A COMPARISON OF PEER EVALUATION SYSTEMS IN TEAM-BASED LEARNING

Thomas L. Ruble, Rider University
Sigfredo A. Hernandez, Rider University
William J. Amadio, Rider University

Zero-sum peer evaluation systems have been advocated to counter problems of "free-riding" in students' group work. In this paper, we suggest that the zero-sum peer evaluation system may encourage behavior that runs counter to many objectives of team-based learning. We propose that an alternative peer evaluation system, non-zero-sum with caps is preferable to the zero-sum system in facilitating group process and minimizing rating biases. Participants in this study were 168 students working on group projects. As part of the course requirements, students completed a non-zero-sum peer evaluation with upper and lower boundaries. Following the required peer evaluation, the students completed a "hypothetical" peer evaluation using a zero-sum system with an attitude survey comparing the two peer evaluation systems. The attitude survey indicated a very slight preference for the zero-sum system. However, the zero-sum system produced greater self-enhancement and some gender biases. A peer evaluation system needs to consider the broader impact of such evaluation on group process rather than merely focusing on countering the "free-rider" problem.

INTRODUCTION

Individual accountability is a key problem affecting the productivity of learning teams and group projects (Michaelsen 1994, Slavin 1995, Millis and Cottell 1998, Johnson and Johnson 1999). To counter problems of "free-riding" in students' group work, Grieb and Pharr (2001) proposed a grade system based on a pure zero-sum peer evaluation scheme. Their game theory argument assumes that individual team members try to maximize their own GPA by managing their limited time an effort among various classes and projects. Given this assumption of GPA maximizing students, Grieb and Pharr (2001) further argued that pure zero-sum peer evaluation scheme is superior to a maximum limit (nonzero-sum) scheme in obtaining individual accountability in group projects.

In this paper, we suggest that the zero-sum peer evaluation system may encourage individual behavior that runs counter to many objectives of team-based learning. The position espoused by this paper is grounded in the group dynamics and cooperative learning literatures. A team member's effort is a function of an environment that facilitates/hinders the development of cohesive and productive groups. Highly productive teams tend to be highly cohesive (Robbins 1989, Mullen and Copper 1994, Johnson and Johnson 2000). Group cohesion is desirable because it forms strong bonds between team members and thereby influences them to work

harder to achieve team goals (Oyster 2000). Team members derive valuable rewards, in addition to their grades or GPA, from participating in a cohesive team. (Michaelsen 1994, Slavin 1995, Johnson and Johnson 2000, Oyster 2000). Thus, an important attribute of a peer evaluation scheme is that, in addition to promoting individual accountability of group members, it helps to develop group cohesiveness. We propose that an alternative peer evaluation scheme, nonzero-sum with ceiling and floor caps is desirable to the pure zero-sum scheme in contributing to group process and the development of team cohesiveness.

METHOD

Participants in this study were 168 students working on group projects as part of a course requirement. Included in the study were 55 MBA students enrolled in a night program and 113 undergraduates enrolled in both evening and day courses. The classes covered management, marketing, and computer sciences. The students formed 37 different teams ranging in size from three to six. The gender mix was about equal with 86 males and 82 females.

All the students participated in a peer evaluation that was part of the group project grading system. The system was a nonzero-sum evaluation with an upper boundary equal to 100% of group project score and a lower boundary equal to 70% of the group project score. Students rated themselves as well as the other members of their group. Immediately following the actual peer evaluation, the students were asked to complete a hypothetical peer evaluation using a zero-sum system with 100% serving as the individual base line. Following the allocation of points based on the zero-sum system, students completed a seven item attitude survey comparing the two peer evaluation systems. The two peer evaluation forms are included in the Appendix.

RESULTS

The study analyzed student ratings and attitudes toward both peer evaluation schemes. The two systems produced the same ratings in 44% of the cases. That is, nearly half of the students gave every group member the same 100% of the group project grade. Still, the two peer evaluation systems produced some differences that could impact a student's grade.

The results of the peer evaluations for oneself and others are reported in Table 1. Under the nonzero-sum system, ratings given to oneself were slightly higher than ratings given to others in the group ($M=97.9$ versus $M=96.5$, $t=2.24$, $p=.027$). In comparison, under the zero-sum system, self ratings were noticeably higher than ratings of others ($M=102.8$ versus $M=99.2$, $t=3.48$, $p<.001$). In terms of points, the discrepancy between grades given to oneself versus grades given to others increased from 1.4 in the nonzero-sum system to 3.6 in the zero-sum system. A similar bias was present in comparing one's self ratings with ratings received from other group members. For the nonzero-sum system, self ratings exceeded ratings received from other members by less than one point ($M=97.9$ versus $M=97.2$, $t=2.21$, $p=.028$). In comparison, under the zero-sum system, the rating by self versus rating by others discrepancy increased to nearly three points ($M=102.8$ versus $M=100.0$, $t=3.99$, $p<.001$). Clearly, the zero-sum system yielded stronger self-serving grading biases.

Insert Table 1

In answering the survey following the two peer evaluation systems, students were not aware of the possible biases reported above. Without this data, the attitude survey indicated a very slight preference for the zero-sum system. Results of the post-evaluation survey are reported in Table 2. Students felt the zero-sum system provided a more accurate assessment of relative contributions, was better for rewarding outstanding contributions, and was better for withholding points for below average contributions. On the other hand, students felt that the zero-sum scheme was easier to abuse (enrich oneself at the expense of others) than the nonzero-sum scheme. Overall, students preferred the zero-sum system by a small margin over the nonzero-sum system.

Insert Table 2

Despite the apparent preference for the zero-sum system, other possible problems surfaced with further analysis of the rating data. Table 3 presents results of a comparison of ratings by gender. For the nonzero-sum system, there were no statistical differences in ratings of teammates based on gender differences of the raters. However, for the zero-sum system, there were differences in ratings of male and female targets based on the gender of the rater. Specifically, under the zero-sum system, males gave male teammates higher ratings than females gave male teammates ($M=101.0$ versus $M=98.4$, $t=2.38$, $p=.019$). Further, females gave female teammates higher ratings than males gave female teammates ($M=100.6$ versus $M=98.2$, $t=2.51$, $p=.013$). Thus, under the zero-sum system there was a bias favoring same gender teammates. This bias favoring same genders in peer evaluations was not present under the nonzero-sum system with upper and lower caps.

Insert Table 3

IMPLICATIONS FOR THE DEVELOPMENT OF TEAMS

In advocating a zero-sum peer evaluation system coupled with a formal team charter, Grieb and Phar (2001) suggest that such a system would improve the identification of free-rider behavior, provide incentives for non free-riders and help establish the expectation that free-riders would be held accountable for their behavior. However, focusing on simply managing free-rider behavior misses the broader implications of peer evaluation. A peer evaluation method, as a desirable incentive system for group projects and cooperative learning, needs to be consistent with an environment that facilitates the development of productive and cohesive teams.

Our data shows that students were more likely to show self-enhancement in their peer evaluations. Moreover, students recognized that the zero-sum scheme is easier to abuse than the nonzero-sum scheme with caps. Because of the greater capacity of a team member to enrich oneself at the expense of others, the pure zero-sum scheme seems more likely to promote negative interdependence or competition among team members. Instead, a desirable peer evaluation scheme should encourage positive interdependence among team members. Positive interdependence is a necessary condition for the development of cohesive and productive teams (Michaelsen 1994, Millis and Cottell 1998, Johnson and Johnson 1999, Johnson and Johnson 2000).

The larger variation of scores and the potential for individual enrichment that results from the zero-sum scheme are not likely to facilitate group processing. Group processing refers to the group discussion about what members' actions are helpful and unhelpful in accomplishing team goals (Johnson and Johnson 1999). Group processing helps teams become more productive and cohesive (Johnson and Johnson 1999, Johnson and Johnson 2000). A desirable peer evaluation scheme should not hinder group processing. The larger variation of scores and the potential for individual enrichment typical of the pure zero-sum scheme are more likely to elicit greater feelings of inequity among the lower scoring members. In such a situation, group processing is less likely to occur naturally in the team and if required by the instructor, communications are more likely to be less open and more emotionally charged, and group conflict is more likely to result.

Students indicated a slight preference for the zero-sum peer evaluation system. However, this preference was expressed before knowing the scores contained self-enhancement and gender biases. If students were aware of the biases associated with the zero-sum system, it is quite possible that their attitudes toward the two systems would change. As faculty members concerned with developing a positive environment for team performance, we are still looking for an alternative to the pure zero-sum system of peer evaluation

REFERENCES

Grieb, T. and S. W. Pharr. " Managing Free-Rider Behavior in Teams," Journal of the Academy of Business Education, Fall 2001, 37-47.

Johnson, D.W. and R.T. Johnson. Learning Together and Alone: Cooperative, Competitive, and Individualistic Learning, (1999), Boston: Allyn and Bacon.

Johnson, D.W. and F.P. Johnson. Joining Together: Group Theory and Group Skills, (2000), Boston: Allyn and Bacon.

Michaelson, L.K. "Building Learning Teams: The Key to Harnessing the Power of Small Groups in Higher Education," in Collaborative Learning: A Sourcebook for Higher Education, K.W. Prichard and R.M. Sawyer (Eds.), (1994), Westport, CT: Greenwood.

Millis, B. J. and P.G. Cottell. Cooperative Learning for Higher Education Faculty, (1998), Phoenix, AZ: Oryx.

Mullen, B. and C. Cooper. "The Relationship Between Group Cohesiveness and Performance: An Integration," (1994), Psychological Bulletin, 115, 210-227.

Oyster, C. K. Groups: A User's Guide, (2000), Boston: McGraw Hill.

Robbins, S.P. Organizational Behavior: Concepts, Controversies, and Applications, (1989), Englewood Cliffs, NJ: Prentice Hall.

Slavin, R.E. Cooperative Learning, (1995), Boston: Allyn and Bacon.

TABLE 1

SELF AND OTHERS RATINGS

	ZERO-SUM	NONZERO SUM
SUBJECTS' SELF RATINGS	102.8	97.9
SUBJECTS' RATINGS OF OTHERS	99.2	96.5
RATINGS OF SUBJECT BY OTHERS	100.0	97.2

TABLE 2

RATINGS OF ZERO-SUM(ZS) COMPARED TO NONZERO SUM

ITEM	MEAN RATING
1. ZS is Easier	2.55
2. ZS allows More Accurate evaluations	2.20
3. ZS provides better chance to Reward	1.96
4. ZS provides better chance to Punish	2.14
5. ZS is easier to Abuse	2.36
6. ZS provides a Fairer evaluation	2.51
7. ZS is Preferred over alternative	2.77

Note: 1 = Strong Agreement
3 = Neutral
5 = Strong Disagreement
n = 166

TABLE 3

GENDER DIFFERENCES IN RATINGS

	ZERO-SUM	NONZERO SUM
MALE SUBJECTS' RATINGS OF MALES	101.0	97.0
FEMALE SUBJECTS' RATINGS OF MALES	98.4	96.7
MALE SUBJECTS' RATINGS OF FEMALES	98.2	97.5
FEMALE SUBJECTS' RATINGS OF FEMALES	100.6	97.8

APPENDIX

REQUIRED COURSE PEER EVALUATION

Name _____ Team _____

Rate the contributions of your teammates, including yourself, to the team project.
Base your ratings on the norms, ground rules, and goals established by your team for the project.
Use the following scale:

Excellent, above average contribution . . . . = 10
Good, average contribution . . . . . = 9
Below average contribution . . . . . = 8
Deficient contribution . . . . . = 7

NAME RATING COMMENTS

1. Myself

2.

3.

4.

5.

6.

PEER EVALUATION -- ALTERNATE FORM

This form represents another approach to evaluating the team members. In this case, you have either 400, 500, or 600 fixed total percentage points to assign (depending on whether the team has 4, 5, or 6 members). If everyone contributed equally, you would assign 100 percent of the team grade to each member. However, if some members contributed more than others, you could allocate greater than 100 percent to some members while reducing other allocations below 100 percent as long as the total equaled the appropriate base points. Then, we multiply the team grade by the percentage allocations for each member to arrive at a final individual grade.

Team # ____ Number in team (circle) 4 5 or 6 Total points per team _____ POINTS

Myself (name) _____ _____
A. (name) _____ _____
B. (name) _____ _____
C. (name) _____ _____
D. (name) _____ _____
E. (name) _____ _____

Please compare this Alternative peer evaluation system with the Original. Indicate your agreement or disagreement with the following statements: (circle your choice)

	Strongly Agree	Moderately Agree	Neutral	Moderately Disagree	Strongly Disagree
1. This approach is easier.	1	2	3	4	5
2. This approach provides a more accurate assessment of the relative contributions.	1	2	3	4	5
3. This approach is better for rewarding outstanding contributions.	1	2	3	4	5
4. This approach is better for withholding points for below average contributions.	1	2	3	4	5
5. This approach is easier to abuse (enrich oneself at the expense of others).	1	2	3	4	5
6. Overall, this approach is a fairer system.	1	2	3	4	5
7. I prefer this approach to the original.	1	2	3	4	5