Analyzing Peer Assessment Claims for Team Projects in Engineering Courses

RYAN R. SENKPEIL
rsenkpei@purdue.edu
Purdue University

DANIEL M. FERGUSON
dfergus@purdue.edu
Purdue University

MATTHEW OHLAND
ohland@purdue.edu
Purdue University

RUTH WERTZ
rwertz@purdue.edu
Purdue University

ABSTRACT

Ten distinct claims about peer assessment system user perceptions or actions were found to have been made by researchers regarding peer assessments in a team based project. The claims include: the presence of "Free Riders", the ability for peer assessments to motivate students, biased assessments because of social acceptance, raters noticing peers' lack of appropriate skills, positive rater reactions to the specific peer assessment system used, improvements in confidence and skill in rating because of practice rating exercises, biased assessments because of Halo error and social pressures, raters resisting peer assessments because of perceived bias, raters perceiving peer assessments to be unhelpful, and raters hesitation to assess peers because of perceived lack of authority. Several thousand anonymous and volunteer verbatim comments from a peer assessment system used in engineering courses were categorized based on the above claims, and analyzed in an attempt to support or dispute these claims. Only two claims in this study were found to be strongly supported in the verbatim comments made when completing peer reviews (Free Riders and Missing Skills) and no claims on peer assessment can be disputed based on the available evidence. The presence of Free Riders is the most significant peer assessment problem identified in the verbatim comments that instructors must address.

INTRODUCTION

In teaming oriented classes a major goal for instructors is to facilitate the team's interactions in order to keep an environment conducive to learning for all students. One method to assist in facilitating teams is via peer- and self-evaluations[10]. The Comprehensive Assessment of Team Member Effectiveness, or CATME, is a tool used by 4,000+ instructors to assess how effectively team members are contributing to their teams[10]. Instructors utilizing CATME configure a web-based survey to study any combination of five areas of team interaction. Those areas are: contributing to the teams work, interacting with teammates, keeping the team on track, expecting quality, and having relevant knowledge skills and abilities. Student evaluators are asked to rate each team member and themselves in each area indicated by the instructor using a behaviorally anchored rating scale. Student evaluators are also given the option of providing open-ended comments at the end of the survey. These open-ended comments do not have any required form and can be made regarding any subject of the evaluator's choosing. Instructors can use these comments to require evaluators to justify their ratings, or to facilitate discussions between the instructor and students[10].

Given that the comments are volunteered and open-ended our research question was "did student evaluators when writing open-ended comments expose any specific issues they are having within their teams that were previously identified by researchers as common to team interactions and peer reviews". By studying verbatim copies of the volunteered comments we attempted to support or dispute claims about peer reviews and team interactions made by peer evaluation researchers. Any previously identified peer assessment issues identified by examining these volunteer comments will give instructors using a peer assessment system with this comment feature another tool to use to manage their student team interactions and the teaming process.

Literature Review

While often not specifically a discourse on peer review comments, there is much research that discusses the expectations of team interactions and peer reviews. Ohland et al states that "Research shows that many raters, particularly average and below average performers, do not differentiate in their ratings of team members when it is warranted, sometimes because they worry that providing accurate ratings would damage social relations in the team"[10] and also that "[the] 'big five' model, therefore, assumes that team members will have the skills and motivation to contribute effectively to the team, yet the required skills and motivation are frequently key deficiencies in student teams" [10]. Salas, Sims, and Burke explains that the "big five" in teamwork are the core teamwork components and include team leadership, mutual performance monitoring, backup behavior, adaptability, and team orientation[12].

Similar quotes discussing either team interactions or peer reviews, often within undergraduate engineering courses, were used to investigate in this study a several claims regarding

student perceptions or participation in peer reviews. In prior research these claims were made regarding peer reviews by students:

1. Free Riders – The presence of team members that don't contribute, or contribute significantly less than other team members [3,4,5,7,9].
2. Motivate Students – The ability of peer reviews and self-evaluations to cause students to work harder by giving them explicit expectations and a definite goal[13].
3. Social Acceptance – The occurrence of a student rating their teammates higher than expected for fear of being rejected socially[11].
4. Missing Skills – The presence of team members who don't have the necessary skill or knowledge to provide a meaningful contribution to the team[10].
5. Rater/Ratee Reactions – The occurrence of students reacting positively to the CATME system compared to other peer review systems[8].
6. Like Teammates – The occurrence of students rating certain team members higher based on how much they like those team members[10].
7. Why do they Like Teammates - Students will explain to what extent they like other teammates, as a person, teammates, or friend[6].
8. Practice Rating – Students will benefit from repeatedly practicing with the CATME system[2].
9. Social Pressure – Students are being pressured by teammates to give higher rating than were earned[14].
10. Halo Error – The occurrence of students rating teammates higher than expected because they perceive them to be better teammates than they are[15].
11. Keeping the Team on Track – The occurrence of students discussing how well they and their team members keep the team on track[10].
12. Resisting and Bias in Peer Evaluations – The occurrence of students resisting peer evaluations because they feel they are biased because of friendship, popularity, or other factors[14].
13. Peer Feedback Unhelpful – The occurrence of students commenting on how they feel that the feedback they give or receive is unhelpful[1].
14. Impact and Consideration of Grades – The occurrence of students changing their ratings because of how they can affect team member's grades[1].

The specific quotes that correspond to each claim and the locations within each paper where they were found can be seen in Table 3 in the Appendix.

Methodology

The first step of this study was to decide which of the fourteen claims mentioned in prior research should be considered during our analysis of the data. Using our knowledge from observing several years of student peer assessments we focused on the following ten peer assessment claims:

1. Free Riders
2. Missing Skills
3. Resisting and Bias in Evaluations
4. Halo Error
5. Impact on Grades
6. Social Pressure
7. Rater/Ratee Reactions
8. Reviews Motivating Students
9. Practice Ratings Benefitting Raters
10. Peer Feedback being Unhelpful

The 4,000 verbatim comments we used to study these claims came from CATME and were given by students from several freshman and higher level engineering and business classes over the course of four years. The verbatim comments data used in this study were previously counted and categorized in a predecessor project[16]. This prior data categorization organized the verbatim comments into 9 major categories and then into 0-45 sub-categories within each major category. The number of comments within each sub-category was counted. The breakdown of the student comments categorization was as follows:

CATME: Specialization, format, comment box, framing of choices, repetitive, survey format, example, project related questions, schedule format

Faculty Comment: role assignment, survey timing, grade assignment, grade guideline, criterion, task assignment, task definition, task division, supervision

Faculty Practice: team size, role assignment, CATME purpose, CATME, self-rating, task assignment, survey timing, supervision, lack of supervision, team formation, guidance, grade guideline, lack of guidance, number of surveys, criterion, measurement quality

Peer Evaluation Philosophy: measurement quality, rating, self-rating, schedule conflict, general, CATME rating, privacy, lack of professionalism, lack of interest, lack of commitment, lack of contribution, philosophy, CATME, friendship

Team Comment: Collaboration, formation, general, scheduling conflict, communication problem, commitment, intellectual diversity, leadership conflict, conflict, team composition, team combination, character, team characteristics, task assignment, role assignment, coordination,

combination, skill diversity, compensation, team size, lack of coordination, lack of communication, lack of commitment, lack of interest, lack of skills, lack of motivation, lack of team spirit, lack of knowledge, lack of consensus, lack of project understanding, sharing, conflicting options, maturity, CATME feedback, personality diversity, philosophy, language concerns, knowledgeable, motivation, partially interested, cooperation, supervision, considerate, personality diversity, self-role assignment

Team Formation: philosophy, criterion, schedule conflict, team size, gender, teaming quality, diversity, general, CATME, compensating

Team Strategy: sharing, lack of coordination, time management, freeloaders, overachievers, task assignment, role assignment, compensating, distribution of work

System Interface: CATME, CATME progress, design, logo, CATME login

Outlier Comments: *no sub-categories*

A detailed breakdown of major and sub-categories along with the number of comments within each category can be seen in Table 4 in the Appendix.

The analysis began by reading each comment within each major category. If the comments supported one of the ten claims the comment was assigned to the category or categories that it supports. After all comments were read the number of comments that support each claim were counted within individual sub-categories and then compared to the total number of comments in those sub-categories. Finally, based on the ratio of comments that related to each claim to the total number of comments within sub-categories in which supported comments were found, we concluded whether or not each claim, for which evidence was identified, was strongly supported, weakly supported, or not supported based on the verbatim comments we had. No evidence was found to dispute any of the claims identified by peer assessment researchers.

Findings

The number of supporting comments for each claim, along with the total number of comments in each subcategory, are shown in Table 4. The ratios of supporting comments vs. total comments in a category are shown in Table 1. Based on the data in Table 1 we determined whether or not each of the claims were strongly supported, weakly supported, or not supported by the verbatim comments collected from students using CATME. Our judgments about whether the evidence gathered from verbatim comments found in CATME supports the identified claims are shown in Table 2 below. Only two peer review and team interactions claims were strongly supported: Free Riders and Missing Skills in the CATME data that we examined.

Some of the strongest verbatim comments supporting the three claims for which we found supporting evidence are displayed below:

Claim	Number Counted	Number Possible	Ratio	Per-centage	Categories / Subcategories
Free Riders	72	609	0.118	11.8	Peer_Evaluation_Phil: rating, lack of contribution Team_Comment: compensating, collaboration, communication problem, lack of coordination, lack of communication, lack of interest, lack of motivation
Missing Skills	11	82	0.134	13.4	Team_Comment: skill diversity, combination, lack of knowledge, team combination
Resisting Evaluations	5	180	0.028	2.8	Peer_Evaluation_Phil: rating
Halo Error	10	340	0.029	2.9	Peer_Evaluation_Phil: meas-urement quality
Impact on Grades	5	180	0.028	2.8	Peer_Evaluation_Phil: rating
Social Pressure	1	180	0.006	0.6	Peer_Evaluation_Phil: rating
Rater/Ratee Reaction	0	0	0.000	0.0	None
Motivate Students	0	0	0.000	0.0	None
Practice Ratings	0	0	0.000	0.0	None
Peer Feedback Helpful	0	0	0.000	0.0	None

Table 1: Count of all claims and categories/subcategories in which they were found

Claim	Judgment
Free Riders	Strong Support
Missing Skills	Strong Support
Resisting Evaluations	Weak Support
Halo Error	Weak Support
Impact on Grades	Weak Support
Social Pressure	No Support
Rater/Ratee Reaction	No Support
Motivate Students	No Support
Practice Ratings	No Support
Peer Feedback Helpful	No Support

Table 2: Judgment of each claim

Free Riders: *"I wanted to say that this project was completely done by X and myself. X and the other guy (I don't know his name because [he] never came to class) made absolutely no contributions to this project."*
"I ended up doing the majority of the work on my own and end up having to get feedback from other groups because my group is unmotivated. X contributes nothing at all and all Y does is try to look up answers to problems on the internet."
"If this were a company, I'd fire two of the other 4 people on my team due to poor excuses for missing weekly meetings, unwilling to do any extra work outside of class time, lack of direction unless I tell them SPECIFICALLY and EXACTLY what to do (which defeats the point of having another brain in the operation), and lack to take any sort of control over something in the project."

Missing Skills: *"The only issue is that X does not really possess the same set of analytical skills that I feel the rest of us have developed as individuals. X relies heavily on the rest of us, and while X does make some contribution (organizing the powerpoint, organizing data in a spreadsheet, etc), X is rarely involved in the actual calculations or grinding analysis that allows us to generate results for our team."*
"From the Beginning I figure that our group would have some problems or were not capable of doing such a project, because of the skills many of our members [did not] have."
"Team members failed to demonstrate basic electronics skills or understanding of diagrams and could not begin project requirements much less complete them in a timely manner."

Resisting Evaluations: *"People will get bored with this survey and start putting anything. This survey is especially bad if it can change someone's grade at all. If someone doesn't know someone else, and they [give a] neutral instead of a 5, or SA or whatever, it will impact that person's grade."*
"This survey is bullshit. If I had complaints about my teammates, I would have gone to the teacher about it."

Halo Error: *"I am not taking this class to prevent my classmates from passing this course. Therefore, if asked at any point to rate my peers I will [give] anyone full credit regardless of how they performed. Just seeing the courage of my classmates to speak in front of the class and that they are committed to doing the work is enough for me."*
"I forgot who I met, so I just gave everyone a good grade."

Impact on Grades: *"I am not taking this class to prevent my classmates from passing this course. Therefore, if asked at any point to rate my peers I will [give] anyone full credit regardless of how they performed. Just seeing the courage of my classmates to speak in front of the class and that they are committed to doing the work is enough for me."*
"I think it is ridiculous that my individual grade is harmed by the poor evaluations of half of my group. The two members of my group, X and Y, always take our project and do it all by themselves, refusing to allow Z and I to participate. They do not answer our calls or tell us when we

are working and for the final project, did not even tell us when we were testing. I am angry enough that I did not get to participate in the project, but it is infuriating that I should be penalized."
"I would hate to lower their grades or have my own grade lowered based on an essentially superficial assessment."

Social Pressure: *"Following the revelation of Peer Review 2 excel spreadsheet file by our TA (in which we can see what [ratings] each of our team member gave to each other), I was left in a situation to give my group members good marks (both voluntary and involuntary), so that we don't get into further fights."*

Conclusions:

There are two claims that are commented on much more often than the others, Free Riders and Missing Skills. These are two major teaming issues that need to be addressed by instructors in teamwork oriented courses.

The first issue is Free Riders, or the presence of team members that don't contribute, or contribute significantly less than other team members. The second issue is Missing Skills, or the presence of team members who don't have the necessary skill or knowledge to provide a meaningful contribution to the team. Each issue should be focused on by instructors when arranging and managing student teams as they are issues that are cited as major teaming issues in student peer assessments.

Issues regarding team interactions evoke the strongest student comments as opposed to comments on the peer assessments themselves. We conclude that students are more likely to comment on the status of their team when there is strong dysfunctionality present in their team than they are concerned about the peer assessment system itself.

Recommendations:

Verbatim comments as found in student peer review systems are not a sufficient source of data for assessing the quality of peer interaction in a team. While many types of verbatim comments were provided by student participants in peer reviews the variety of types of comments and the focus of their comments are so varied as to not serve as a reliable source of data on team interactions. On the other hand when team situations are severely out of line-as in the presence of a Free Rider- verbatim comments could serve as a 'fire alarm' for the instructor.

Instructors need to take particular note of 'Free Riders' or students who are not fully or appropriately participating in the team activities in their courses. Multiple peer reviews during a

term can identify this ‘fire alarm’ phenomena early in the term and give instructors an opportunity to address the problem with the students[s] or with the team with sufficient time for the student or the team to recover and correct their performance relative to the expected course outcomes.

Next Steps:

Future studies of peer assessment could analyze the relationship between student’s and teams’ peer review ratings and the verbatim comments. If strong correlations are present then additional uses of verbatim comments to aid team management may be found.

Another possibility would be to observe student’s or a team’s performance in class and compare that to their peer assessment comments. Observations of this sort will give us clues on the behavior students are using to form the perceptions or team interactions that inform their comments and judgments.

A third type of future research would be to expand this type of study to practicing engineers’ verbatim comments on peer views of teammates. We presume the same issues that are seen within student engineers working in teams carry over to practicing engineers working in teams but confirmation of this hypothesis could lead supervisors of teams to use the same correction strategies as student instructors when examining peer review verbatim comments.

Claim	Quote
Free Riders	"Instructors often use peer evaluations to deter or remediate these problems, especially free riding, and to assign grades fairly based upon students' contributions."[10] Or above quotes
Motivate Students	"In addition to motivating students to contribute to their teams, using self- and peer evaluations shows students what is expected of them and how their team contributions will be evaluated"[10]
Social Acceptance	"Research shows that many raters, particularly average and below average performers, do not differentiate in their ratings of team members when it is warranted, sometimes because they worry that providing accurate ratings would damage social relations in the team"[10]
Missing Skills	"The "big five" model, therefore, assumes that team members will have the skills and motivation to contribute effectively to the team, yet these are frequently key deficiencies in student teams."[10]
Rater/Ratee Reactions	"In addition, instruments with descriptive anchors may generate more positive rater and ratee reactions, have more face validity, and offer advantages for raters from collectivist cultures"[10]
Like Teammates	"We expected that scores on the CATME-B would be positively associated with the degree to which teammates like the student and would want to work with the student again."[10]
Why do they like teammates	"We used two items from Jehn and Mannix (2001) and created a third item to measure the extentto which teammates like the student. Thesewere (1) I like this person as an individual; (2) Iconsider this person to be a friend; and (3) I enjoyspending time with this person"[10]
Practice Rating	"Repeated use of a peer-evaluation system increases students' confidence and skills in rating their peers...The practice-rating exercise should help students to improve their confidence and rating skill before they rate their actual teammates"[10]

Social Pressure	"Students in these studies did not appear to use the full range of the scale, resulting in a restriction of range problem with the data. Although this is a common problem in peer evaluation research for a variety of reasons, including social pressures to give high ratings"[10]
Halo Error	"The high correlations among the CATME-B dimensions, however, may also indicate the presence of halo error, which occurs when peers' perceptions of a teammate as a good or bad team member affect their ratings in specific areas. A metaanalysis found that correlations among different dimensions of job performance rated by peers are inflated by 63% due to halo error"[10]
Keeping the team on track	"The fact that rater effects are most substantial for the dimensions "Interacting with Teammates" and "Keeping the Team on Track" supports this interpretation because these dimensions are more idiosyncratic to specific teammates"[10]
Resisting and bias in peer evaluations	"Individuals whoare required to participate in peer-evaluation systemsoften resist the systems because they areconcerned that peer evaluations will be biased byfriendships, popularity, jealousy, or revenge. Recentresearch suggests that these concerns may bewell-founded"[10]
Peer feedback unhelpful	"...the team members agreed in their interviews that the peer feedback was unhelpful…"[1]
Impact and consideration of grades	"While these students agree with researchers on the characteristics of effective feedback, authority over grades seemed to matter more than the quality of feedback." AND "Students showed their desire for specific feedback, but they focused more on their grades than their learning."[1]

Table 3: Original claims list with quotes and citations.

Major Category	Sub-Category	# of Comments
CATME		**115**
	Specialization	8
	format	18
	comment box	2
	framing of choices	40
	repetitive	5
	survey format	2
	example	1
	project related questions	1
	schedule format	38
Faculty Commet		**539**
	role assignment	30
	survey timing	94
	grade assignment	1
	grade guideline	17
	criterion	358
	task assignment	20
	task definition	1
	task division	1
	supervision	17
Faculty Practice		**1160**
	team size	58
	role assignment	36
	CATME purpose	8
	CATME	142
	self-rating	2
	task assignment	41
	survey timing	127
	supervision	29
	lack of supervision	4
	team formation	2
	guidance	11
	grade guideline	21
	lack of guidance	1
	number of surveys	1
	criterion	512
	measurement quality	165
Peer Evaluation Phil		**1159**
	measurement quality	180
	rating	340
	self-rating	24
	scheduling conflict	154
	general	28
	CATME rating	4
	privacy	1
	lack of professionalism	1
	lack of interest	26
	lack of commitment	15
	lack of contribution	2
	phil	260
	CATME	123
	friendship	1

Table 4a: Full list of major categories and sub-categories with comment counts.

Major Category	Sub-Category	# of Comments
Team Comment		**1526**
	Collaboration	16
	formation	2
	general	29
	scheduling conflict	179
	communication problem	67
	commitment	69
	intellectual diversity	2
	leadership conflict	1
	conflict	1
	team composition	1
	team combination	3
	character	1
	team characteristics	1
	task assignment	43
	role assignment	39
	coordination	268
	combination	5
	skill diversity	65
	compensating	76
	team size	60
	lack of coordination	5
	lack of communication	1
	lack of commitment	15
	lack of interest	27
	lack of skills	1
	lack of motivation	13
	lack of team spirit	1
	lack of knowledge	8
	lack of consenus	1
	lack of project understanding	1
	sharing	175
	conflicting options	0
	maturity	1
	CATME feedback	1
	personality diversity	9
	phil	261
	language concerns	2
	knowledgeable	4
	motivation	28
	partially interested	1
	cooperation	5
	supervision	27
	considerate	1
	personality diversity	9
	self-role assignment	1
Team Formation		**1187**
	phil	275
	criterion	527
	schedule conflict	26
	team size	57
	gender	2
	teaming quality	1
	diversity	78
	general	28
	CATME	123
	compensating	70

Team Strategy		**383**
	sharing	178
	lack of coordination	47
	time management	5
	freeloaders	2
	overachievers	1
	task assignment	40
	role assignment	37
	compensating	71
	distribution of work	2
System Interface		**32**
	CATME	13
	CATME progress	8
	design	6
	logo	2
	CATME login	3
Outlier Comments		**2**

Table 4b: Full list of major categories and sub-categories with comment counts.

Citations:

1. Authors et al., (submitted). *International Journal of Problem Based Learning:* 19-25.
2. Brutus, S., Donia, M. B., & Ronen, S. (2013). Can Business Students Learn to Evaluate Better? Evidence From Repeated Exposure to a Peer-Evaluation System. *Academy of Management Learning & Education*, *12*(1), 18-31.
3. Chen, Y., & Lou, H. 2004. Students' perceptions of peer evaluation: An expectancy perspective. *Journal of Education for Business,* 79: 275–282.
4. Felder, R. M., & Brent, R. 2007. Cooperative learning. In P. A. Mabrouk, ed., *Active learning: Models from the analytical sciences,* ACS Symposium Series, 970: 34–53. Washington, DC: American Chemical Society.
5. Fellenz, M. R. 2006. Toward fairness in assessing student groupwork: A protocol for peer evaluation of individual contributions. *Journal of Management Education,* 30: 570–591.
6. Jehn, K. A., & Mannix, E. A. 2001. The dynamic nature of conflict: A longitudinal study of intragroup conflict and group performance. *Academy of Management Journal,* 44: 238–251.
7. Kaufman, D. B., Felder, R. M., & Fuller, H. 2000. Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education,* 89: 133–140.
8. MacDonald, H. A., & Sulsky, L. M. 2009. Rating formats and rater training redux: A context-specific approach for enhancing the effectiveness of performance management. *Canadian Journal of Behavioural Science,* 41: 227–240.

9. Michaelsen, L. K., Knight, A. B., & Fink, L. D. (Eds.). 2004. *Team based learning: A transformative use of small groups in college teaching.* Sterling, VA: Stylus Publishing.

10. Ohland, Matthew W., et al. "The Comprehensive Assessment of Team Member Effectiveness: Development of a Behaviorally Anchored Rating Scale for Self-and Peer Evaluation." Academy of Management Learning & Education 11.4 (2012): 610-625.

11. Saavedra, R. & Kwun, S. K. 1993. Peer evaluation in selfmanaging work groups. *Journal of Applied Psychology*, 78: 450–462.

12. Salas, E., Sims, D. E., & Burke, C. S. (2005). Is there a "big five" in teamwork?.*Small group research*, *36*(5), 555-599.

13. Sheppard, S., Chen, H. L., Schaeffer, E., Steinbeck, R., Neumann, H., & Ko, P. 2004. *Peer assessment of student collaborative processes in undergraduate engineering education.* Final Report to the National Science Foundation, Award Number 0206820, NSF Program 7431 CCLI-ASA.

14. Taggar, S., & Brown, T. C. 2006. Interpersonal affect and peer rating bias in teams. *Small Group Research*, 37: 87–111.

15. Viswesvaran, C., Schmidt, F. L., & Ones, D. S. 2005. Is there a general factor in ratings of job performance? A metaanalytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, 90: 108–131.

16. Wertz, Ruth, 2010, Unpublished research project to categorize 4,000 anonymous verbatim comments posted in the CATME system and received 2006-2010 during student peer reviews.