ASEE AMERICAN SOCIETY FOR ENGINEERING EDUCATION

Paper ID #45426

Comparison of Teamwork Assessment Methods in Engineering Classes

Mr. Adam Barnes, University of Virginia

Adam Barnes graduated with a BS and MS in Electrical Engineering from Virginia Tech. He worked in small business and industry for 18 years before returning to Virginia Tech to teach engineering as a professor of practice. He then moved to Charlottesville and began teaching at the University of Virginia in the Electrical and Computer Engineering Department, where he has now been for 5 years.

Comparison of Teamwork Assessment Methods in Engineering Classes

Introduction

The ability to work well in teams is consistently one of the most sought-after skills by employers, and so deserves serious attention in higher education courses [1]. To evaluate the effectiveness of any program or module designed to teach teamwork requires a valid assessment tool. This is particularly important in engineering disciplines which must assess teamwork as part of their ABET accreditation. Because teamwork assessment is often subjective, it can be difficult to evaluate rigorously.

Assessing the work produced by a team is not necessarily a good indication of the teamwork itself [2]. Peer assessment offers a more direct measurement of teamwork, often done with a survey of team members. Ideally, surveys would be conducted at multiple points during the project, although frequently they are only performed at the end. The survey is relatively easy to implement, and there are a number of survey tools that streamline the process and make it applicable for a wide range of class sizes and subjects. However, the accuracy of the survey assessment by the students themselves can be problematic. First, results can be skewed by the phrasing of the survey questions [3]. Second, many students don't know what good teamwork really looks like; either they have not worked on a good team in the past or their team was not challenged to the point that the quality of the teamwork came into play. Third, students' assessment of their peers may be too lenient, because they don't want to hurt their teammates' grade or feelings. Conversely it may be too harsh if the team members had a serious falling out or if one of the team members is prejudiced.

An alternative teamwork assessment is for the instructor to evaluate the students on their contributions. This provides a more objective view but is often difficult to implement as it requires a considerable time commitment on the part of the instructor, which is hard to justify for classes where teamwork occurs but is not the primary thrust of the class, as is the case in many engineering classes. Additionally, the instructor may not be privy to internal team dynamics.

In this paper I will compare the survey method with direct observation by the instructor in an effort to determine how reliable each method is, as well as their relative merits and limitations.

Methodology

The data for this paper is taken from the Fall semester 2024 Capstone Senior design course. Part of the project process is for each team to meet with the instructor on a weekly basis. This afforded the opportunity to regularly interact with each team and directly assess team performance. In parallel, the students filled out teamwork surveys, providing a basis for comparison. This provides two assessment tools that can be compared, one based on the instructor's evaluation, and one based on the students' evaluation.

The class consisted of 87 students broken up into 19 teams of 4 to 5 students each. Students in our program are encouraged in their third year to begin thinking about their Capstone projects so that they don't spend too much time at the beginning of the Capstone class trying to figure out what they want to do for their project. As a result of this preplanning, most groups were self-assembled, although a handful of unattached students were assigned to teams in the first week

and a half of classes. The expectation is that students will be more lenient evaluating their friends, skewing assessment results, which is part of the impetus for this study.

Weekly Evaluations

The instructor met with each team every week to discuss progress on their project, answer questions, and model meeting protocol. Students were expected to fill out a brief (one paragraph) summary of work they did during the week, and what they intended to accomplish in the coming week. The weekly report and instructor evaluation of student engagement provided the basis for instructor assessment of teamwork. Each student was assessed in three categories each week, chosen to promote professional behavior in the students and provide them with graded feedback throughout the semester. They were rated on a Likert scale from 1 to 5 in the following categories:

- *Punctuality:* Team member is present and ready to start at the agreed upon meeting time. This category was included partly for logistical reasons, but also as a proxy for organization and respect for the team. When students were late, the rest of the team was asked where the missing student was to see if they had communicated with their team as to why they might be running late.
- *Meeting Preparation*: Team member is fully prepared for the meeting, knowing what needs to be discussed, and with relevant files/materials readily available. The individual weekly report is completed. This category also acts as a proxy for organization and respect for the team, as well as making the meeting run more smoothly.
- *Weekly Progress*: A check to see if goals from the previous week were completed satisfactorily. This category was used to keep students on schedule and check individual contributions throughout the semester.

Team and Team Member Assessment

Each student will be evaluated by their fellow teammates and by the instructor in the following categories:

- *Professionalism*: How well did each team member maintain the appropriate level of professionalism throughout the project? Behaviors to consider: shows respect towards team members, arrives at meetings on time, meets deadlines, uses applicable technology, etc.
- *Communication*: How well did each team member communicate in the appropriate manner for this project? Behaviors to consider: uses academic dialogue, listens to other team members, offers and accepts feedback constructively, facilitates positive interactions among team members, etc.
- *Content Knowledge and Skills*: Did the team member have or acquire the knowledge and skills necessary to contribute to the successful completion of the project? Behaviors to consider: understands common vocabulary, completes background readings or research, knows how to use necessary equipment or technology, has ability to fill multiple roles within the group, etc.
- *Work Ethic*: Was the team member dedicated to completing the amount of work necessary to achieve the goals of the group?
- *Overall Contribution*: How much did each team member contribute to the overall successful completion of the project?

- *Feedback*: Comments and justifications for scores provided in the other categories. This had no numerical value but provided insight into the students' thought processes.

Additionally, the instructor and each student rated the team performance in the following categories:

- *Work Distribution*: Individuals on a team might be doing different jobs, but everyone has to pull their weight. How well was the work distributed in your group?
- *Team Organization*: Good teamwork requires coordination and communication. How well did the team work together?
- *Team Dynamics*: A good team experience often depends on how well team members get along. How would you rate your interactions with your team? Consider the overall dynamic; if there was just one problem team member that should be detailed in that individual's evaluation.

Each category is rated on a Likert scale from 1 to 9, except for the Feedback category which was a short answer/essay format. In previous semesters, the use of a 1 to 5 scale resulted in a lot of 'perfect' scores of 5 as students seemed reluctant to dock their teammates by 20% for minor infractions. The 1 to 9 range was chosen to give the students a more granular rating scale. Prompts were given for each of the odd Likert entry, as detailed in the Appendix.

The evaluation was done twice, once in the middle of the semester and then again at the end. The first evaluation was intended to expose the students to the teamwork assessment rubric, give them a feeling for how their team was initially performing, and allow time for course correction if their performance was substandard. The final assessment factors more heavily into their grade and is used for ABET assessment. The data presented is based on the final assessment.

Assessment Comparison

The two teamwork assessment techniques are not interchangeable, so in comparing them it is important to identify how they differ on each of the teamwork categories being rated. In general, because the instructor is an outside party observing the team, those evaluations will be more objective. However, because the instructor is not working directly with the students, those evaluations are more of a snapshot and may not be as comprehensive as student evaluations of each other.

- *Professionalism*: Completing work on time correlates well between the instructor and student teamwork assessments. It is expected that students may be more respectful around the instructor than they are with each other however, which may cause this category to be rated higher by the instructor than by the students.
- *Communication*: The instructor is not privy to most of the communication that goes on between team members, so the instructor's rating on this category is more anecdotal from student to student, which makes it less robust than the student rating.
- *Content Knowledge and Skills*: The instructor should get a good feel for students' technical skills competence based on the work they accomplish, skills they learn, questions they ask, and troubleshooting they perform. There should be a good correlation between the instructor and student evaluation of this category.
- *Work Ethic*: This category will be similar to Professionalism in that the students' attitude may present differently to the instructor than the other team members. However, there

should be a strong correlation between work ethic and work accomplished, so the instructor should get a good feel for the student performance in this category, posing notwithstanding.

- *Overall Contribution*: This is an important category for the students, because it can reflect skills like leadership and creativity that are important for team function but difficult to pin down in a list of accomplishments. Like Communication, the instructor perception of this category will be more anecdotal.
- *Work Distribution*: This category can be difficult to assess because it often compares tasks that are very different from each other. Some tasks will be technically challenging and some will be tedious, some will be quick while others will be time consuming. The instructor's greater experience should give a better feel for the relative effort and time commitment required for various tasks in a broad sense, although the students will better understand the effort needed for the specific task in which they are engaged.
- *Team Organization*: The instructor should be able to evaluate this category well based on the weekly progress of each group and how well they coordinate their efforts and work in parallel on different tasks. The correlation between instructor and student assessment is expected to be good.
- *Team Dynamics*: Team dynamics may differ when the instructor is present, so evaluation of this category will be more subtle for the instructor than the student in most cases.

Results and Conclusions

Results for both individual work and group work as rated by the students and the instructor are shown in the bar graphs in Figures 1 through 3. Overall scores were quite high, which may be unsurprising since the students are all seniors who would not have made it to their senior design course without a high level of maturity and competence. The individual work is broken out into two graphs to highlight the difference between the large number of perfect scores at one end, and a few dramatically lower scores at the bottom end. Figure 1 shows the averages for the whole class. Figure 2 shows only the low-end ratings, which include the few under-performing students. No similar figure for high end ratings is included since they are all perfect scores. Those perfect scores to some extent represent the hard work and dedication of the students, but relative to the instructor ratings, the peer ratings are decidedly less critical. Indeed, students awarded their peers perfect scores over 90% of the time in every category.

Although the Likert scale for all the ratings ran from 1 to 9, for all the graphs the axes only show from 4 to 9 to allow differences between peer and instructor ratings to be more evident.

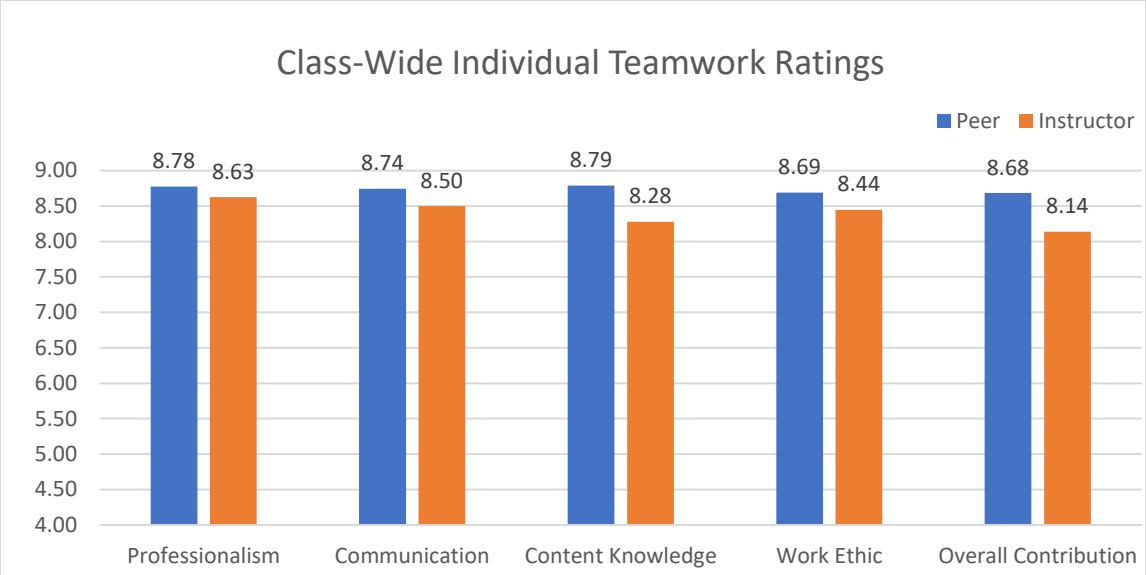

Figure 1. Average ratings of individual students on teamwork metrics across the entire class

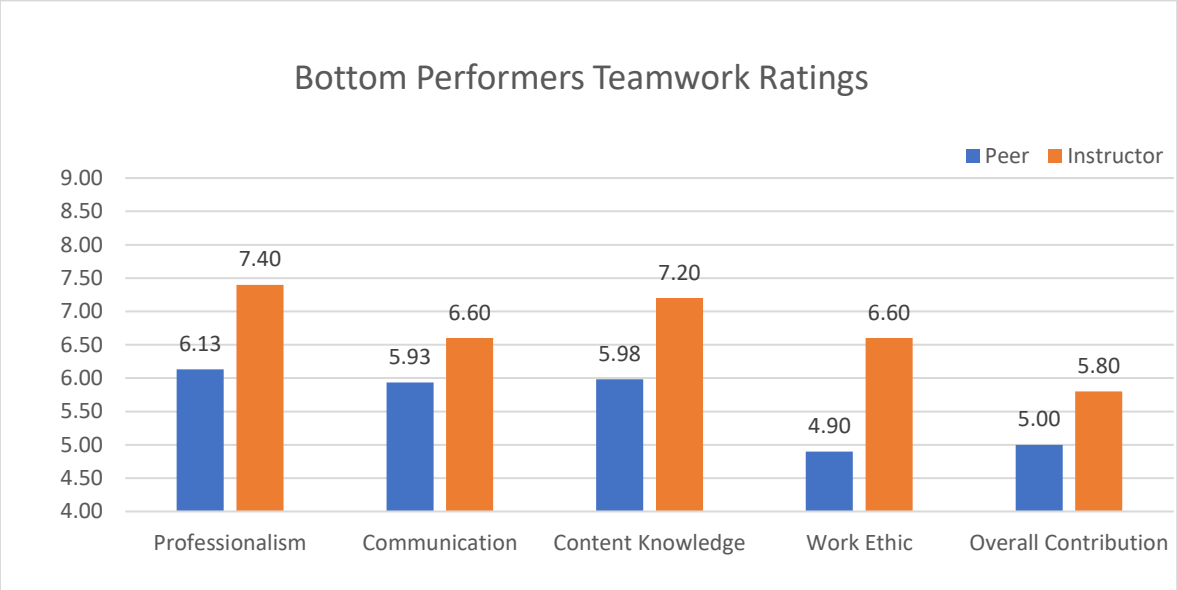

Figure 2. Average ratings of 5 lowest-ranked individual students on teamwork metrics

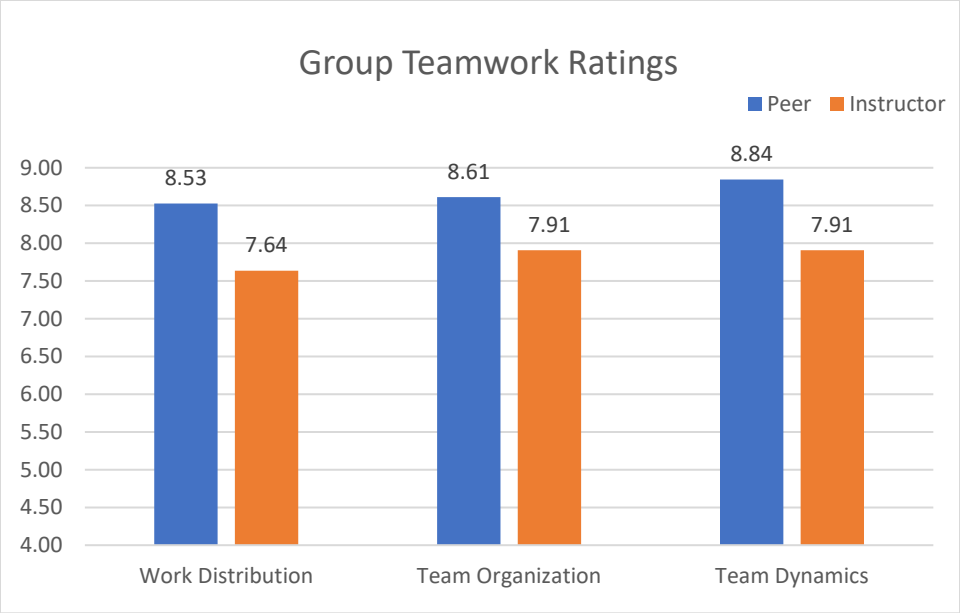

Figure 3. Average ratings of group teamwork performance

Likert Scale: Offering a 1 to 9 scale to provide more granularity did not seem to change the students' inclination to give their teammates perfect scores; over 90% of all the peer-rated scores were perfect. While many of these perfect scores may be justified since they mirrored instructor ratings, the instructor ratings were much more evenly distributed, suggesting that the students' ratings were not as discriminating.

Individual Teamwork Scores: Overall, student ratings were higher than that of the instructor. To some extent this was anticipated since many students were rating their friends. The bias was higher than expected, with over 90% of the class receiving perfect scores from their teammates across the board. The exception is the lowest rated students, where conflict arose on the team. In these cases, some of the ratings were clearly low-balled to 'punish' the under-performing student (e.g. receiving a rating of 1 for every single category). This dichotomy suggests that the instructor ratings were generally more nuanced, being more critical of performance, but still recognized contributions, however slight, from under-performing students.
The two categories that saw the largest difference between peer and instructor ratings were Content Knowledge and Overall Contribution. This presumably arose because the instructor had greater experience with the technologies and tasks at hand, and so was better able to determine when a student was performing well and when they were muddling through. This would argue that having subject matter experts to be involved in the ratings is important.

Teamwork vs. Individual Work: If we set aside the score inflation for a moment, we see the lowest rated metrics were Work Ethic, Overall Contribution, and Workload Distribution. One may surmise that students felt they were doing more work than their teammates, even if their teammates were otherwise performing well, or at least that was the thing that bothered them enough they did not award a perfect score. The instructor ratings of the groups were notably lower than the individual ratings, largely because while individual work was generally of high caliber, the teamwork skills were not as good. Students work well by themselves but are less adept at collaborating with others.

Self-Assembled vs. Assigned Teammates: There were seven students that were added to teams by the instructor rather than self-assemble. The hypothesis was that they would be evaluated more critically because they were less likely to be friends with the group ahead of time. Of the seven, three were rated with perfect scores across the board. Two were rated highly, but with some room for improvement. The remaining two were rated as the worst two students in the class, both by the instructor and their fellow teammates. Disregarding the poorly performing students, the top five of the seven had average scores that roughly track with the average scores for the class overall. Combined with the results of the expanded Likert scale noted above, this would suggest that students are simply not very critical of their peers, or at least that they were able to establish enough camaraderie that they did not want to penalize their teammates. The sample size is small, however, so this result is not conclusive.

Peer vs. Instructor Assessment: Both the peer and instructor assessments of teamwork have their limitations. For this study, the instructor assessment is arguably more accurate simply because it discriminated between good work and superlative work, while the peer assessment largely did not. Additionally, when students did penalize their under-performing peers, they tended to go overboard. Overall, data from this study does not recommend the peer survey of teamwork as an accurate, or at least a discerning, measure of teamwork. This may be in part a fault of the assessment rubric, which was phrased in a way that may have inflated scores. A more carefully worded rubric may produce a more nuanced result. Spending more time in class discussing what differentiates good teamwork from adequate teamwork and/or doing class exercises to rate teamwork might also produce more thoughtful teamwork ratings. Changing the assessment format may provide better results as well, for example, distributing a fixed pool of points among teammates according to the amount they contributed to the project. In any event, this warrants more study, because the peer survey is much easier and less time consuming to implement than an instructor rating of the teams, so having a reliable, accurate peer survey method would be valuable, especially for large classes.

References

[1] Elayyan, Shaher. 2021. “The Future of Education According to the Fourth Industrial Revolution.” Journal of Educational Technology and Online Learning 4 (1): 23–30. https://doi.org/10.31681/jetol.737193

[2] Francis, N., Pritchard, C., Prytherch, Z. and Rutherford, S. (2025), “Making teamwork work: enhancing teamwork and assessment in higher education”. FEBS Open Bio. https://doi.org/10.1002/2211-5463.13936

[3] Van Duzer, Eric; McMartin, Flora, “Building better teamwork assessments: A process for improving the validity and sensitivity of self/peer ratings”, ASEE Annual Conference Proceedings, p 1047-1057, 1999, 1999 ASEE Annual Conference and Exposition: Engineering Education to Serve the World

Appendix

Likert Scales for team and teammate evaluations

Work Distribution RATING PROMPT

1-9 Rating Scale

Individuals on a team might be doing different jobs, but everyone has to pull their weight. How well was the work distributed in your group?

9	Work was very well distributed, with everyone contributing nearly equally.
8	
7	Work was well distributed, although some contributed more than others
6	
5	Work was reasonably distributed, but at least one member underperformed
4	
3	Work was not well distributed, with more than one member underperforming, or only one or two members carrying the group.
2	
1	Work was so poorly distributed that this was not really a group project

Team Organization RATING PROMPT

1-9 Rating Scale

Good teamwork requires coordination and communication. How well did the team work together?

9	The team worked like a well-oiled machine
8	
7	Team worked together well with only minor hiccups
6	
5	Team mostly worked well together, but could have been better organized or communicated better
4	
3	The team often did not work well together, either organizing tasks poorly or not communicating clearly.
2	
1	The team did not work well together at all

Team Dynamics RATING PROMPT

1-9 Rating Scale

A good team experience often depends on how well team members get along. How would you rate your interactions with your team? Consider the overall dynamic; if there was just one problem team member that should be detailed in that individual's evaluation.

9	Excellent camaraderie, we got along well and enjoyed each other's company
8	
7	The team relationship was professional and respectful
6	
5	The team was courteous but not particularly friendly
4	
3	The team relationship was strained and challenging
2	
1	The team relationship was dreadful

Professionalism RATING PROMPT

1-9 Rating Scale

How well did each team member maintain the appropriate level of professionalism throughout the project? Behaviors to consider: shows respect towards team members, arrives to meetings on time, meets deadlines, uses applicable technology, etc.

9	This team member acted in a professional way at all times.
8	
7	This team member acted in a professional way most of the time.
6	
5	This team member acted in a professional way some of the time.
4	
3	This team member rarely acted in a professional way.
2	
1	This team member never acted in a professional way.

Communication RATING PROMPT

1-9 Rating Scale

How well did each team member communicate in the appropriate manner for this project? Behaviors to consider: uses academic dialogue, listens to other team members, offers and accepts feedback constructively, facilitates positive interactions among team members, etc.

9	This team member communicated appropriately at all times.
8	
7	This team member communicated appropriately most of the time.
6	
5	This team member communicated appropriately some of the time.
4	
3	This team member rarely communicated appropriately.
2	
1	This team member never communicated appropriately.

Content Knowledge and Skills RATING PROMPT

1-9 Rating Scale

Did the team member have or acquire the knowledge and skills necessary to contribute to the successful completion of the project? Behaviors to consider: understands common vocabulary, completes background readings or research, knows how to use necessary equipment or technology, has ability to fill multiple roles within the group, etc.

9	This team member's knowledge and skills that contributed to the project were excellent.
8	
7	This team member's knowledge and skills that contributed to the project were good.
6	
5	This team member's knowledge and skills that contributed to the project were fair.
4	
3	This team member's knowledge and skills that contributed to the project were substandard.
2	
1	This team member's knowledge and skills that contributed to the project were poor.

Work Ethic RATING PROMPT

1-9 Rating Scale

Was the team member dedicated to completing the amount of work necessary to achieve the goals of the group?

9	This team member's work ethic was excellent.
8	
7	This team member's work ethic was good.
6	
5	This team member's work ethic was fair.
4	
3	This team member's work ethic was substandard.
2	
1	This team member's work ethic was poor.

Overall Contribution RATING PROMPT

1-9 Rating Scale

How much did each team member contribute to the overall successful completion of the project?

9	This team member's contribution was excellent.
8	
7	This team member's contribution was good.
6	
5	This team member's contribution was fair.
4	
3	This team member's contribution was substandard.
2	
1	This team member's contribution was poor.