Journal of the Scholarship of Teaching and Learning, Vol. 13, No. 1, February 2013, pp. 45 – 56.

Cooperative learning and peer evaluation: The effect of free riders on team performance and the relationship between course performance and peer evaluation

Molly J. Dingel[1], Wei Wei[2], Aminul Huq[1]

Abstract: Cooperative learning has gained popularity in higher educational settings. However, assigning grades equitably to all team members in a way that rewards them for their contributions remains challenging. In this paper, we ask whether having free riders on a team lowers the quality of submitted work, and whether students' course performance correlates with peer evaluations. In an introductory sociology course, 101 students were organized into 20 teams; each team completed a set of three papers. Students were evaluated by the quality of work submitted and peer evaluations. For statistical analysis, we organized the teams into one of three groups, among which we compared average paper grades: group one contained teams with free riders, group two contained teams that may have free riders, and group three contained teams with no free riders. We found that teams with free riders did not submit significantly lower quality work than teams without free riders. We also statistically analyzed the relationship between student course performance and their peer evaluation, and found that students' performance in the course showed little correlation with their peer evaluations. Our results indicate that free riders do not depress the grades of non-free rider students. We also found that students' peer evaluations do not correlate with their course performance, a finding which warrants further research.

Keywords: cooperative learning, collaborative learning, project based learning, peer evaluation.

I. Introduction.

Many teachers and scholars emphasize the importance of collaborative learning. However, it is a challenge to assess team work in an effective and equitable way. In this article, we describe a unique, interdisciplinary project that required teams of students to write a set of three papers. We investigate how to best reward students for their contributions to collaborative assignments by exploring the usefulness of peer evaluations. As such, we ask whether having a "free rider," someone who does little or no work on team projects, lowers the overall quality of submitted project papers. We also explore whether students' peer evaluation correlate with their course performance.

[1] Center for Learning Innovation, University of Minnesota Rochester, 300 University Square, 111 S. Broadway, Rochester, MN 55904
[2] Mathematics Department, Metropolitan State University, Saint Paul, MN 55106

A. Collaborative learning.

A large body of research has demonstrated the benefits of collaborative learning strategies (Schroeder, Scott, Tolson, Huang, & Lee, 2007). These benefits include student reports that team work is beneficial, motivating (Bartle, Dook, & Mocerino, 2011), and confidence- and responsibility-building (Caulfield & Persell, 2006). In addition, evidence suggests that collaborative learning strategies, like discussing material in teams, can increase both students' test scores and their connection to classmates (Rau & Heyl, 1990). Some evidence indicates that collaborative learning may be beneficial specifically towards the aim of improving quantitative literacy (Caulfield & Persell, 2006).

Collaborative educational work can be conceptualized along a continuum from casual classroom use, to intentional and structured activities, to those designed to support "high-performance" teams that "engage in significant learning tasks" (Fink, 2002, p. 8). As collaborative activities move from "casual" to "team-based," instructors integrate the work into the course structure and devote increased attention to team formation and to accountability within the team. The most structured collaborative strategies are called "team-based learning" (Fink, 2002), or "cooperative learning" (Johnson, Johnson, & Holubec, 1986; Millis & Cottell, 1998). Millis and Cottell (1998) describe "cooperative learning" as establishing positive interdependence, which allows students to achieve mutual goals and rewards through structured tasks and interdependent roles.

B. Evaluating collaborative learning.

Some teachers argue that all group members should be assigned one grade. These teachers argue that in the workplace, teams are often rewarded for the quality of the project, and not for individual contributions to the project (Kagan, 1995). In addition, they argue that assigning one grade should motivate students to work together by creating "positive reward interdependence" (Johnson et al., 1986; Kagan, 1995). However, some evidence suggests that students easily become frustrated with group work when all group members earn the same grade regardless of effort (Conway & Kember, 1993). This frustration can cause students to lose motivation and become disengaged from class activities (Ruel, Bastiaans, & Nauta, 2003).

There are other good reasons to reject the arguments that all team members should receive the same grade (Kagan, 1995). Especially in more structured learning teams like those engaging in cooperative learning, peer evaluations[3] provide a way for students to be held accountable for their participation in, and contributions to, the teamwork. Given the importance of engagement from all group members, many scholars argue that peer evaluations are an essential part of successful cooperative learning strategies (Cheng & Warren, 2000; Conway & Kember, 1993).

As Kane and Lawler (1978) point out, there are different methods of peer assessment: peer nomination, where team members nominate the highest and lowest performing team member for a variety of particular characteristics; peer ranking, where each group member ranks all others from best to worst on a variety of factors; and peer rating, where each team member rates all other team members on a scale. In their review, they found that peer rating was most

[3] In this paper, we distinguish "peer evaluation" from "peer assessment," where peer assessment involves "grading" or providing feedback on a student's work or presentation, while peer evaluation involves rating the relative contribution and effort of peers working together in a team.

useful for collecting feedback about team members' specific behaviors. In this paper, we focus on peer ratings because of their usefulness in providing feedback about student participation in a group, and because peer ratings are commonly used in conjunction with collaborative learning projects. Throughout this paper, we use the term "peer ratings" synonymously with "peer evaluations."

Many peer evaluation strategies involve weighted peer evaluations (Cheng & Warren, 2000; Conway & Kember, 1993; Gatfield, 1999), which are calculated by dividing a student's individual effort rating with an average effort rating for the team. The resulting individual weighting factor controls for situations where all team members are evaluated particularly high or low. Other authors recommend single-item behaviorally anchored assessment (Kaufman, Felder, & Fuller, 2000; Oakley, Felder, Brent, & Elhajj, 2004), which has been shown to be as statistically reliable as a ten-item instrument (Ohland, Layton, Loughry, & Yuhasz, 2005).

While the research described above explores the question of how to integrate peer evaluations into students' final grades, to our knowledge, no research exists that explores whether the presence of a "free rider"—a student who fails to shoulder their part of the work load—decreases the quality of the team product. This question is pertinent because instructors use peer evaluations for the purpose of assigning grades to individuals in an equitable way. Peer evaluations help instructors adjust grades to ensure that students receive credit consistent with their contributions. In other words, peer evaluations help instructors assign higher grades to students who do more work on a project, compared to those who did very little on the project, and therefore foil the attempt of a student to reap unearned rewards (Roberts & McInnerney, 2007).

The presence of a free rider causes workload to shift onto the other group members, and this is a pertinent issue. However, even more concerning is that if the *product* of team work declines in teams with free riders, then, above shouldering additional work, students' grades may suffer as a result of other's (in)actions. If confirmed, this result may cause many instructors to question whether it is acceptable to ask students to engage in cooperative learning even if they use peer evaluations to modify individuals' grades. We believe that the issue of product quality is as important as the issue that some students put in more time and effort on group projects, because even if course assessment includes *no* team work, some individuals in classes will work harder than other students. Given that students are graded on the quality of their work, it is not usually of concern for instructors that some students put more time and effort into assignments with potentially higher learning outcomes (Huang, Yang, Kuo, & Yang, 2007; Webb, 1993). Assuming that within teams, students have similar ability to score full points on a particular assignment, some students working harder for more points within teams (via positive peer evaluations) should not generally raise concern among instructors. Indeed, many scholars recommend constituting teams with enough people, generally between three and six members (e.g., Oakley et al., 2004; Rau & Heyl, 1990) so that if one student is a "no-show," there remain enough team members to balance out the loss of effort from that team member. Further, evidence suggests that peer evaluations relieve students' frustration about shouldering more work because of free riders. In their study of 123 nursing students, Shiu and colleagues (2012) found that students had generally positive attitudes about peer evaluations and believed that peer evaluations reduced free riding among peers.

For instructors who use peer evaluations, another important question is whether the peer evaluations are valid and reliable. Though there are a number of studies that examine inter-rater reliability of peer evaluations in the work place (c.f. Conway & Huffcutt, 1997), there is very little research that explores correlations between a variety of variables and peer evaluation in the context of educational team work. Therefore, an open question is: what influences student evaluation of their peers? Persons (1998) found that students with high grade point averages (GPAs), higher participation grades, and higher homework grades received higher peer evaluation ratings than students with lower GPAs. This finding indicates that students are accurately identifying and evaluating students relative to their course performance, but little other research has been done in this area.

In this paper we explore group projects and peer evaluations through two questions. First, does the presence of free riders negatively affect the quality of the paper the team submits? In other words, are papers of lower quality if team members evaluate each other, or even one member, poorly? Second, do students who perform better in the course have higher peer evaluations? We explore this latter question by examining whether those who had low peer evaluations had low course grades.

II. Methods.

A. Project overview.

Students who participated in the project were enrolled in an introductory sociology class in the Center for Learning Innovation at the University of Minnesota Rochester. The vast majority of the students in this department are pursuing a Bachelor of Science in the Health Sciences. In their first year, students enroll in a common set of courses, including statistics, sociology, and a one-credit writing course. This structure allows faculty to work closely together to create integrated assignments that require students to draw on knowledge from multiple courses. Institutional Review Board (IRB) approval and student consent were obtained at the beginning of the fall 2011 semester. The study included 101 (of 106) students enrolled in one of two sections of an entry-level sociology course (one student declined to participate in the study, and their team was omitted from analysis).

In creating this set of assignments, we as instructors were driven by a pedagogical desire to create a learning environment where students had to work together to actively ask questions, test a hypothesis, and engage with course concepts. We organized students into teams of four to six, such that, when possible, all team members were in the same section of sociology, statistics, and writing. Beyond this criterion, teams were randomly constituted. A total of 20 teams were formed for the project.

Students were given a subset of census data with several variables. In their teams, they chose two variables about which to hypothesize a relationship. They then wrote a series of three papers: in the first, administered and graded in the sociology course, students articulated their hypothesis and statistically described each variable. In the second paper, administered through the statistics course but counting towards their grade in both sociology and statistics, students tested the hypothesis using skills learned in their statistics course. In the third paper, administered and graded in the sociology course, students contextualized their statistical findings with sociological concepts. For each paper, each student evaluated themselves and each of their teammates. Students also spent time in their writing course doing peer review of and revising these papers. In the sociology course, these cooperative data papers made up 20% of students' course grades. Individual assessments (in-class exams, essays, quizzes, and participation) made up the remaining 80% of their grades.

B. Grading.

Each team paper was assigned a “team grade” by the instructor. For peer evaluations, we used the process described by Oakley and colleagues (2004). Students were asked to rate themselves and each of the peers on their team using a behaviorally anchored, nine scale rating (Excellent = 100, Very good = 87.5, Satisfactory = 75, Ordinary = 62.5, Marginal = 50, Deficient = 37.5, Unsatisfactory = 25, Superficial = 12.5, No show = 0) that was converted into numbers. From these, we averaged the peer evaluations for each student, and calculated the team peer evaluation average. For each student, we calculated an “adjustment factor,” capped at 1.05, by dividing the individual average evaluation by the team average evaluation. Individual students’ grades were calculated by multiplying their adjustment factor by the team paper grade. This way of grading helps ensure that the free riders receive lower grades than the team paper grade.

C. Free riders and analysis methods.

To evaluate whether a free rider lowers the quality of the team project, we divided the 20 teams into three groups according to the standard deviation of peer evaluations of each team. We then performed a statistical analysis to compare the average paper grades among the three groups. Note that throughout, we use “teams” to refer to the student teams of 4 to 6; “groups” refers to sets of teams grouped by standard deviation of peer evaluations. We consider free riders to be students whose peer evaluations were significantly lower than their team members. Teams with free riders therefore had larger standard deviations of peer evaluations because some team members received high peer evaluations and others received low evaluations.

Based on the peer evaluation data, we observed that the standard deviation of the peer evaluations of a team was larger than or equal to 10 if at least one team member was rated lower than 70 and most other team members were rated much higher than 70, except that one team had a standard deviation of 9.3 (<10) but one member was rated at 68.8 for the last paper. The other observation was that the standard deviation of the peer evaluations of a team was smaller than 5 if all the team members were rated higher than 90.

Based on the above two observations, we divided the 20 student teams into three groups according to the standard deviation of peer evaluations: group 1 contains the teams whose peer evaluations’ standard deviation is larger than or equal to 10; group 2 contains the teams whose peer evaluations’ standard deviation is larger than or equal to 5 but less than 10; and group 3 contains the teams whose peer evaluations’ standard deviation is less than 5. In other words, teams in group 1, with large standard deviations of team member peer evaluations, had at least one free rider; teams in group 2, with moderate standard deviation of team member peer evaluations, may or may not have had a free rider; and teams in group 3, with small standard deviation had no free riders. Statistical analyses including the D’Agostino-Pearson normality test and Analysis of Variance (ANOVA) were performed to compare the average paper grades for the three groups.

D. Analysis of peer evaluation and course grade.

We calculated and recorded each student’s overall course performance, which included the evaluation of student work in homework, quizzes and exams before the due date of paper 1 and between the due date of paper 1 and paper 2. Prior to the due date of data paper 1, students took 11 quizzes, 1 exam, and wrote 1 essay. Between data papers 1 and 2, students took 3 quizzes, 1 exam, and wrote 1 essay. Between data papers 2 and 3, students only took one quiz. After data paper 3, students took two quizzes, and either wrote one additional essay or took a final exam.

Because there was minimal formal assessment between the due date of paper 2 and paper 3, we only analyzed the relationship between the overall course grades and peer evaluations for the first two papers. We performed the D'Agostino-Pearson normality test and calculated the Spearman correlation coefficient to test whether a correlation existed between the peer evaluations and student overall course grades. All analyses were conducted using the Statistical Analysis Systems (SAS 9.2) software (SAS Institute Inc., Cary, NC).

III. Results.

A. Free riders and team performance.

For each paper, all 20 student teams were assigned to 1 of 3 groups according to the standard deviation of the peer evaluations. The D'Agostino-Pearson normality test, on the basis of an examination of the skewness and kurtosis, showed that the paper grades for each group follow normal distributions. Therefore, the assumption of normality for performing a two-way analysis of variance (ANOVA) was satisfied. We then calculated the mean and standard deviation of paper grades for each group (table 1) and conducted a two-way ANOVA, including the factors of group and paper, to assess whether there were statistically significant differences of average paper grades among the three groups.

Table 1. Group average (mean) and standard deviation (SD) of paper grades.

Teams grouped by standard deviation of peer evaluation	Paper 1 (Mean ±SD)	Paper 2 (Mean ±SD)	Paper 3 (Mean ±SD)
Group 1 (at least one free rider)	84.50 ±7.84	84.55 ±8.57	89.60 ±5.03
Group 2 (may or may not have a free rider)	84.44 ±8.53	91.35 ±7.99	90.17 ±4.79
Group 3 (no free riders)	87.00 ±8.54	86.06 ±6.09	87.22 ±4.89

Table 1 shows no consistent pattern that any group performed significantly better across all the three papers. The results of the two-way ANOVA showed no interaction between the factors of paper and group (p=0.6074), no significant effects of the factor paper on average group grades (p=0.1422) and no significant effects of the factor group on average group grades (p=0.9165). In other words, at a significance level of 0.05, the average paper grade for any one group was not statistically significantly different from any other group and the average group grade for one paper was not statistically significantly different from any other papers. Therefore, a free rider does not statistically significantly affect the quality of the submitted paper (figure 1). Figure 1 shows that the average paper grades are very similar among the three groups.

B. Peer evaluation and course grade.

To test whether there is a strong correlation between the peer evaluations and the overall course performance for each student, we first conducted the D'Agostino-Pearson normality test to examine whether the course grades and the peer evaluations followed normal distributions. The results of the normality test showed that the peer evaluations of the whole class did not follow a normal distribution. Therefore, we calculated the nonparametric Spearman correlation coefficient with the two factors: course overall grade and peer evaluation (tables 2 and 3). We only conducted the analysis for paper 1 and paper 2 because there was minimal formal course assessment between papers 2 and 3.

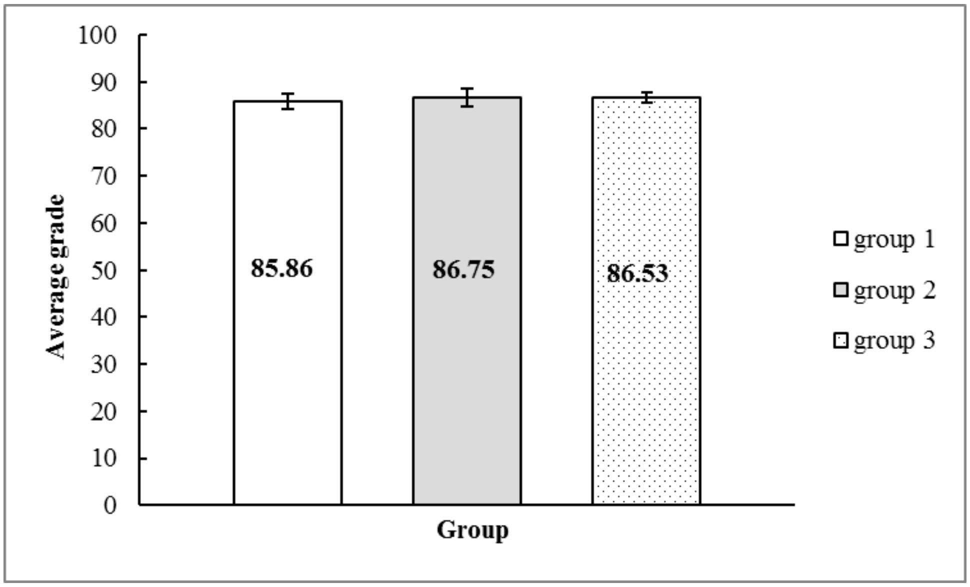

Figure 1. Average grade by group. The white bar represents the average grade of three papers for group 1, the grey bar represents the average grade of three papers for group 2 and the dotted bar represents that for group 3. The error bar represents one standard error or the average. The numbers on the bars represent the mean grade from each group.

Tables 2 and 3 show that the correlation coefficients between the overall course grade and peer evaluation are 0.4758 and 0.3626 for paper 1 and paper 2. Both of the correlation coefficients are very small (a coefficient of 0.7 is usually considered a moderate correlation), thus there is no strong correlation between the overall course performance and the peer evaluation.

Table 2. Spearman correlation coefficient between the course grade and peer evaluation for paper 1.

	Course grade	Peer evaluation
Course grade	1.0000	0.4758
Peer evaluation	0.4758	1.0000

Table 3. Spearman correlation coefficient between the course grade and peer evaluation for paper 2.

	Course grade	Peer evaluation
Course grade	1.0000	0.3626
Peer evaluation	0.3626	1.0000

Figures 2 and 3 represents scatter plots between peer evaluation and course overall grade for both paper 1 (figure 2) and paper 2 (figure 3) for every student. There are points showing very high peer evaluations but very low course grade, or very low peer evaluations but high course grade for both papers. Those points verified the very weak correlation between peer evaluation and overall course performance.

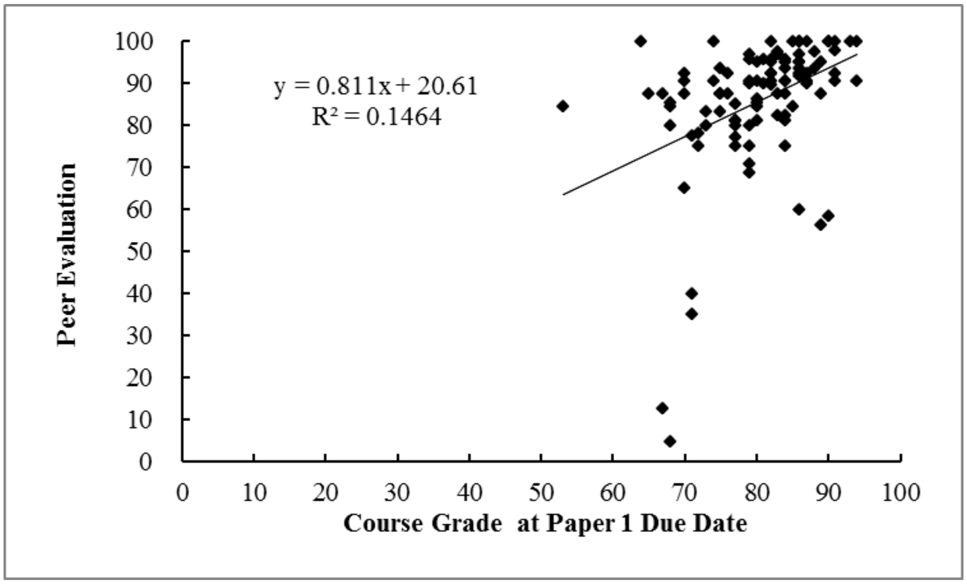

Figure 2: Average student peer evaluations for paper 1 versus course grade at paper 1 due date

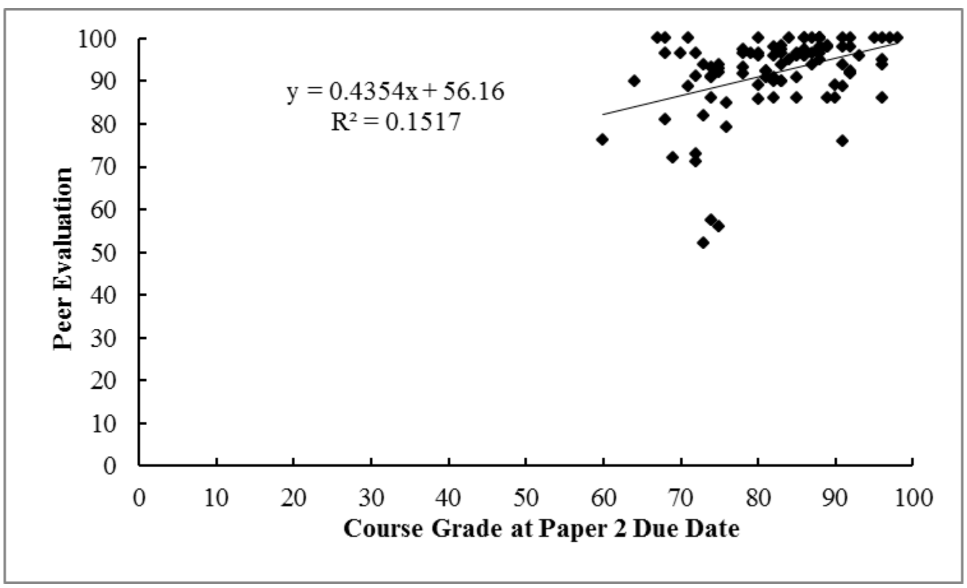

Figure 3: Average student peer evaluations for paper 2 versus course grade at paper 2 due date

IV. Conclusion.

Collaborative learning is a strategy with much empirical support (Schroeder et al., 2007), and is beneficial even in larger classes (c.f., Longmore, Dunn, & Jarboe, 1996; McKinney & Graham-Buxton, 1993). Collaborative learning includes a variety of activities spanning from instructors pairing up students and having them share with one another or complete simple cooperative exercises (Rao & DiCarlo, 2000) to creating "learning teams" that work together for an extended time to complete more complex learning activities. However, using learning teams for complex activities that make up a non-trivial part of a student's grade creates challenges for instructors seeking to equitably assign grades for these projects. Peer evaluations in the form of peer rating provide a good way to collect feedback about the involvement of individual students in team activities (Kane & Lawler, 1978). Grading schemas that do not account for individual variability may undermine student motivation and violate individual accountability (Kagan, 1995).

In this paper, we begin to explore whether individual students can be graded equitably when engaging in collaborative learning. We first investigate whether free riders decrease the quality of team papers. Based on the statistical analysis of project performance for the three groups (teams with at least one free rider, teams that may or may not have a free rider, and teams without free riders), we concluded that there are no significant differences among the performance of those three groups. Hence, free riders do not decrease the quality of project papers.

We then explore whether students' evaluation of their peers accurately mirrors those peers' performance in the course. Based on the correlation analysis between peer evaluations and student course performance, we found that students' evaluations of their peers do not accurately mirror the students' course grades. Students also provided brief qualitative feedback about their peers. Anecdotal evidence from this feedback suggests that students evaluated their peers based on *effort.* Common feedback about students with high peer evaluations include statements like: "Helped with the organization of the group and worked very well with everyone to get her section done" and "[Student] was always on time to meetings. She was active in group discussions and offered advice and opinions in an effective way." Common feedback about students with low peer evaluations include statements like: "Didn't get the section done when needed. Didn't seem very into the project. But helped edit the paper" and "Never showed up to any meetings beside final submission meeting nor contributed to group efforts outside meetings." These statements lend support to the claim that students evaluate their peers based on effort instead of skill, and presents a plausible reason why students' peer evaluations do not correlate with their grade in the course. These statements also support the claim that peer evaluations create equitable individual grades, because students seemed to evaluate their peers based on their efforts on the team project.

Our findings suggest that grades on collaborative activities can be equitably assigned. In that way, instructors should feel emboldened to use collaborative learning techniques, provided they use peer evaluations to adjust students' grades. Our findings also suggest that additional work needs to be done to explore what factors students use to evaluate their peers. These findings provide a sound basis to build upon and identify factors that influence peer evaluations and collaborative learning.

V. Discussion and Implications.

Team projects may inspire frustration in students (Ruel et al., 2003), who are often used to working individually and being awarded grades based on work that they alone produce (King & Behnke, 2005). Instructors (and students) may question whether free riders sabotage the final product of team-based classroom endeavors, therefore decreasing a students' grade, for which individual students cannot compensate. Evidence suggests that in the absence of peer evaluation, free-riding can increase the perception of inequality in team project grades (Ruel et al., 2003).

Our data support the use of team projects. Though engaged students in teams with free riders undoubtedly have to shoulder additional work, our analysis indicates that final paper grades do not suffer from the presence of free riders. This result should be encouraging to faculty who use or hope to use cooperative learning techniques. We therefore argue that peer evaluations protect the equity of individual grades by ensuring that free riders do not benefit from their lack of participation. Being engaged in classroom activities and "working harder" for a higher grade are generally not considered to be problems for individual projects, and, especially for teams of 3-6 students, we believe that a similar logic should also be applied to team projects. Future research should explore students' perception of free riders and whether peer evaluations quell students' concerns about free-riding peers (Shui et al., 2012).

If the skills and knowledge of students who are actively engaged increase more than free riders, then that would provide additional justification for the use of team projects and peer evaluation. In fact, some evidence indicates that learning outcomes are different for students who are actively engaged when compared to those who are not: Huang and colleagues (2007) found that, when compared to free riders, students who were actively involved in collaborative activities better assimilated knowledge and developed stronger leadership skills. In a similar study (Webb, 1993), middle-school students worked in groups to solve math problems, and then were subsequently individually tested on those problems. Webb found that the students who actively participated in the problem-solving groups earned higher individual test scores than those who had "loafed" during group work. Both studies indicate that students who actively participate in teams reap educational benefits, and therefore support the use of peer evaluations to adjust individual grades.

Given that peer evaluations are commonly used by instructors to reward individual students for good work within teams, in order to ensure the fairness of peer evaluation it is important to understand what factors students use to evaluate their teammates. To begin to explore this issue, we asked whether students' peer evaluations correlate with course grades. In our sample, students are not evaluating their peers consistently with their peers' grades in the course. This finding runs counter to other research (Persons, 1998), and could result from a variety of reasons: 1) students may be evaluating each other based on *effort*, which does not necessarily correlate with grade; 2) students are basing evaluation on other, as yet unknown, factors (King & Behnke, 2005). While we present anecdotal evidence that suggests peers evaluate one another on effort, there is little research that explores this issue. One study suggests that important factors that may influence peer evaluations are a student's age, race, and tendency to help others (Watson, BarNir, & Pavur, 2010). It is also possible that sociometry—someone's attraction, aversion, or neutrality towards another person—influences their evaluation (Kane & Lawler, 1978). This area is ripe for additional research and is an important area to continue to study.

References

Bartle, E.K., Dook, J., & Mocerino, M. (2011). Attitudes of tertiary students towards a group project in a science unit. *Chemistry Education Research and Practice, 12*, 313-311.

Caulfield, S.L., & Persell, C.H. (2006). Teaching social science reasoning and quantitative literacy: The role of collaborative groups. *Teaching Sociology, 34*, 39-53.

Cheng, W., & Warren, M. (2000). Making a difference: Using peers to assess individual students' contributions to a group project. *Teaching in Higher Education, 5*(2), 243-256.

Conway, J.M., & Huffcutt, A.I. (1997). Psychometric properties of multisource performance ratings: A meta-analysis of subordinate, supervisor, peer, and self-ratings. *Human Performance, 10*(4), 331-360.

Conway, R., & Kember, D. (1993). Peer assessment of an individual's contribution to a group project. *Assessment & Evaluation in Higher Education, 18*(1), 45-57.

Fink, L.D. (2002). Beyond small groups: Harnessing the extraordinary power of learning teams. In L. K. Michaelsen, A. B. Knight & L. D. Fink (Eds.), *Team-based learning: A transformative use of small groups* (pp. 3-25). Westport, CT: Praeger.

Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment & Evaluation in Higher Education, 24*(4), 365-377.

Huang, H., Yang, S., Kuo, Y., & Yang, S. (2007, September 3-7). *The study of "Free-rider" In problem-based learning situation*. Paper presented at the International Conference on Engineering Education, Coimbra, Portugal.

Johnson, D.W., Johnson, R.T., & Holubec, E.J. (1986). *Circles of learning: Cooperation in the classroom*. Edina, MN: Interaction Book Company.

Kagan, S. (1995). Group grades miss the mark. *Educational Leadership, 52*(8), 68-71.

Kane, J.S., & Lawler, E.E.I. (1978). Methods of peer assessment. *Psychological Bulletin, 85*(3), 555-586.

Kaufman, D.B., Felder, R.M., & Fuller, H. (2000). Accounting for individual effort in cooperative learning teams. *Journal of Engineering Education, 89*(2), 133-140.

King, P.E., & Behnke, R.R. (2005). Problems associated with evaluating student performance in groups. *College Teaching, 53*(2), 57-61.

Longmore, M.A., Dunn, D., & Jarboe, G.R. (1996). Learning by doing: Group projects in research methods classes. *Teaching Sociology, 24*(1), 84-91.

McKinney, K., & Graham-Buxton, M. (1993). The use of collaborative learning groups in the large class: Is it possible? *Teaching Sociology, 21*(4), 403-408.

Millis, B.J., & Cottell, P.G.J. (1998). *Cooperative learning for higher education faculty*. Phoenix, AZ: American Council on Education and The Oryx Press.

Oakley, B., Felder, R.M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning, 2*(1), 9-34.

Ohland, M.W., Layton, R.A., Loughry, M.L., & Yuhasz, A.G. (2005). Effects of behavioral anchors on peer evaluation reliability. *Journal of Engineering Education, 94*(3), 319-326.

Persons, O.S. (1998). Factors influencing students' peer evaluation in cooperative learning. *Journal of Education for Business, 73*(4), 225-229.

Rao, S.P., & DiCarlo, S.E. (2000). Peer instruction improves performance on quizzes. *Advances in Physiology Education, 24*(1), 51-55.

Rau, W., & Heyl, B.S. (1990). Humanizing the college classroom: Collaborative learning and social organization among students. *Teaching Sociology, 18*(2), 141-155.

Roberts, T.S., & McInnerney, J.M. (2007). Seven problems of online group learning (and their solutions). *Educational Technology & Society, 10*(4), 257-268.

Ruel, G., Bastiaans, N., & Nauta, A. (2003). Free riding and team performance in project education. *International Journal of Management Education, 3*(1), 26-38.

Schroeder, C., Scott, T.P., Tolson, H., Huang, T.-Y., & Lee, Y.-H. (2007). A meta-analysis of national research: Effects of teaching strategies on student achievement in science in the United States. *Journal of Research in Science Teaching, 44*(10), 1436-1460.

Shui, A., Chan, C., Lam, P., Lee, J., & Kwong, A. (2012). Baccalaureate nursing students' perceptions of peer assessment of individual contributions to a group project: A case study. *Nurse Education Today, 32*(3), 214-218.

Watson, W.E., BarNir, A., & Pavur, R. (2010). Elements influencing peer evaluation: An examination of individual characteristics, academic performance, and collaborative processes. *Journal of Applied Social Psychology, 40*(12), 2995-3019.

Webb, N.M. (1993). Collaborative group versus individual assessment in mathematics: Processes and outcomes. *Educational Assessment, 1*(2), 131-152.