Journal of Management Education
Volume 32 Number 5
October 2008 580-612

10.1177/1052562907310641
http://jme.sagepub.com
hosted at
http://online.sagepub.com

An Expectancy Theory Motivation Approach to Peer Assessment

Barry A. Friedman
Pamela L. Cox
Larry E. Maher
State University of New York at Oswego

Group projects are an important component of higher education, and the use of peer assessment of students' individual contributions to group projects has increased. The researchers employed an expectancy theory approach and an experimental design in a field setting to investigate conditions that influence students' motivation to rate their peers' contributions to team projects. Two questionnaires were also developed and tested by the researchers. This research found that rating format and rating frequency significantly interacted to influence student motivation and their perceptions of their team. Research findings reveal peer assessment to be a complex process in need of further study. Two peer assessment instruments and peer assessment training materials are provided along with suggestions for future research.

Keywords: *peer assessment; teamwork; expectancy theory; motivation; group projects; evaluation; group dynamics*

Group projects have become an important component of higher education, and the use of peer assessment of students' individual contributions to final group projects has increased (Brooks & Ammons, 2003; Falchikov & Goldfinch, 2000; Gueldenzoph & May, 2002; Hanrahan & Issacs, 2001; Pond & Ul-Haq, 1997; Topping, 1998). Instructors often cannot directly observe the contributions students make in groups, and it is therefore useful to use peer observations to supplement instructor observations

Authors' Note: We wish to thank the editor, John Stark, and the reviewers for their helpful comments and suggestions. Supplemental peer assessment materials can be obtained from the authors by contacting Pamela L. Cox, School of Business, 327 Rich Hall, State University of New York at Oswego, Oswego, NY 13126; *phone*: (315) 312-2658; *fax*: (315) 312-5440; *e-mail*: pcox@oswego.edu.

(Chen & Lu, 2004). Proponents of peer assessment advocate its use to improve student learning (Ballantyne, Hughes, & Mylonas, 2002) and its positive effect on student achievement and attitudes (Topping, 1998).

For peer assessments to have a positive impact, the peer assessment process should be effective and provide meaningful ratings. By effective, we mean that the assessment process leads to outcomes desired by the students and the instructor. By *meaningful ratings*, we mean that the student has taken the rating process seriously and spent time considering each rating. Effective and meaningful peer assessment requires that students have the ability and motivation to rate other students, yet few studies investigate students' motivation to rate peers. Because of end-of-course time constraints, instructors often use peer ratings to determine grades without providing feedback to students or holding students accountable for the quality of the ratings they provide. Therefore, students may have little motivation to provide meaningful ratings. The purpose of this research is to investigate the impact of specific peer rating practices on student motivation to rate their peers and to understand student perceptions of the peer rating process.

Literature Review

Incorporating teamwork in the classroom is often problematic because of the uneven effort, participation, and contribution of individual team members (Gueldenzoph & May, 2002). Many students have negative perceptions of group projects because of experience with dysfunctional groups (Fiechtner & Davis, 1985) and free riding or social loafing problems (Brooks & Ammons, 2003; Mello, 1993; Strong & Anderson, 1990; Williams, Beard, & Rymer, 1991). One solution may be to incorporate some type of peer evaluation of individual effort to team projects. Peers work closely together and may therefore have a greater number of accurate behavioral observations (DeNisi, Randolph, & Blencoe, 1982; Greguras, Robie, & Born, 2001). Evaluation of individual contributions to group work can positively influence a group's ability to work well together (Druskat & Wolff, 1999) and help to ensure that team members contribute their fair share of the work (Erez, Lepine, & Elms, 2002; Karou & Williams, 1993).

Students' Motivation to Rate Peers

Expectancy theory provides a theoretical context to explore students' motivation to rate peers. Originally conceptualized by Vroom (1964) and

Porter and Lawler (1968), expectancy theory attempts to explain individuals' motivation to make choices, expend effort, and complete a task. According to expectancy theory, motivation consists of three components: (a) individual expectancy that effort will lead to performance (effort-to-performance, or E-P, expectancy), (b) the belief that performance will lead to certain outcomes (instrumentality), and (c) the value (valence) placed on these outcomes (Rynes & Gerhart, 2000). An individual feels motivated when he or she feels that effort will lead to an acceptable level of performance, performance will lead to some outcome(s), and the outcome(s) are personally valued (Isaac, Wilfred, & Douglas, 2001).

Although used extensively in the workplace, expectancy theory has recently been applied to academic environments (Isaac et al., 2001). Chen and Hoshower (2003) used expectancy theory to explore the validity of students' evaluations of their teachers. Chen and Lu (2004) used expectancy theory to assess key factors that may motivate students to participate in the peer evaluation process. They outlined outcomes attractive to students such as determining peers' grades, reducing uneven work distribution, enhancing group productivity, and improving peers' performance. Multiple regression analysis examined the force model of expectancy theory. The force model explained students' motivation for participation (R^2 = .72). However, the attractiveness of the peer evaluation system (outcome attractiveness) was more important than the likelihood that the students' efforts would lead to a successful contribution. In other words, students are willing to exert effort in evaluating their peers if they find the outcomes of the peer evaluation attractive to them. Chen and Lu contend that "if students are kept ignorant of the use of peer evaluation, if peer evaluations are used for purposes that students do not value, or if they see no visible results from their participatory efforts, they will cease to give meaningful input" (p. 281).

The methodology employed by Chen and Lu (2004) and elsewhere (Campbell, 2003) uses a within-subject decision-making methodology suggested by Stahl and Harrell (1981). This methodology uses fictitious scenarios to manipulate E-P expectancy level, performance-to-outcome (P-O) instrumentality, and valence. In a typical within-subjects study, participants respond to between 20 and 30 fictitious scenarios. Within each subject, the manipulated expectancy theory components are then regressed on the participants' reported level of attractiveness and effort they would exert across the fictitious situations. An average R^2 statistic is then calculated across subjects. However, the sample size for each subject is only 20 to 30, depending on how many scenarios are presented. As a result, this small sample size may take advantage of spurious multiple correlations related to small sample size

(Aiken & West, 1991). Additionally, student motivation to rate peers has not yet been tested in a field experiment using actual peer ratings.

Peer Assessment Instruments

Researchers have explored two types of assessment instruments to determine individual weighting factors: holistic and categorical. In holistic assessment, each student awards only one score to each of his or her team members, reflecting an overall impression of contributions to the group effort (Lejk & Wyvill, 2001). In categorical peer assessment, students score each other on a number of categories, criteria, or dimensions. The scores are combined to arrive at a percentage contribution for each group member (Lejk & Wyvill, 2001). With both of these methods, some sort of weighting factor is derived that can then be multiplied with the group project grade to determine each individual's grade.

Studies using categorical approaches have employed a wide variety of criteria. Gueldenzoph and May (2002) recommend that the specific criteria match the objectives of the course or group project. There is some evidence that specific criteria related to the goals of the course may increase assessment accuracy (Levi & Cadiz, 1998).

Studies comparing holistic versus categorical peer assessments of individual contributions to a group project have yielded mixed results. In support of categorical approaches, it was found that the use of specific performance criteria may improve team member performance (Erez & Somech, 1996; Harkins, 1987). Brooks and Ammons (2003) found that students perceived that free rider problems declined when evaluations provided specific feedback.

However, Lejk and Wyvill (2001) found that students preferred holistic over categorical instrument approaches. A holistic assessment was used to calculate individual grades, and a category-based assessment was for research purposes only. Lejk and Wyvill (2001) compared the extent of agreement among peers using the two methods. The holistic method resulted in higher interrater agreement among peers and less change in students' individual grades when the holistic weight was applied to the group project grade. The holistic assessment was also more effective at dealing with outstanding and weak group contributors. The authors concluded that the holistic assessment enabled the students to reflect an overall impression and thus resulted in many groups giving each other equal grades. They contend that the categorical assessment was less effective in reflecting contributions to the group effort because students tend to assign different aspects of the group projects to each other and therefore will not have the opportunity to demonstrate all the criteria assessed. The method also implies that all categories are

equally important. Students may, in fact, avoid tasks they dislike. Lejk and Wyvill (2001) contend that holistic peer evaluation supports the goals of summative assessment (i.e., grading) and that categorical peer assessment may be useful for formative feedback (i.e., developmental feedback).

Lejk and Wyvill (2001) did not use an experimental design whereby comparisons between holistic and categorical peer assessment conditions were possible. Knowing that the holistic assessment determined final grades, students' may have been reluctant to penalize other students and thus assigned equal ratings to their peers. In fact, a number of studies cite the unwillingness to differentiate among team members for fear of hurting interpersonal relationships as one of the problems with peer evaluation (Greguras et al., 2001; Liden & Mitchell, 1983; Mohrman & Lawler, 1983; Naiper & Latham, 1986).

Rating Frequency

The frequency and timing of peer assessment administration may also be critical. Students typically provide peer ratings at the end of the semester only. Bacon, Stewart, and Silver (1999) contend that end-of-semester peer assessment may actually encourage undesirable behaviors. Students may tolerate undesirable behaviors thinking that they can punish the poor performers at the end of the semester. Brooks and Ammons (2003) found support for peer assessment that provides feedback at both early and multiple points during the semester in reducing social loafing problems and increasing students' satisfaction with group work. Completing peer assessments multiple times has the added benefit of giving students practice with the technique. Frequent feedback gives poor performers a chance to improve (Feichtner & Davis, 1992). Research has not yet investigated the influence of frequency or timing of peer assessment on student motivation to rate.

Student Perceptions of Peer Assessment

Student have mixed reactions to peer assessment, ranging from positive (e.g., fair and provides formative information) to negative (e.g., it is socially embarrassing and cognitively challenging) (Topping, 1998). Students have expressed skepticism about the value of peer comments (Orsmond, Merry, & Reiling, 1996) and are not always convinced that their peers will grade them fairly (McDowell, 1995). Student leniency may be the result of a desire not to cause friction and preserve friendships (Cheng & Warren, 1997; Greguras et al., 2001). Kelmar (1993) found the instructor's average marks to be significantly lower than peer marks. Other studies similarly found that students

Table 1
Research Experimental Conditions

		Feedback Specificity	
		Holistic	Categorical
Peer feedback frequency	Single	Students provide an overall peer rating at end of semester. $N = 45$ (27.1%)	Students provide five dimensional ratings once at the end of the semester. $N = 37$ (22.3%)
	Multiple	Students provide an overall rating three times during the semester. $N = 40$ (24.1%)	Students provide five dimensional ratings three times during the semester. $N = 44$ (26.5%)

were reluctant to award low marks to peers even when they were deserved (Brindley & Scoffield, 1998; Cheng & Warren, 1997; Falchikov, 1995).

In this article, we examine the impact of the following on student motivation to provide peer ratings and student satisfaction with peer assessment and group work: frequency of peer assessment and feedback (three times versus once at the end of the semester) and peer assessment instrument (holistic versus categorical). See Table 1 for a description of the experimental conditions.

Hypotheses

The research reported here uses expectancy theory and student perceptions of the rating process as dependent variables. The two independent variables are rating format (holistic or categorical) and rating frequency (single at end of course or three times during course). Specific hypotheses are offered below.

Hypothesis 1: Increased frequency of peer feedback will increase student motivation to provide peer ratings.

Hypothesis 2: Increased frequency of peer feedback will increase student satisfaction with their teams and the peer rating process.

Rather than team members waiting until the end of the semester to "punish" poor performers, multiple peer assessment and feedback encourages early communication of desirable and undesirable behaviors. This communication affords lagging team members the opportunity to improve (Feichtner & Davis,

1992). Conducting peer evaluation earlier in the semester is especially critical because it may have an impact on the early stages of team development when teams are deciding how to conduct work and team member roles (Druskat & Wolff, 1999). Peer evaluation at early and multiple points during the semester may reduce social loafing problems and increase student satisfaction with teamwork (Brooks & Ammons, 2003). Repeated practice and familiarity with the rating process afforded to students in the frequent feedback experimental condition will increase students' confidence in their ability to rate (E-P expectancy) and their perception that performance will lead to certain outcomes (instrumentality). Increases in student confidence to rate their peers and improved team functioning (for example, less social loafing) may lead to increased motivation to provide peer ratings, increased team satisfaction, and increased satisfaction with the peer rating process.

Hypothesis 3: Categorical rating formats will increase student motivation to provide peer ratings.

Categorical peer assessment instruments with specific criteria may provide students with clearer expectations as to performance standards and expected behavior (Brooks & Ammons, 2003; Van Velsor & Leslie, 1991; Young & Henquinet, 2000). Categorical rating formats provide raters with more specific information than do holistic rating formats and may increase students' perceived competence to rate peers (E-P expectancy). Categorical ratings provide students with greater depth of behavior feedback, serving to increase students' perceived valence associated with the ratings. Specific categories may be easier to assess than global overall ratings that require raters to consider a multitude of factors to determine a holistic rating. The use of specific performance criteria may be associated with improved team member performance (Erez & Somech, 1996; Harkins, 1987) and decreased social loafing (Brooks & Ammons, 2003). Lejk and Wyvill (2001) contend that the holistic assessment may enable students to reflect an overall impression. On the other hand, category-based peer assessment may be more useful for formative (developmental) feedback than holistic assessment. Categorical assessment was also found to be more effective at fine tuning small differences in individual contributions to team projects (Lejk & Wyvill, 2001, p. 69).

In proposing Hypothesis 3, it is our contention that students who receive specific (categorical) feedback multiple times during the semester will use that feedback in a formative or developmental manner, improving on individual weaknesses and capitalizing on strengths. The greater performance feedback specificity will have a greater impact on student expectancies when

it leads to improved individual performance and improved team functioning. As students perceive team functioning to improve, their motivation to provide peer ratings will also increase, as will their satisfaction with their teams and the peer rating process.

Hypothesis 4: Students who use categorical rating formats frequently will experience higher motivation to rate peers and experience greater satisfaction with the peer rating process.

Hypothesis 4 states that the categorical rating format and multiple rating frequency will interact to increase motivation. Previous research has not employed an experimental design whereby the interaction of rating format and frequency can be assessed. It is our contention that the provision of greater depth and clarity of information afforded by the categorical rating format, combined with the greater opportunity to affect change using multiple ratings, will result in higher motivation and satisfaction with the rating process relative to the other experimental conditions.

Method

The sample consisted of 166 students enrolled in an introductory business course offered at a northeastern state university. The course introduces students to the fundamentals of business, emphasizing business communication and critical thinking skills. Fifty-five percent of the students were male and ninety-seven percent were White. Fifty-four (32.7%) students reported that they had never rated their peers in past courses, and 104 (63%) reported that they had only rated their peers once or twice in past courses. All students enrolled in the course were first-semester freshmen. Therefore, any experience with rating peers most likely occurred during high school.

Students completed a skill assessment instrument, and teams of four or five students with diverse skill sets were formed using the results (see Cox & Bobrowski, 2004, for a more detailed account of the instrument and team formation process). Instructors assigned a project to each team, where it chose a company and conducted a detailed company analysis during the course of the semester. The project was divided into three parts. Students completed the first part of the project 2 months into the semester, the second part 3 months into the semester, and the third part the last week in the semester. Instructors graded each part of the project and provided students with feedback within 2 weeks of the submission. The written team project constituted 30% of students' final grade with a final team presentation of the project constituting 10% of the final grade (Cox & Bobrowski, 2004).

Based on an expectancy theory instrument used by Nadler and Lawler (1977), the researchers developed the Peer Assessment Expectancy Questionnaire, or PAEQ (Appendix A, Exhibit 1). A focus group consisting of 38 business school undergraduate students discussed their experiences with peer assessments and discussed expectancy theory as it relates to peer assessment. The themes that emerged from the focus group not only influenced the design of our research and the development of the PAEQ but also helped us to define what students consider an effective and meaningful peer assessment process (see Friedman & Lechner, 2005, for a more detailed description of the pilot study and development of the PAEQ).

In designing the study, we made every effort to ensure that students took the ratings seriously and took time to carefully rate their peers. Students were shown ahead of time exactly how the ratings would affect their grade. We also sought to provide anonymity by having the students complete the ratings outside of class, returning them to the instructor in an unmarked, sealed envelope. Feedback was provided in aggregate form so that individuals could not be associated with specific ratings. It was difficult to prevent students from being overly lenient or harsh. However, we felt that the average rating would balance out lenient or harsh individual ratings. We were also unable to prevent students from engaging in "gamesmanship" as they were aware of the peer assessment method ahead of time.

An additional outcome was added to the PAEQ even though it did not emerge from the focus groups: "developing your own rating skills." This outcome was based on our review of the literature. The peer assessment literature suggests that by providing students with multiple opportunities to rate their peers, they become more adept with using the peer assessment instrument (Brooks & Ammons, 2003; Feichtner & Davis, 1992).

The researchers tested Vroom's (1964) original two models in this study: the valence and force models. The valence model represents students' perceived attractiveness of rating their peers, and the force model represents students' motivational force to rate their peers. Question 1 of the PAEQ contains seven P-O items (i.e., the degree that students feel that rating peers result in developing better skills). The seven items are second-level outcomes (Vroom, 1964) linked to the first order outcome of doing a good job of rating peers. Question 2 of the PAEQ contains seven items that measure the importance that the students assigned to each of the second order outcomes. The valence model is calculated by multiplying students' perceived probability that the second order outcomes (Question 1) will result from doing an especially good job rating one's peers (i.e., P-O expectancy) multiplied by the importance that students attach to each of the same outcomes. As adapted from Campbell (2003),

$V_j = \Sigma\ (Imp_k\ I_{jk})$
$k = 1$
where
V_j = the valence of providing peer ratings (first level outcome),
Imp_k = the importance of the second- level outcomes
I_{jk} = the perceived performance-to-outcome probability that the first level outcome, providing peer ratings (V_j) will lead to second level outcomes (V_k), and
n = the number of potential second level outcomes. (p. 4)

We operationalized the force model by multiplying valence (V_j) as defined above with the expected probability that increased effort will lead to meaningful peer ratings. Question 3 of the PAEQ contains three items that measure students' belief that effort will lead to performance (e.g., "working hard to rate my peers will result in ratings that accurately reflect each peer's contribution"). Campbell (2003) states:

$F_i = (E_{ij}\ V_j)$
Where:
F_i = the motivational force to perform act i,
E_{ij} = the expectancy that act i will result in outcome j, and
V_j = *the valence of outcome j* [that is, the valence of providing peer ratings (first-level outcome)]. (p. 4)

Using the Statistical Package for the Social Sciences to analyze the responses, the PAEQ demonstrated excellent internal consistency, with an average Cronbach's alpha of .83 ($p \leq .001$). Table 2 contains reliability statistics for the PAEQ subscales.

The researchers developed and administered a final questionnaire at the end of the semester: Student Peer Feedback Questionnaire (SPFQ). The questionnaire was administered after the final peer ratings were completed by the students (on a different day but during the same week) to minimize any interaction effects. The SPFQ measured students' satisfaction with their groups and the rating process and the level of effort they expected to exert to complete the peer ratings (see Table 3). The results section describes this questionnaire and its factor structure.

Experimental Conditions

Feedback specificity (holistic and categorical) and feedback frequency (one and three feedback iterations) served as the two experimental conditions

Table 2
Peer Assessment Expectancy Questionnaire Internal Consistency[a]

	Cronbach's Alpha		
Scale	2004 Pilot (N = 57)	2005 Pretest (N = 166)	2005 Posttest (N = 166)
Performance to outcome	.85	.82	.86
Valence	.86	.77	.85
Effort to performance	.82	.74	.79

a. All Cronbach alphas listed are significant ($p \geq .001$).

in the study, resulting in a 2 × 2 experimental design (see Appendix A, Exhibits 2 and 3 for the peer assessment instruments). The introductory business course consisted of 12 sections of approximately 19 students each. Seven different instructors taught the course. To minimize any potential instructor effect, we assigned three different instructors to each experimental condition. All students received a standardized presentation at the beginning of the semester (and prior to any group work taking place) that outlined the nature of the research and peer feedback, with emphasis on the importance of feedback for group and individual effectiveness, and the impact that peer ratings have on grades. The only difference between the presentations was that students in the categorical condition were shown examples using the categorical format and students in the holistic condition were shown examples using the holistic format. We provide the holistic format presentation (see Appendix A, Exhibit 3) with the categorical format presentation, available from the authors on request. In both presentations, we explained in detail how peer ratings affect student grades prior to completing the peer evaluations. Several examples were provided so that students could see the impact of both good and poor peer evaluations on their grade. Students were provided with the formula that would be used to determine average peer ratings and shown how averaged peer ratings would be used to calculate individual grades (see Appendix A, Exhibit 3).

The potential impact of the peer assessments on course grades was held constant across the different experimental conditions. The only differences between the conditions were whether average ratings were calculated categorically or holistically and how many times students rated their peers. The curriculum of the course was standardized across sections, as was the determination of final grades. Standardized grading rubrics also helped to ensure

Table 3
Student Peer FeedNback Questionnaire Three-Factor Solution Using Principle Components Analysis Extraction With Varimax Rotation

		Factors[a]		
		Team Functioning	Effort and Assessment Fairness	Team Concerns
1	Members on my team worked well together.	.85		
2	My team functioned efficiently.	.85		
3	By the end of the project, everyone on my team did his or her fair share of the work.	.81		
4	I would do another project with my team if given the opportunity.	.77		
5	I am proud of the project my team produced.	.74		
6	Not everyone on my team did his or her fair share of the work.	–.71		
7	Teamwork improved on my team as the semester progressed.	.69		
8	I enjoyed working on my team.	.67		
9	I think it is important to take part in a team project as part of my studies.		.65	
10	I tried my best to contribute to the team project.		.63	
11	Team projects are a good way to learn.		.57	
12	I put a great deal of thought into my assessment of the members of my team.		.57	
13	Peer assessment methods are a fair way of distributing marks in a team project assessment.		.56	
14	Automatically granting each team member the same grade for a team project is unfair.		.54	
15	Before undertaking a team assignment, students should learn how to work as a team.		.51	
16	The feedback I received will help me contribute more to the team.		.50	
17	Team assignments tend to penalize able students.			.80
18	I would rather undertake an individual assignment than a team assignment.			.73
19	Team assessments methods are not as reliable as individual assessments.			.69
20	Team assignments allow students to “hide” more easily than individual assignments.			.65

NOTE: Students responded to each item using a 5-point Likert-type scale, where a = *strongly agree*, b = *agree*, c = *not sure/neutral*, d = *disagree*, and e = *strongly disagree*.
a. For purposes of clarity, only highest loadings are displayed.

consistency across sections (see Cox & Bobrowski, 2004, for a more detailed discussion).

One of the concerns that emerged from our focus group discussions was that many students did not believe that professors actually use peer ratings in their determination of final grades. Thus, trust in the instructor to use peer ratings effectively is an important prerequisite for students taking the process seriously. Our standardized presentation to students and examples of the impact peer ratings would have on their grade (Appendix A, Exhibit 3) were intended to establish trust by providing students prior to rating their peers with the algorithms that would be used to calculate their grades. We also provided students with answers to questions they might have regarding the peer assessment process (see Appendix A, Exhibit 3).

On the same day as the presentation (Week 2 of the semester), the PAEQ was administered to students. The initial PAEQ data served as the pretest measurement of student expectancies. Several weeks later (Week 6), following completion of the first part of the project, students in the holistic and categorical multiple conditions were given a peer assessment instrument to complete on their own with instructions to return the completed peer assessment instrument to their instructors in a provided envelope. Students in the multiple conditions also completed the peer assessment instrument after the second part of the project was submitted (Week 10) and at the end of the semester after the third part of the project was submitted (Week 14). Students in the holistic and categorical single conditions completed the peer assessment instrument once at the end of the semester (Week 14) following submission of the third part of the project.

In the "holistic/single" and "holistic/multiple" experimental conditions, students rated their peers using a holistic instrument based on the instrument designed by Lejk and Wyvill (2001). This instrument asks students to divide $100 among the members of their team (including themselves), reflecting the contributions of each team member to the team project. The average individual team member allocation multiplied by the team grade determined the students' individual project grades. Following submission of their ratings, students received feedback indicating the average rating from their peers and the impact of the rating on their final project grade.

Feedback specificity was greater in the "categorical/single" and "categorical/multiple" experimental conditions. Students provided their peers with feedback on five categories: attendance at team meetings, completion of fair share of work, participation at team meetings, meeting deadlines, and encouraging and supporting other team members. The researchers instructed students to allocate $100 among themselves and their peers for each category, according to

the contributions of each team member to the team project. Rating formats and student instructions appear as Exhibit 2 (categorical ratings) and Exhibit 3 (holistic ratings) in Appendix A. We delivered feedback in the single and multiple categorical conditions in the same manner as described earlier for the holistic conditions.

With respect to both the holistic and categorical formats, behavioral anchors were not given. However, with both formats, students were given the same behavioral examples for both conditions (e.g. attend meetings) during the training session. However, only students using the categorical format gave separate ratings for each category. We improved on the Lejk and Wyvill (2001) design by using the same rating method across all four experimental conditions (dividing $100 among team members). Lejk and Wyvill had students divide 100% among team members in the holistic conditions and used a 5-point scale for each category in the categorical condition.

The researchers trained course instructors to administer the PAEQ and SPFQ and instructed students on how to provide peer ratings. Instructors were instructed on how to use the peer ratings to calculate final grades and were provided with an Excel program to automate the process as much as possible (available from the authors on request). The researchers provided students with the following peer feedback: group project grade, average peer rating, self-rating, and their fair share of the group project score based on their peer ratings. Students were assured that their names would be omitted from any publications that emerged from the research, with their data used in aggregate form. Students were not aware of the specific hypotheses the research was designed to study.

At the end of the semester (14th week), students completed two questionnaires. The PAEQ was administered a second time to ascertain changes in students' motivation to rate their peers as a function of the experimental conditions. It was administered after the final peer assessments were completed (the same week but on a different day). In addition, the SPFQ that measured students' perceptions and satisfaction with their teams and the rating process was administered (see Table 3). The SPFQ was administered on the same day as the PAEQ. Although the administration of the SPFQ might have resulted in an interaction effect with the PAEQ, course logistics made it impractical to administrate it on a different day. Of the 208 students enrolled in the course, 166 completed peer assessments, both administrations of the PAEQ, and the SPFQ. Thirty (14%) students did not provide ratings. Some students had dropped the course, while others failed to either complete their peer assessments or were absent from class on the days the PAEQ and the SPFQ were administered. We ran a chi-square analysis to see if the missing data were

related to the experimental conditions. Although the multiple administration, categorical rating condition had the highest missing data (9.6%), there was no significant relationship between the experimental conditions and the missing data ($\chi^2 = .73$).

Results

Questionnaire Reliability and Factor Structure

The PAEQ demonstrated high reliability, with average Cronbach alphas of .78 ($p \leq .001$) and .83 ($p \leq .001$) for the first and second administrations, respectively (see Table 2). To test the internal construct validity of the SPFQ, the researchers' factor analyzed the item responses using a principle components extraction method with a varimax rotation (Rummel, 1970). A three-factor solution generated the clearest factor solution: team functioning, effort and assessment fairness, and team concerns (see Table 3). The three factors accounted for 57.32% of the total variance of the original questionnaire responses. Table 4 contains descriptive statistics for valence and force motivation scores and SPFQ factor scores for each of the four experimental conditions. Valence and force motivation scores were calculated using Campbell's (2003) formulation reviewed earlier and rescaled to the original 7-point Likert-type scales. Across all experimental conditions, the second administration of the PAEQ resulted in lower motivation scores than the first PAEQ administration ($t = 7.37$, $p \leq .001$, and $t = 6.81$, $p \leq .001$ for valence and force motivation scores, respectively). The difference between the first and second PAEQ administrations was not significantly different among the four experimental conditions.

Valence and Force Motivation Anova

Table 5 reports the results of the ANOVA analysis of the valence and force motivation (PAEQ) dependent variables across the rating format (holistic or categorical) and rating frequency (single or multiple) independent variables as well as average performance to outcome, importance, and E-P ANOVA results. In support of the ANOVA randomization assumption, there were no significant differences between the experimental conditions with respect to students' initial valence or force motivation. Instructors administered the PAEQ a second time at the end of the course. The first PAEQ scores served as covariates to control for initial student motivation in

the second PAEQ administration ANOVA (see Table 5). As expected with this type of analysis, the valence and force covariates were significant. More interestingly, the interaction terms for both valence and force motivation were significant (see Appendix B, Figure 1). The highest level of motivation was reported by students in the categorical rating format that provided ratings only once at the end of the course. Students who provided categorical ratings multiple times during the course experienced the lowest motivation. Students who provided holistic ratings on multiple occasions reported the highest motivation (see Appendix B, Figure 2). As with valence motivation, students who provided categorical ratings multiple times during the course experienced the lowest motivation.

To understand further the difference between the valence and force motivation results, the researchers conducted additional ANOVAs using P-O expectancy, importance, and E-P ratings as dependent variables. The E-P probability for students who provided holistic ratings was higher than students who provided categorical ratings ($M = 5.45$, $SD = 1.08$, and $M = 5.08$, $SD = 1.11$, respectively). This rating format main effect was significant ($F = 4.01$, $p \leq .04$). This result is expected, as holistic ratings are less complex than categorical ratings. There were no differences between the experimental conditions with respect to the expected effort students would exert to complete the peer ratings. However, students in the holistic format/multiple occasion condition had a higher expectation that their effort would result in accurate peer ratings (E-P probability), which when multiplied by valence in the Campbell (2003) formulation increased force motivation. There were no other significant main or interaction effects for P-O or importance ratings. The relationship between effort level and force motivation was significant ($r = .64$, $p \leq .05$) in the holistic format/multiple rating condition.

Student Perception Factor Score Anovas

Table 6 contains the ANOVA results for the three SPFQ factor scores. With respect to the student perceptions of the rating process, only the interaction term for the team function factor was significant. As illustrated in Appendix B, Figure 3, students who provided categorical ratings at the end of the course only reported that their team function was the highest. Students in the holistic rating/single rating frequency condition reported the lowest team function ratings.

We correlated the SPFQ factor scores with the PAEQ scores to determine the relationship between students' perceptions of the rating process, team dynamics, and their motivation to rate their peers. Team concerns was positively related to valence ($r = –.30$, $p \leq .05$) and force motivation

Table 4
Descriptive Statistics: Peer Assessment Expectancy Questionnaire (PAEQ) and Student Peer Feedback Questionnaire Factor Score Means and Standard Deviations

	Holistic Rating				Categorical Ratings			
	Single Administration N = 36		Multiple Administrations N = 39		Single Administration N = 34		Multiple Administrations N = 42	
	M	SD	M	SD	M	SD	M	SD
PAEQ–first administration								
Performance-to-outcome expectancy	5.17	0.90	5.24	0.82	4.85	1.18	4.84	1.02
Importance	5.72	0.75	5.88	0.73	5.86	0.77	5.58	0.79
Effort-to-performance expectancy								
Valence	4.21	1.07	4.38	1.09	4.05	1.22	4.01	1.30
Force	3.47	1.28	3.62	1.13	3.25	1.19	3.16	1.53
PAEQ–second administration								
Performance-to-outcome expectancy	4.37	1.34	4.78	1.00	4.24	0.93	4.08	1.16
Importance	5.28	0.93	5.68	0.97	5.42	0.85	5.13	1.06
Effort-to-performance expectancy	5.43	1.10	5.62	0.95	5.26	0.99	4.84	1.25
Valence	3.36	1.37	3.86	1.45	3.55	1.27	3.06	1.27
Force	2.68	1.39	3.15	1.48	2.81	1.21	2.22	1.24
Student Peer Final Questionnaire[a]								
Team function	–0.30	0.77	0.14	1.11	0.19	1.02	0.01	1.02
Assessments fairness	0.11	0.98	–0.17	1.02	–0.11	0.83	0.14	1.10
Team concerns	–0.04	0.99	0.02	1.09	0.09	0.87	–0.05	1.05

a. Standardized factor scores.

Table 5
Two-Way ANOVA on First and Second Peer Assessment ENxpectancy Questionnaire (PAEQ) Administrations

Dependent Variables	Independent Variables	SS	df	MS	F
PAEQ–first administration					
Performance-to-outcome expectancy	Rating frequency	0.03	1	0.03	0.41
	Holistic/categorical	4.90	1	4.90	68.38
	Interaction	0.07	1	0.07	0.07
Importance	Rating frequency	0.12	1	0.12	0.07
	Holistic/categorical	0.24	1	0.24	0.13
	Interaction	1.79	1	1.79	3.06
Effort-to-performance expectancy	Rating frequency	0.16	1	0.16	0.19
	Holistic/categorical	4.42	1	4.42	5.11*
	Interaction	1.27	1	1.27	1.47
Valence motivation	Rating frequency	0.17	1	0.17	0.12
	Holistic/categorical	2.70	1	2.70	1.93
	Interaction	0.41	1	0.41	0.29
Force motivation	Rating frequency	0.02	1	0.02	0.01
	Holistic/categorical	4.29	1	4.29	2.52
	Interaction	0.54	1	0.54	0.32
PAEQ–second administration					
Performance-to-outcome expectancy with covariate[a]	Covariate	45.84	1	45.84	45.90***
	Rating frequency	0.05	1	0.05	0.02
	Holistic/categorical	0.73	1	0.73	0.27
	Interaction	2.75	1	2.75	2.75
Importance with covariate[a]	Covariate	23.08	1	23.08	29.50***
	Rating frequency	0.05	1	0.05	0.02
	Holistic/categorical	0.73	1	0.73	0.38
	Interaction	1.89	1	1.89	2.42
Effort-to-performance expectancy with covariate[a]	Covariate	27.70	1	27.70	27.14***
	Rating frequency	0.96	1	0.96	0.94
	Holistic/categorical	1.49	1	1.49	1.46
	Interaction	0.95	1	0.95	0.93
Valence motivation with covariate[a]	Covariate	89.3	1	89.33	72.89***
	Rating frequency	0.01	1	0.01	0.01
	Holistic/categorical	0.83	1	0.83	0.67
	Interaction	6.55	1	6.55	5.34*
Force motivation with covariate[a]	Covariate	87.61	1	87.61	72.58***
	Rating frequency	0.22	1	0.22	0.18
	Holistic/categorical	1.43	1	1.43	1.18
	Interaction	7.96	1	7.96	6.59*

NOTE: MS = Mean Square; SS = Sum of Squares.
a. The first valence and force PAEQ administrations served as the covariates, thus holding the initial student motivation to rate their peers constant.
*$p \leq .05$. **$p \leq .001$.

Table 6
Two-Way ANOVA on Student Feedback Questionnaire Factor Scores

Dependent Variables	Independent Variables	SS	df	MS	F
Team functioning	Rating frequency	0.73	1	0.73	0.18
	Holistic/categorical	1.43	1	1.43	0.35
	Interaction	4.09	1	4.09	4.18*
Effort and assessments fairness	Rating frequency	0.01	1	0.01	0.01
	Holistic/categorical	0.08	1	0.08	0.02
	Interaction	3.11	1	3.11	3.12
Team concerns	Rating frequency	0.065	1	0.065	0.14
	Holistic/categorical	0.04	1	0.04	0.08
	Interaction	0.46	1	0.46	0.45

NOTE: MS = Mean Square; SS = Sum of Squares.
*$p \leq .05$.

($r = .19$, $p \leq .05$). Effort and assessment fairness was negatively correlated with valence ($r = –.31$, $p \leq .01$) and force motivation ($r = –.30$, $p \leq .01$). Team functioning was negatively related to students' E-P expectancy for the second PAEQ administration ($r = –.17$, $p \leq .05$).

Finally, using a random sample of ratings in the categorical conditions, we looked at descriptive statistics and frequency distribution to determine if there was variance across the five rating categories. The mean across the five rating categories was 20, and the standard deviation was 8.14, indicating that students differentiated between the dimensions.

Discussion

The data show mixed support that rating format and frequency influence students' motivation to rate their peers. The study found no support for Hypotheses 1 to 3, which predicted main effects for rating format and rating frequency. These independent variables, taken alone, did not influence the dependent variables. More interestingly, however, the data offer support for Hypothesis 4, which predicted a significant interaction between rating format and rating frequency. The direction of the interaction was unanticipated. Students who provided categorical ratings multiple times during the course experienced the lowest motivation. Students who provided holistic

ratings on multiple occasions reported the highest motivation. Although it cannot be determined directly from the data presented here, students may have felt that the added effort required to provide categorical ratings multiple times relative to holistic ratings may not have justified its limited benefits, at least with respect to team function. Students may have felt that providing their peers with overall holistic ratings on multiple occasions strikes an appropriate balance between the effort expended and the expected result of those efforts in terms of team outcomes. Perhaps students feel that holistic formats provide adequate peer feedback and that multiple occasions provide ratees more than one time interval to improve their contributions to their group.

Regardless of rating frequency or type of rating, students reported lower motivation to rate their peers at the end of the semester. Additionally, although not statistically significant, students who rated on multiple occasions using the categorical format experienced the largest decrease in motivation. Students in this experimental condition may have experienced greater fatigue, as they rated more frequently and made more judgments than students in the other experimental conditions made. Anecdotal evidence suggests that the "fatigue" factor may have influenced student motivation to rate peers. Instructors and teaching assistants reported that students in the multiple/categorical condition complained toward the end of the semester about the tedium of completing the categorical peer rating forms and indicated that they were "sick" of having to complete yet another rating of their peers.

The relationship between students' perceptions of their teams and the rating process and their motivation to provide meaningful peer assessments should be explored further. In this study, we found that students who were concerned with various aspects of teams (e.g., "team assignments allow students to 'hide' more easily than individual assignments") tended to have higher motivation to rate their peers. Peer ratings may provide students most concerned with the use of teams a mechanism to assign credit to able students and hold other students accountable for low performance. We also found that students who perceived that they themselves worked hard on the team and that peer assessments should be fair had lower motivation to rate their peers. Students may have felt that other students would not work as hard to provide peer assessments, rendering the feedback useless. The converse may also be true—students who put forth less effort and do not see peer ratings as a fair process were less motivated to rate their peers. Last, we found that students' perception of how well their team was functioning

was not related to motivation to rate peers, with the exception that students who reported their teams to function poorly possessed higher motivation to rate their peers.

Limitations

First, differences between instructors with respect to teaching competence, classroom management, and commitment to the research most likely varied somewhat. However, the researchers made several efforts to minimize these sources of error variance through assignment of instructors to the experimental conditions. The instructors were blind to the specific study hypotheses and were trained to administer the study instruments. The researchers also gave standardized presentations to students in each course section that explained the study purpose in broad terms and the instruments and stressed that student confidentiality of instrument responses would be maintained.

Second, the researchers did not investigate variables that may influence motivation, such as ways to increase student accountability for accurate peer ratings, training students to rate accurately, and student goal setting to increase their group contributions using the feedback they receive. Third, this study was not able to guarantee that students conducted a "mindful, careful" consideration of the peers' contributions. In the design of the study, we tried to create an environment whereby students would feel motivated to take the time to carefully consider their ratings of their peers and submit ratings that accurately reflected their perceptions of peer contributions. However, we have no way of knowing whether the students' ratings reflected mindful, careful consideration of their peers' contributions. For some of the students, the primary motivation may have been simply to "check off yet another box" or to complete the assessments as quickly as possible without carefully considering their input. As students were aware of the assessment instrument ahead of time, they could also have engaged in "gamesmanship" (for example, manipulating the ratings to make themselves look good or giving others equal marks regardless of contribution to discourage competition).

Finally, actual team performance may in fact be confounding the results of the study, not as a result of the peer rating system but as a cause, or correlate, of student evaluation of the system. There may be a common method bias with administering both surveys at the end of the semester. With low class sizes ($n \leq 19$), there were a small number of teams (four to five teams

per course section) in each condition so that differences in actual team performance would have a large impact.

Conclusion

The use of peer assessments for developmental feedback and as a component of students' final course grades deserves further research. Given our mixed results, it is difficult to draw any definitive recommendations from our research. The results of this study do serve to illustrate the complexity of the peer assessment process. Future research should continue to investigate ways to maximize student motivation to rate peers. Although the study suggests that holistic ratings given multiple times are viewed more positively by students, our mixed results make us hesitant to recommend this format and method of peer assessment over any other. This study does add to the existing peer assessment literature by using an experimental design in a field setting to assess student motivation to rate their peers' contributions to team projects. Additionally, we provide classroom practitioners with two peer assessment instruments, a presentation that can be used to train students to rate their peers and an Excel program to help automate the calculation of individual grades (available on request). Two questionnaires (PAEQ and SPFQ) were also developed and tested by the researchers.

We recommend that the effort required to provide peer assessments must be weighted against students' perceived benefits. Students' valence and motivation were lowest in the categorical multiple rating condition. Categorical ratings provided several times a semester (or for several courses in the same semester) require much effort and motivation. Motivation in typical classroom settings may not be sufficient to produce meaningful categorical ratings on multiple occasions. In addition, if peers do not improve based on the feedback they receive, students' motivation would decrease. Conversely, students in the holistic multiple administration condition evidenced the highest motivation (force) to rate, suggesting that an overall rating may require an amount of effort commensurate with expected benefits. Unfortunately, holistic ratings do not provide students with specific developmental feedback that they can use to improve individual performance and team functioning.

Further research will increase our understanding of the complexity of the peer assessment process. Specifically, we recommend that future research address the following questions:

- More research is needed to determine the impact of final grades on peer evaluations. What is the impact on the motivation of peer raters when peer evaluation has less or more of an influence on grades? How does the approach used to calculate individual course grades affect students' motivation to rate their peers? How might qualitative feedback (e.g., comments) in addition to the quantitative feedback provided by the holistic and categorical formats affect students' motivation to rate peers and team performance?
- When students are trained to rate their peers and give better feedback, are they more or less motivated to rate their peers? When students set goals based on peer feedback, are they more motivated to increase their individual contributions to group projects?
- What outcomes do students desire most from the peer assessment process? Does changing the PAEQ to reflect different outcomes alter student motivation to rate their peers? For example, Chen and Lu (2004) used expectancy theory to assess which peer assessment outcomes were most attractive to students. The results in order of attractiveness were as follows: determining peers' grades, reducing conflict and uneven work distribution, enhancing group productivity and collaboration, and improving peers' performance and behavior. How does the attractiveness of the Chen and Lu outcomes compare with the outcomes that emerged from our focus group?
- As student experience with teamwork and peer assessment increases, are they more or less motivated to rate their peers? The participants in the study reported here were first-semester freshmen. Would conducting the research with upper classmen more experienced with teamwork and peer assessment yield different results?
- Does varying the dimensions in the categorical assessment change the results? We used dimensions that primarily measured effort. Would the use of performance dimensions (i.e., quality of contributions, analysis, and writing) change the results?
- Does the use of behaviorally anchored rating scales change the results? Although we gave students behavioral examples during the training sessions, behavioral anchored rating scales were not used in either the categorical or holistic formats.
- The final administration of the PAEQ and SPFQ occurred on the same day and during the same week that students rated their peers. Although course logistics influenced the timing of the PAEQ and SPFQ administration, it may have confounded our results. Future research could examine the impact of student perceptions of their team's performance on motivation to rate peers and satisfaction with the peer assessment process.

- Research into specific student motivations to rate their peers is also needed. Conducting focus groups with students participating in the four experimental conditions might provide qualitative feedback that would help to explain the mixed results from this research. Focus groups might also enable us to get a feel for what motivates students as they complete peer assessments. Are they attempting to accurately convey their perceptions of team member contributions? Is their primary motivation merely to complete a course requirement as quickly as possible? Or are they trying to give good quality feedback that will help team members to improve individual performances and overall team performance? Are students engaging in "gamesmanship" to manipulate the system?

The research reported here underscores that simply requiring students to provide peer assessments that influence final grades is not sufficient to result in student motivation to rate. The conditions under which students rate their peers must be carefully considered, including the nature of the ratings themselves, how often ratings are provided, and students' motivational set. This research highlights the complexity of the peer rating process and the need for further research for a better understanding of the peer assessment process.

Appendix A

Exhibit 1
Peer Assessment Expectancy Questionnaire

Answer Questions 1, 2, and 3 by circling the answer that best describes your feelings.

Your responses are absolutely confidential.

	Question 1: Below are some things that could happen if you do an especially good job rating your peers. Look at each item and make a judgment about how likely is it that each outcome would happen to you if you did an ***especially good*** **job rating your peers.**	Not at all Likely		Somewhat Likely		Quite Likely		Extremely Likely
A	You will receive a better grade.	1	2	3	4	5	6	7
B	You will receive praise from your instructor.	1	2	3	4	5	6	7
C	You will feel better about yourself having given feedback to others.	1	2	3	4	5	6	7
D	You will develop better rating skills.	1	2	3	4	5	6	7
E	You will get a feeling of accomplishing something worthwhile.	1	2	3	4	5	6	7
F	Your peers will respect you.	1	2	3	4	5	6	7
G	Your peers will receive the grade they deserve.	1	2	3	4	5	6	7

	Question 2: Different people want different things ***from rating their peers.*** Here is a list of things a person could value. How important is each of the following to you? How important to you is . . .	Moderately important or less			Quite Important			Extremely Important
A	Receiving good grades?	1	2	3	4	5	6	7
B	Receiving praise from your instructor?	1	2	3	4	5	6	7
C	Giving feedback to others?	1	2	3	4	5	6	7
D	Developing your own rating skills?	1	2	3	4	5	6	7
E	Accomplishing something worthwhile?	1	2	3	4	5	6	7
F	Receiving respect from your peers?	1	2	3	4	5	6	7
G	Your peers receiving the grade they deserve?	1	2	3	4	5	6	7

Question 3: Below you will see pairs of factors. Indicate by checking the appropriate number to the right of each pair how often it is true for you personally that the factor on the left leads to the factor on the right. Remember, for each pair; indicate how often it is true by checking the number that seems most accurate.

			Never		Sometimes		Often		Almost Always
Working hard to rate my peers	→	ratings that accurately reflect each peer's contribution	1	2	3	4	5	6	7
Working hard to rate my peers	→	ratings that will help my peers	1	2	3	4	5	6	7
Working hard to rate	→	high quality feedback	1	2	3	4	5	6	7

General Information

How attractive to you is it to provide peer ratings? (circle one)

–5 –4 –3 –2 –1 0 1 2 3 4 5

Very Unattractive **Very attractive**

Indicate the level of effort you expect to exert to complete the peer ratings. (circle one)

0 1 2 3 4 5 6 7 8 9 10

Zero Effort **Great Deal of Effort**

What college year are you? (check one) ☐ Freshman ☐ Sophomore ☐ Junior ☐ Senior

In college, how many courses have you been asked to rate other students' performance? ☐ never ☐ 1-2 courses ☐ 3-4 courses ☐ More than 4 courses

With respect to your overall ☐ 1 ☐ 2 ☐ 3 ☐ 4 ☐ 5 ☐ 6 ☐ 7

college performance so far, how would you rate yourself?	Well Below Average	Below Average	Average	Above Average	Well Above Average

What is your gender? ☐ Male ☐ Female

Which best describes Your race? ☐ White ☐ African American ☐ Asian ☐ Hispanic ☐ Native American ☐ Other

What is your major? ☐ Business Administration ☐ Finance ☐ HRM ☐ Management Accounting ☐ Public Accounting (BS) ☐ Public Accounting (BS/MBA) ☐ Marketing ☐ Management Science

Exhibit 2
Categorical Rating Peer Feedback Instructions

Imagine that you have $100 to divide among yourself and your teammates for each of the following categories:

- Attendance at team meetings
- Completion of fair share of work
- Participation at team meetings
- Meeting deadlines
- Encouraging and supporting other team members

Your allocation of the $100 for *each* category should reflect the contributions of each team member to the Company Report team project. Below is a sample of the peer feedback worksheet:

		Dollars Given to Each Team Member				
Category		Brian	Fred	Sally	Maureen	Mark
Attendance at team meetings	$ 100					
Completion of fair share of work	$ 100					
Participation at team meetings	$ 100					

Meeting deadlines	$ 100
Encouraging and supporting other team members	$ 100

Exhibit 3
Holistic Feedback Presentation with Examples and Answers to Student Questions

The following examples were given to students in the holistic conditions during a peer feedback training session. Examples to students in the categorical conditions were similar but used the categorical format.

Example 1:

Suppose you feel that each of your team members contributed *equally* to the team project (include yourself–let us call you "Brian"). In that case, the $100 would be divided equally with each team member receiving $20.

		Dollars Given to Each Team Member				
		Brian	Fred	Sally	Maureen	Mark
Contribution to the team project	$ 100	20	20	20	20	20

Example 2:

Suppose you feel that you and your team members did not contribute equally. You gave team members different amounts of money based on how you felt each contributed to the Company Report project (*the dollars must add up to 100)*. Study the worksheet below. Sally attended all of the meetings, arriving first and leaving last. She also encouraged others to do a good job on the project, and met all her deadlines for the parts of the project assigned to her by the team. On the other hand, Fred skipped a few meetings, and even left another meeting early. Fred also was late in delivering his assignments to the team. Since the "fair share" is $20, you therefore gave Sally 30 points and Fred only 10 points.

		Dollars Given to Each Team Member				
		Brian	Fred	Sally	Maureen	Mark
Contribution to the team project	$ 100	20	10	30	20	20

Q: What will my feedback look like?

A: You will learn how you and the team rated you on your overall contribution to the team project. You already know how many dollars you gave yourself. You will also learn the average dollar amount your team gave you.

Q: How are grades determined from the peer feedback ratings?

A: Remember, all team members will provide feedback ratings. The average dollar amount is first determined. This average is then divided by the $20 fair share and then multiplied by the grade assigned to the team project.

For example, suppose your team received a team project grade of 90%. If your average dollar amount was $20, then divide $20/$20 (which is 1) × 90%. Since you did your fair share, you receive the entire 90% grade.

Suppose, however, you didn't do your fair share, and your average dollar amount was $17. In this case, you would divide $17/$20 (which is .85). You would therefore receive .85 of the 90%, which equals 77%.

On the other hand, suppose your average dollar amount is $22. You do the math: $22/$20 = 1.1, multiplied by 90% is 99%. You can actually get higher than the assigned grade for the project by being a great team member!

Q: Are my feedback ratings confidential?

A: Absolutely! The only feedback that will be given to students will be average ratings the team gave. No student will know who made what rating, nor will any student see individual ratings. So give honest feedback and rate your team members as you see it!

Appendix B

Figure 1
Student Valence to Rate Their Peers as a Function of Rating Format and Rating Frequency.

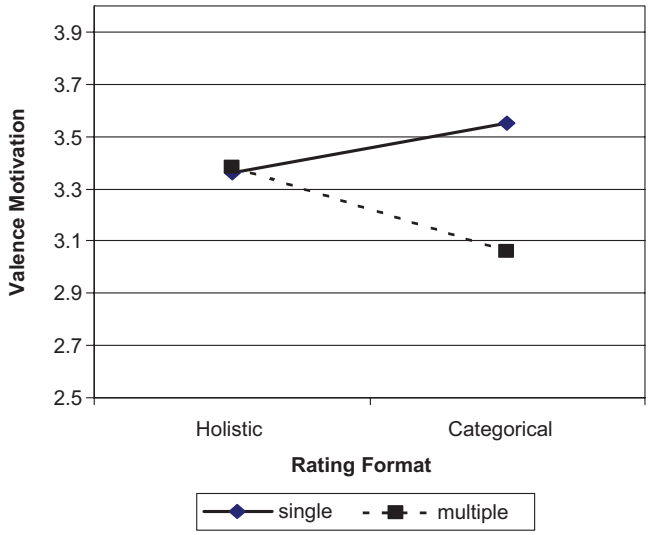


Figure 2
Student Force Motivation to Rate Their Peers as a Function of Rating Format and Rating Frequency.

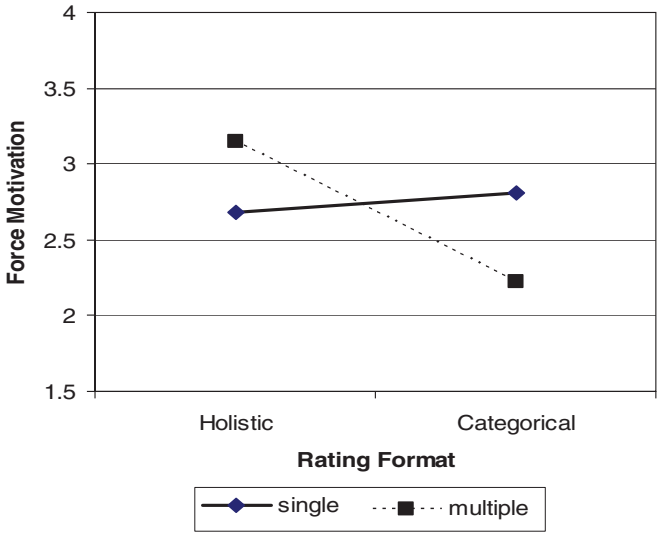
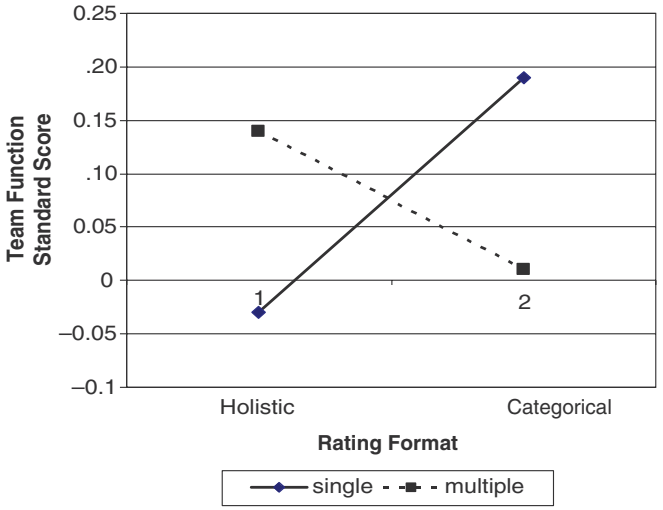


Figure 3
Team Function as a Function of Rating Format and Rating Frequency



References

Aiken, L. S., & West, G. W. (1991). *Multiple regression: Testing and interpreting interactions*. Newbury Park, CA: Sage.

Bacon, D. R., Stewart, K. A., & Silver, W. S. (1999). Lessons from the best and worst team experiences: How a teacher can make a difference. *Journal of Management Education*, *23*(5), 467-488.

Ballantyne, R., Hughes, K., & Mylonas, A. (2002). Developing procedures for implementing peer assessment in large classes using an action research process. *Assessment & Evaluation in Higher Education*, *27*(5), 427-441.

Brindley, C., & Scoffield, S. (1998). Peer assessment in undergraduate programs. *Teaching in Higher Education*, *3*(1), 79-89.

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.

Campbell, S. V. (2003). Using expectancy theory to assess gender differences in Russian accounting student motivation. *Academy of Educational Leadership Journal*, *7*(3), 1-14.

Chen, Y., & Hoshower, L. B. (2003). Student evaluation of teaching effectiveness: An assessment of student perception and motivation. *Assessment and Evaluation in Higher Education*, *28*(1), 71-88.

Chen, Y., & Lu, H. (2004). Students' perceptions of peer evaluation: An expectancy perspective. *Journal of Education for Business*, *79*(5), 275-282.

Cheng, W., & Warren, M. (1997). Having second thoughts: Student perceptions before and after a peer assessment exercise. *Studies in Higher Education*, *22*(2), 233-240.

Cox, P. L., & Bobrowski, P. E. (2004). Power tools for teams: A model for improving the teamwork skills of first-year business students. *Journal of Behavioral and Applied Management*, *5*(3), 204-227.

DeNisi, A. S., Randolph, W. A., & Blencoe, A. G. (1982). Potential problems with peer ratings. *Academy of Management Journal*, *26*, 508-520.

Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology*, *84*(1), 58-74.

Erez, A., Lepine, J. A., & Elms, H. (2002). Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology*, *55*, 929-948.

Erez, M., & Somech, A. (1996). Is group productivity loss the rule or the exception? Effects of culture and group-based motivation. *Academy of Management Journal*, *39*(6), 513-537.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Innovations in Education & Training International*, *32*(2), 175-187.

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research*, *70*(3), 287-322.

Feichtner, S. B., & Davis, E. A. (1985). Why some groups fail: A survey of students' experiences with learning groups. *Organizational Behavior Teaching Review*, *9*(4), 75-88.

Feichtner, S. B., & Davis, E. A. (1992). *Collaborative learning: A sourcebook for higher education*. University Park, PA: National Center of Postsecondary Teaching, Learning and Assessment.

Friedman, B. A., & Lechner, T. (2005, March). *The development of a goal setting and expectancy questionnaire to identify at risk students*. Paper presented at the National

Academic Advising Association (NACADA) Region 1 Conference, Montreal, Canada. (Available from the authors on request; friedman@oswego.edu).

Greguras, G. J., Robie, C., & Born, M. P. (2001). Applying the social relations model to self and peer evaluations. *The Journal of Management Development*, *20*(6), 508-525.

Gueldenzoph, L. E., & May, G. L. (2002). Collaborative peer evaluation: Best practices for group member assessments. *Business Communications Quarterly*, *65*(1), 9-21.

Hanrahan, S. J., & Issacs, G. (2001). Assessing self and peer-assessment: The students' views. *Higher Education Research and Development*, *20*(1), 53-70.

Harkins, S. G. (1987). Social loafing and social facilitation. *Journal of Experimental Social Psychology*, *23*, 1-18.

Isaac, R. Z., Wilfred, J. P., & Douglas, C. (2001). Leadership and motivation: The effective application of expectancy theory. *Journal of Managerial Issues*, *13*(2), 3045-3695.

Karou, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integrations. *Journal of Personality and Social Psychology*, *65*(4), 681-706.

Kelmar, J. H. (1993). Peer assessment in graduate management education. *International Journal of Educational Management*, *7*(2), 4-7.

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.

Levi, D., & Cadiz, D. (1998). *Evaluating team work on student projects: The use of behaviorally anchored scales to evaluate student performance* (ERIC Document Reproduction Service, ED 424250). Washington, D.C.: Educational Resources Information Center.

Liden, R. C., & Mitchell, T. R. (1983). The effects of group interdependence on supervisor performance evaluations. *Personnel Psychology*, *36*, 643-653.

McDowell, L. (1995). The impact of innovative assessment on student learning. *Innovations in Education and Training International*, *32*(4), 302-313.

Mello, J. A. (1993). Improving individual member accountability in small work group settings. *Journal of Management Education*, *17*(2), 253-259.

Mohrman, A. M., & Lawler, E. E. (1983). Motivation and performance appraisal behavior. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 173-189). Hillsdale, NJ: Lawrence Erlbaum.

Nadler, D. A., & Lawler, E. E. (1977). Motivation: A diagnostic approach. In J. R. Hackman, E. E. Lawler, & L. Porter's (Eds.), *Perspectives on behavior in organizations* (pp. 67-87). New York: McGraw-Hill.

Naiper, N., & Latham, G. (1986). Outcome expectancies of people who conduct performance appraisals. *Personnel Psychology*, *39*, 827-837.

Orsmond, P., Merry, S., & Reiling, K. (1996). The importance of marking criteria in the use of peer assessment. *Assessment & Evaluation in Higher Education*, *21*, 239-250.

Pond, K., & Ul-Haq, R. (1997). Learning to access students using peer review. *Studies in Educational Research*, *23*(4), 331-348.

Porter, L. W., & Lawler, E. E. (1968). *Managerial attitudes and performance*. Burr Ridge, IL: Irwin.

Rummel, R. J. (1970). *Applied factor analysis*. Evanston, IL: Northwestern University Press.

Rynes, S. L., & Gerhart, B. (2000). *Compensation in organizations: Current research and practice*. San Francisco: Jossey-Bass.

Stahl, M. J., & Harrell, A. M. (1981). Modeling effort decisions with behavioral decision theory: Toward an individual differences model of expectancy theory. *Organizational Behavior and Human Performance*, *27*(3), 303-325.

Strong, J. T., & Anderson, R. E. (1990). Free-riding in group projects: Control mechanisms and preliminary data. *Journal of Marketing Education*, *12*, 61-67.

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, *68*(3), 249-276.

Van Velsor, E., & Leslie, J. B. (1991). *Feedback to managers: A review and comparison of 16 multirater feedback instruments*. Greensboro, NC: Center for Creative Leadership.

Vroom, V. C. (1964). *Work and motivation*. New York: John Wiley & Sons.

Williams, D. L., Beard, J. D., & Rymer, J. (1991). Team projects: Achieving their full potential. *Journal of Marketing Education*, *76*(1), 56-60.

Young, C. B., & Henquinet, J. A. (2000). A conceptual framework for designing group projects. *Journal of Education for Business*, *76*(1), 56-60.