

Effect of Peer Evaluation Format on Student Engagement in a Group Project

Jack T. Tessier¹
SUNY Delhi, Delhi, NY, 13753

Abstract

Active participation in classrooms often involves group work. In order to examine the effect of using peer evaluations as part of that experience, this study measured the influence of four formats of peer evaluation on students' perceptions of fairness of the peer evaluation method, its impact on peer engagement, and peer evaluation scores. The privately delivered peer evaluation format, where each student indicated what type of letter of recommendation they would write for each of their peers, was most effective in increasing perceptions of peer engagement and scores. These results suggest that a real-world style peer evaluation is most effective at promoting participation in groups, which should help students to become prepared for activities they will experience in the work world.

Keywords: Career preparation, group work, letter of recommendation, peer assessment, peer evaluation.

Results of education research encourage educators to employ active learning techniques in their classrooms, where students take a central role in the learning rather than simply being recipients of information (National Research Council, 1997). Active learning is effective because it promotes investigation, critical thinking, and collaboration (McNeal & D'Avanzo, 1997; National Research Council, 1997). Common forms of active learning include discussions, inquiry-based learning, problem-based learning, and presentations (Brown, Abell, Demir, & Schmidt, 2006; Downing, Kwong, Chan, Lam, & Downing, 2009; George et al., 1996; Grady, Gouldsborough, Sheader, & Speake, 2009; National Research Council, 1997; Park Rogers & Abell, 2008; Rettig & Smith, 2009; Schmidt, Cohen-Schotanus, & Arends, 2009).

These active techniques lend themselves to group work as opposed to individual work. Group work promotes a sense of community (Summers, Beretvass, Svinicki, & Gorin, 2005), improves communication and teamwork (Payne, Monk-Turner, Smith, & Sumter, 2006), and leads to higher levels of thought and learning than can be accomplished by individuals (Michaelsen, Bauman Knight, & Fink, 2004; Saleh, 2011). One of the challenges to group work, though, is minimizing the number of students who will ride on the coat-tails of the group, and not do much of the work.

¹ Corresponding author's email: tessiejt@delhi.edu

A regularly used method of minimizing this problem of social loafing is the integration of peer evaluation, whereby students in the group grade each other relative to their contributions to the group effort. Many teachers encourage peer evaluation (Elliot & Higgins, 2005; Li, 2001; Michaelsen et al., 2004) so long as it is couched in a collaborative classroom environment (Gueldenzoph & May, 2002). Students like being able to evaluate their peers (Chen & Lou, 2004; Gatfield, 1999), especially those students who have work experience (Gatfield, 1999) and know that when everyone contributes the group's product is much better than it would be if only some of the people contribute. Further, data show that peer evaluations are effective at reducing social loafing in group work (Brooks & Ammons, 2003; Chen & Lou, 2004).

Despite the utility of peer evaluations, few studies have examined the importance of the format of the peer evaluation (but see Lejk & Wyvill, 2001, 2002). The question addressed in this study was: what format of peer evaluation will be most effective at promoting student engagement and learning? I hypothesized that peer evaluation formats that are similar to real life evaluations will resonate with students and motivate them. I predicted that perceptions of fairness, group participation, and grades would all increase with peer evaluations based on real-world scenarios.

Methods

This research was approved by the SUNY Delhi Institutional Review Board and was conducted in an environmental issues course with approximately 35 students each semester. The students were primarily non-science majors. The course is certified for natural sciences general education credit by the State University of New York, and meets three times per week for lecture with no laboratory component. I established heterogeneous groups at the beginning of the semester. Specifically, students who had previously taken one of my classes served as group leaders. Their groups were then populated as evenly as possible by environmental studies majors, students who grew up in the area (a rural county in central New York State), students who grew up in New York City, and students who had spent noteworthy time outside the United States. Groups developed a collaborative atmosphere (Gueldenzoph & May, 2002) because they worked together in class throughout the semester on case studies, discussions, and learning quizzes (which mined their knowledge of environmental topics).

Each group was required to give a presentation on a national park of their choosing. The groups had to present on the location of the park, its ecosystems, the reason it was established as a park, threats to the sustainability of the park, and solutions to those threats. Each individual in the group provided a peer evaluation of each group member after the presentation.

I employed four formats of peer evaluation over successive semesters in the course (Table 1) from the spring 2009 semester to the fall 2010 semester. The peer evaluations were all holistic (a single, combined grade for effort related to all aspects of the project) as recommended by Lejk and Wyvill (2001, 2002). The points system (Table 1) was the most academic in nature and required students to grade each other from 0 to 100. The

Table 1. Comparison of formats for peer evaluations used in an environmental issues class for non-majors.

	Format of Peer Evaluation			
	Points	Letter of Recommendation	Letter and Change	Letter, Change, and Email
Description	Grade peer 0-100 Mean peer evaluation multiplied by group grade	Very good = +5 Good = 0 Negative = -10 Mean score added to/ subtracted from group grade	Very good = +5 Good = 0 Negative = -10 Mean score added to / subtracted from group grade Groups rearranged based on peer evaluations	Very good = +5 Good = 0 Negative = -10 Mean score added to / subtracted from group grade Groups rearranged based on peer evaluations
Potential Point Change	Up to 100 points lost	-10 to +5	-10 to +5	-10 to +5
Delivery of Peer Evaluation	Written and delivered in class	Written and delivered in class	Written and delivered in class	Emailed within 24 hours

shift to a realistic model of peer evaluation started with asking students to indicate what type of letter of recommendation they would write for each peer (Table 1): very good, good, or negative. This approach mimics what happens in the work world when a person tries to move to a new job, either immediately following school or as a shift in workplace or role. The second modification was to change the groups for the second part of the course based on the peer evaluations (Table 1): the students who earned the best letters of recommendation were put together in their new groups on down to those earning the least flattering letters having to work together. In the second part of the course, a different project was assigned, and it was advantageous to work with better people. This change is similar to what happens in the work world: employees who can garner the support of their peers get promoted and move on to better jobs, and those who are not highly regarded by peers tend to stagnate or lose their jobs. The final modification was to have the peer evaluations delivered to me via email rather than in class (Table 1). This condition is more consistent with the real world, because letters of recommendation are written away from the prying eyes of peers. In none of these approaches did the students know which new groups had students receiving the most very good letters of the most negative letters.

I collected data for this study using a voluntary survey and course grades. The survey was delivered at the end of the semester and asked students to indicate if they felt the format of the peer evaluation (1) was fair and (2) encouraged their peers to participate. All data were analyzed using Mintab Statistical Software version 15 (a computer program

that performs statistical tests on data that are entered; Minitab, Inc., State College, PA, USA) at $\alpha = 0.05$. Analysis of Variance was used to compare the four peer evaluation formats for perceived fairness, perceived encouragement of participation, group grades on the presentation, and the amount of grade change to individuals based on the peer evaluation.

Results

There was not a significant difference in the degree to which students felt the peer evaluation was fair among formats ($F = 0.78$, $p = 0.505$; Figure 1a). Approximately 80% of students felt the peer evaluation was fair regardless of format.

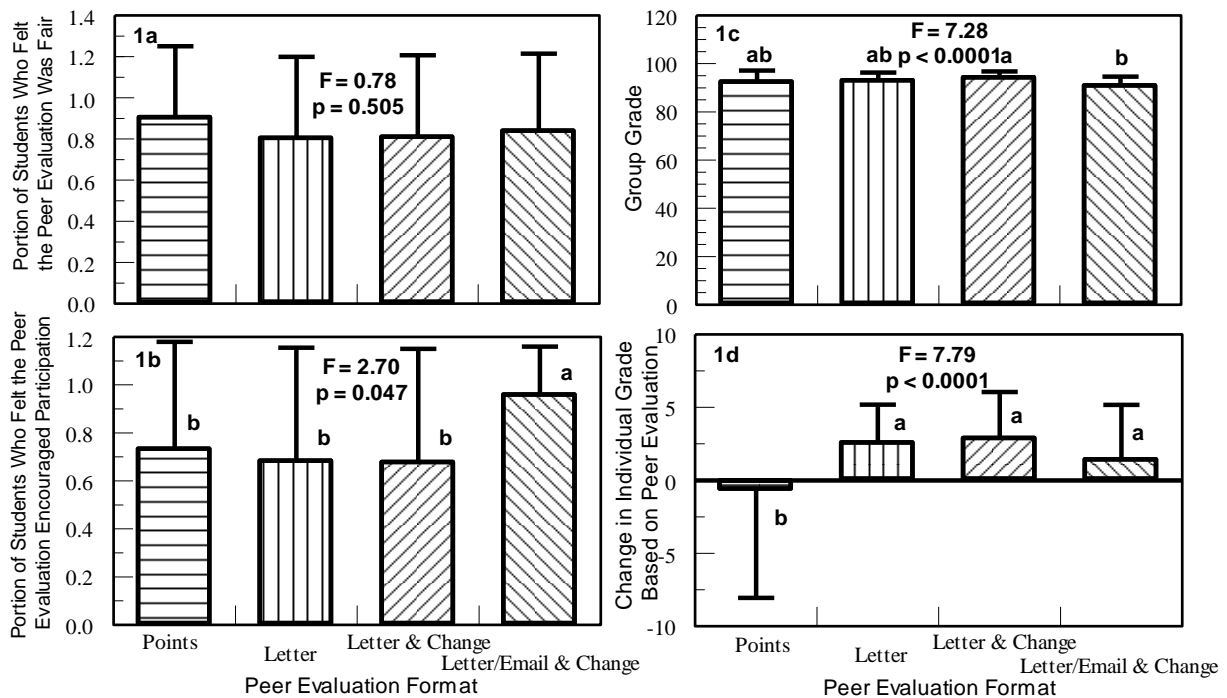


Figure 1. Comparison of a. perceptions of fairness, b. perceptions of encouragement for participation, c. group grades, and d. change in individual grade based on the peer evaluation for four peer evaluation formats in an environmental issues class. Means within a frame with different letters are significantly different at $\alpha = 0.05$. Error bars represent one standard error above the mean. F is the statistic generated by the Analysis of Variance (ANOVA) test. The higher the F statistic, the greater the differences among the treatments. The p-value is the probability of randomly seeing a result as extreme as the F statistic from an ANOVA. The α value is the probability of finding a significant difference among treatments when there is not one (Type I error).

Significantly more students felt that the peer evaluation promoted engagement by peers with the emailed letter of recommendation with change in group than in any other format ($F = 2.70, p = 0.047$; Figure 1b). In the other formats, approximately 70% of students felt that the peer evaluation encouraged engagement, but that number swelled to around 95% with the most real-life format of peer evaluation.

Group grades were significantly different between the letter of recommendation with group change format compared to the emailed letter of recommendation with group change format ($F = 7.28, p < 0.0001$; Figure 1c), but there was no significant difference among any of the other formats. The average letter grade from these formats was an A-, making the statistically significant difference a meaningless functional difference.

The change in individual grade based on the peer evaluation was different for the points format compared to the other formats ($F = 7.79, p < 0.0001$; Figure 1d), which were not significantly different from each other. The average student lost 1 point with the points format and typically gained 2 points with each of the letter of recommendation formats.

Discussion

The results of this research collectively suggest that the real-world format of the privately provided letter of recommendation encouraged student engagement in the group activity and led to positive interactions among peers. Regardless of format, students felt that the peer evaluations were fair (Figure 1a). Students tend to like peer evaluations (Chen & Lou, 2004; Gatfield, 1999), so it is not surprising that the format of the peer evaluation did not have a significant impact on students' perceptions of fairness of the peer evaluation. More than 80% of students viewed the peer evaluations as fair, which speaks to their importance in group activities regardless of format (Elliot & Higgins, 2005; Li, 2001; Michaelsen et al., 2004).

Students felt that the privately written letter of recommendation was the peer evaluation format that most encouraged their peers to contribute (Figure 1b). This encouragement may be the most important facet of the peer evaluation. We would hope, as teachers, that peer evaluations are not simply a means for contributing students to punish the group members who do not contribute (Saavedra & Kwun, 1993), but as a means of motivating students throughout the process to take an active part in their education and the group activity collectively. The capacity to deliver the recommendation in private seems to have made students feel that the peer evaluation would be truthful, and that they must contribute in order to earn a positive recommendation.

Data from the students' grades also indicate that the letter of recommendation formats generally encouraged participation by peers. While there was not an important effect of the format of the peer evaluation on student grades (Figure 1c), the letter of recommendation formats resulted in positive changes in students' grades (Figure 1d). Certainly, the points format permitted the lowest peer evaluation scores by far, but students experiencing the letter of recommendation formats had the opportunity to rate peers negatively yet consistently rated peers in the positive range. This repeated scoring of peers in the posi-

tive to very positive range indicates that groups were experiencing laudatory interactions with peers (Brooks & Ammons, 2003; Chen & Lou, 2004) that were encouraged by the format of the peer evaluation, which corroborates the students' views that the private letter of recommendation format encouraged peer participation (Figure 1b).

While I did not collect data regarding peer evaluations in other courses, I have employed the letter of recommendation with group change based on an email in other science courses that did include science majors. The behaviors of the students in those courses were not evidently different from those in the non-majors course, and I am confident that this format for peer evaluation can be used in any course and discipline. Regardless of discipline, students will seek employment in the work world and will be dependent on letters of recommendation from peers. Further, while students were not informed as to which group was made up of those receiving the worst letters of recommendation, they seemed to understand who they were. Anecdotally, these students seemed to realize the error of their ways and worked harder after the new groups were formed. Inevitably, there have been a few students who also performed poorly with their new group. The new groups filled with the highest rated students have always done very well after the new groups were formed.

Along with providing a well-rounded education for students, our efforts as teachers should help students prepare for their future in the work world (Danielson & Berntsson, 2007; Spowart, 2006). While some students do not like working in groups (Felder, Felder, & Dietz 2002; Gardner & Korth, 1998), most will be required to work in groups in their professional careers (Ezzamel & Willmott, 1998; Stevens & Champion, 1994). The model of a realistic format for peer evaluation supported by the data in this study serves the need of preparing students for their future in evaluating colleagues, encouraging participation, and enhancing the group work experience.

The results of this study lead to new questions related to improving the experience of students working in groups. First, does the original composition of the group affect peer evaluations? Students often rate their friends highly regardless of their contributions. Peer evaluations may be less meaningful in situations where students choose their own group members. Second, how would the results differ if students had to actually write the letter of recommendation rather than simply indicating what type of letter it would be? Students may participate less in the peer evaluation process if more work was required, but they may also take the task even more seriously if there was a higher level of input expected. Third, is the letter of recommendation the most real-world format in today's society? With the increasing use of Web 2.0 features, other venues may become the standard means of determining how a person is viewed by their peers.

In conclusion, the realistic format of peer evaluation that included private delivery of the type of letter of recommendation that a student would write for her/his peers was the most effective format in encouraging participation in group work and led to the highest peer evaluation scores. Real-world experiences, such as these, should help to prepare students for their experiences in the work world following their formal education.

Acknowledgments

I thank the students of BIOL 110 Environmental Issues and Sustainability at SUNY Delhi for their participation in the study, the SUNY Delhi Dean's Council for the Professional Development Grant that supported the dissemination of the research, and Peter Daempfle for constructive comments on the manuscript.

References

- Brooks, C. M., & Ammons, J.L. (2003). Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assessments. *Journal of Education for Business*, 78 (5), 268–272.
- Brown, P. L., Abell, S. K., Demir, A., & Schmidt, F. J. (2006). College science teachers' views of classroom inquiry. *Science Education*, 90 (5), 784–802.
- Chen, Y., & Lou, H. (2004). Students' perceptions of peer evaluation: an expectancy perspective. *Journal of Education for Business*, 79 (5), 275–282.
- Danielson, E., & Berntsson, L. (2007). Registered nurses' perceptions of educational preparation for professional work and development in their profession. *Nurse Education Today*, 27 (8), 900–908.
- Downing, K., Kwong, T., Chan, S.-W., Lam, T.-F., & Downing, W.-K. (2009). Problem-based learning and the development of metacognition. *Higher Education*, 57 (5), 609–621.
- Elliott, N., & Higgins, A. (2005). Self and peer assessment – does it make a difference to student group work? *Nurse Education in Practice*, 5 (1), 40–48.
- Ezzamel, M., & Willmott, H. (1998). Accounting for teamwork: a critical study of group-based systems of organizational control. *Administrative Science Quarterly*, 43 (2), 358–396.
- Felder, R. M., Felder, G. N., & Dietz, E. J. (2002). The effects of personality type on engineering student performance and attitudes. *Journal of Engineering Education*, 91 (1), 3–17.
- Gardner, B. S., & Korth, S. J. (1998). A framework for learning to work in teams. *Journal of Education for Business*, 74 (1), 28–33.
- Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment and Evaluation in Higher Education*, 24 (4), 365–377.
- George, M. D., Bragg, S., de los Santos Jr., A. G., Denton, D. D., Gerber, P., Lindquist, ... & Meyers, C. (1996). *Shaping the Future: New Expectations for Undergraduate Education in Science, Mathematics, Engineering, and Technology*, Washington, DC: National Science Foundation.
- Grady, R., Gouldsborough, I., Sheader, E., & Speake, T. (2009). Using innovative Group-work activities to enhance the problem-based learning experience for dental students. *European Journal of Dental Education*, 13 (4), 190–198.
- Gueldenzoph, L. E., & May, G. L. (2002). Collaborative peer evaluation: best practices for group member assessments. *Business Communication Quarterly*, 65 (1), 9–20.
- Lejk, M., & Wyvill, M. (2001). Peer assessment of contributions to a group project: a comparison of holistic and category-based approaches. *Assessment and Evaluation in Higher Education*, 26 (1), 61–72.

- Lejk, M., & Wyvill, M. (2002). Peer assessment of contributions to a group project: student attitudes to holistic and category-based approaches. *Assessment and Evaluation in Higher Education*, 27 (6), 569–577.
- Li, L. K. Y. (2001). Some refinements on peer assessment of group projects. *Assessment and Evaluation in Higher Education*, 26 (1), 5–18.
- McNeal, A. P., & D’Avanzo, C. (1997). *Student-Active Science: Models of Innovation in College Science Teaching*, New York, NY: Saunders College Publishing.
- Michaelsen, L. K., Bauman Knight, A., & Fink, L. D. (2004). *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*, Sterling, VA: Stylus Publishing.
- National Research Council. (1997). *Science Teaching Reconsidered: A Handbook*, Washington DC: National Academy Press.
- Park Rogers, M. A., & Abell, S. K. (2008). The design, enactment, and experience of inquiry-based instruction in undergraduate science education: a case study. *Science Education*, 92 (4), 591–607.
- Payne, B. K., Monk-Turner, E., Smith, D., & Sumter, M. (2006). Improving group work: voices of students. *Education*, 126 (3), 441–448.
- Rettig, J. E., & Smith, G. R. (2009). Class research projects in ecology courses: methods to “un-can” the experience. *Journal of College Science Teaching*, 38 (5), 38–42.
- Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, 78 (3), 450–462.
- Saleh, T. A. (2011). Statistical analysis of cooperative strategy compared with individualistic strategy: an application study. *Journal of Effective Teaching*, 11 (1), 19–27.
- Schmidt, H. G., Cohen-Schotanus, J., & Arends, L. R. (2009). Impact of problem-based, active learning in graduation rates for 10 generations of Dutch medical students. *Medical Education*, 43 (3), 211–218.
- Spowart, J. (2006). Hotel school students’ views of their preparation for work-integrated learning: an exploratory study. *Asia-Pacific Journal of Cooperative Education*, 7 (2), 10–15.
- Stevens, M. J., & Campion, M. A. (1994). The knowledge, skill, and ability requirements for teamwork: importance for human resource management. *Journal of Management*, 20 (2), 503–530.
- Summers, J. J., Beretvas, S. N., Svinicki, M. D., & Gorin, J. S. (2005). Evaluating collaborative learning and community. *The Journal of Experimental Education*, 73 (3), 165–188.