an opportunity to voice their opinion and frustrations of working with others.

As an instructor, there are also benefits to asking for peer evaluations. The peer evaluations give an opportunity to evaluate how cadets are doing in the lab program. The cadets who conduct minimal work can be identified and then placed in a group together. By doing this, it forces at least one cadet to figure out what is going on in the lab group, and decreases frustrations in other labs groups who had previously worked with these low performers. The points associated with peer evaluations is often not enough to change performance but knowing that they may be forced to work with others who are also minimal contributors would force at least one person into contributing more.

Peer evaluations also provide additional insight on some of the cadets. This information can be used for assessing interpersonal characteristics such as motivation and effort and evaluating their potential to become a commissioned officer. Many of the cadets that received low ratings for their peer evaluations were delinquent in another area. These cadets were reviewed by a board to determine their potential to be a future officer and remain at the academy. When peers state that a cadet is lazy, not a team player, and did little to no work, these characteristics are likely to carry over in other areas and are indications that they may not be a suitable officer. Often cadets will appear to be outgoing in lab and class having positive interaction with the instructor and other cadets. As an instructor, the comments on the peer
evaluations can be extremely useful in evaluating potential to serve as an officer in the United States Army. Peer evaluations could also be used to evaluate potential for graduate work. Additionally, the peer evaluations can also be used for a positive recommendation for cadets who are deficient in some areas

Based on experience in the military, I have observed other officers who will put forth minimal effort into their duties, often pawning off their responsibilities on others. I am able to identify with the frustration of cadets in dealing with others who contribute very little and receive the same grade, or in the case of the Army, the same salary. I feel that the peer evaluations can be used to identify these people and used as a tool to assess future performance.

Others may not see a benefit of the peer evaluations. As identified earlier, there are limitations with conducting peer evaluations. In addition in the military and in society in general, one may often find themselves working with someone else who puts in a significantly different amount of effort or has significantly different motivation. By forcing cadets to work together we can prepare them for future experiences with unequal workloads [10].

Each instructor in the Physics Department has the flexibility to establish lab groups as they choose. Some will randomly assign lab groups once at the beginning of the course, others may randomly assign new lab groups for each lab. In addition, due to cadet absences, some cadets may make up the lab with another group limiting the amount of time they work with the same group. Because of this, there is no way to standardize the use of the peer evaluations throughout the course.

**Conclusion.** Based on the weight of peer evaluations in the course, peer evaluations have limited impact to motivate students to work harder, and influence behavior. However when peer evaluations are conducted midway through the semester, they offer the instructor a tool to modify lab groups and minimize free riders in the classroom, and instill a sense of fairness in the grading for some. At the beginning of the semester, I plan on sharing my experience of using the peer evaluations and include some of the negative comments as well as my plan to put free riders together in an attempt to discourage
the unwanted behavior. As identified in other studies, peer evaluation can be used to assess interpersonal skills, including motivation and effort, which can be used to assess a cadet’s future performance in the military.
Works Cited


