

# Journal of Management Education

<http://jme.sagepub.com/>

---

## **Peer Assessment in Small Groups: A Comparison of Methods**

Diane F. Baker

*Journal of Management Education* 2008 32: 183 originally published online 30  
November 2007

DOI: 10.1177/1052562907310489

The online version of this article can be found at:  
<http://jme.sagepub.com/content/32/2/183>

---

Published by:



<http://www.sagepublications.com>

On behalf of:



[OBTS Teaching Society for Management Educators](#)

**Additional services and information for *Journal of Management Education* can be found at:**

**Email Alerts:** <http://jme.sagepub.com/cgi/alerts>

**Subscriptions:** <http://jme.sagepub.com/subscriptions>

**Reprints:** <http://www.sagepub.com/journalsReprints.nav>

**Permissions:** <http://www.sagepub.com/journalsPermissions.nav>

**Citations:** <http://jme.sagepub.com/content/32/2/183.refs.html>

>> [Version of Record](#) - Mar 10, 2008

[OnlineFirst Version of Record](#) - Nov 30, 2007

[What is This?](#)

# PEER ASSESSMENT IN SMALL GROUPS: A COMPARISON OF METHODS

---

Diane F. Baker  
*Millsaps College*

*This article describes and evaluates several peer evaluation tools used to assess student behavior in small groups. The two most common methods of peer assessment found in the literature are rating scales and single score methods. Three peer evaluation instruments, two using a rating scale and one using a single score method, are tested in several management courses to examine their effectiveness. All three instruments demonstrate acceptable levels of reliability and are found to be correlated with individual performance measures. The article concludes with a discussion of the advantages and disadvantages of each instrument.*

**Keywords:** *peer assessment; group; appraisal; feedback*

Those of us who frequently make group assignments in the classroom are faced with the dilemma of how to evaluate fairly the contribution of each group member. We also struggle with finding ways to help our students become better group members. Whether we seek to evaluate students and/or help them improve their group skills, we must first find an effective way to collect information about how they supported their group's efforts. Instructors may choose to measure outcomes, such as a student's part in a group presentation or paper, arguing that the final product is the most critical measure of performance. Although an assessment of overall performance is important, when the instructor focuses simply on the end result of a group project, much information is lost about specific task and relationship behaviors that affect group success, such as the extent to which each group member took initiative, researched the issues, contributed ideas, met group deadlines, facilitated

---

**Author's Note:** Address correspondence to Diane F. Baker, Millsaps College, 1701 North State Street, Jackson, MS 39210; e-mail: bakerdf@millsaps.edu.

JOURNAL OF MANAGEMENT EDUCATION, Vol. 32 No. 2, April 2008 183-209

DOI: 10.1177/1052562907310489

© 2008 Organizational Behavior Teaching Society

problem solving, and helped resolve group conflict. For the sake of equity, this information is appropriately considered when determining a student's grade. For the sake of learning, this information can be shared with the students to help them gain greater understanding about the extent to which their knowledge, work products, and behaviors contributed to the group's effectiveness and overall performance.

Peer evaluation instruments are commonly used to get input from the group about each member's contributions. These instruments come in many different forms; instructors must therefore make several decisions regarding the type of instrument to use, the behaviors that will be measured, and the extent to which the peer evaluation results will affect grades. Threats to validity should also be considered and are the same as those associated with other types of performance appraisals, including rater bias, halo effect, central tendency, leniency, and strictness effects (Druskat & Wolff, 1999; Greguras, Robie, & Born, 2001). Despite the challenges and validity concerns associated with peer evaluations, research indicates that they can provide valuable information about group member performance (e.g., Druskat & Wolff, 1999; Erez, Lepine, & Elms, 2002; Fedor, Bettenhausen, & Davis, 1999; Fox, Zeev, & Yinon, 1989; Gatfield, 1999; Haas, Haas, & Wotruba, 1998; Mumford, 1983). The goal of this article is to examine different approaches to peer assessment in general and then to compare and evaluate three specific peer evaluation instruments that can be used to assess student performance in groups. The intent of this analysis is to help readers choose or design a method that is suited to their particular needs and context.

#### DEVELOPMENTAL AND EVALUATIVE PURPOSES OF PEER ASSESSMENT

In work and educational settings, both developmental and evaluative approaches have commonly been used to provide feedback about an individual's performance (Druskat & Wolff, 1999; Farh, Cannella, & Bedeian, 1991; Fedor et al., 1999). The primary goal of developmental feedback is to enhance performance of the individual and/or group by identifying the discrepancies between a member's expected and actual performance, thereby giving the member the opportunity to take corrective action. In contrast, evaluative assessment is used for administrative decisions, usually those involving the distribution of rewards; in the classroom, these rewards are grades.

*Using peer evaluation for development.* Instructors cannot assume that students will develop team skills simply by participating in group projects; learning the skills that improve group performance requires practice and feedback (see Whetten & Cameron, 2002, for a summary of the research). If feedback is provided during the middle of the semester, students have the opportunity to improve their team skills before the group finishes its tasks.

For example, Druskat and Wolff (1999) make a case for structured, developmental peer appraisal after the group has been together for a short period and just before members begin work on a major project because this early feedback can have a positive effect on task planning, communication, and motivation. Brooks and Ammons (2003) suggested another approach, conducting a peer evaluation at the end of each major learning unit. Researchers have found that collecting and sharing peer feedback with students increases self-awareness, workload sharing, likelihood of speaking in the group, cooperation among members, and as a result, higher group performance (Brooks & Ammons, 2003; Druskat & Wolff, 1999; Erez et al., 2002; Greguras et al., 2001). It is interesting that some evidence indicates that peer feedback used solely for developmental purposes may be more accurate than that used for evaluation (Farh et al., 1991).

Peer evaluation is not the only way to help students assess their strengths and weaknesses as a group member. Meyer (1991) suggested that self-appraisal is an effective way to encourage skill development, largely because it may increase one's commitment to change. In the work setting, employee participation in the appraisal process is related to satisfaction (Cawley, Keeping, & Levy, 1998). Given the extent to which self-appraisals are used in organizations (Cawley et al., 1998; Keeping, 2003; Meyer, 1991), it may be helpful to give students experience with self-assessment in the classroom (Johnston & Miles, 2004). Many instruments that are designed for peer evaluation can also be used for self-appraisal.

The argument against self-appraisal is that students may be susceptible to self-serving biases that diminish their ability to assess themselves accurately. For example, Johnston and Miles (2004) found that members rated themselves higher than their teammates did, and there was no correlation between self and peer assessments. Nevertheless, Johnston and Miles decided to continue using self-ratings because they thought it was important to encourage self-reflection.

*Using peer assessment for evaluation.* When peer assessment is used for evaluation purposes, it provides accountability for an individual's contributions to group assignments. Though instructors may question the appropriateness of allowing students to influence the grades of their peers, researchers have justified the use of peer ratings for administrative purposes because peers are frequently in the best position to observe relevant behaviors and ratings can be aggregated across peers to increase reliability (Greguras et al., 2001; Mumford, 1983; Murphy & Cleveland, 1995). Students concerned with the fair distribution of work among group members found increased satisfaction with group work when peer assessment was used to reward those who made a greater contribution to group performance (Chapman & Van Auken, 2001; Erez et al., 2002). However, the use of peer evaluations does not ensure the

group experience will be a positive one for students (Bacon, Stewart, & Silver, 1999).

When used for evaluative purposes only, peer assessment is typically conducted at the end of the project after members have completed their responsibilities (e.g., Beatty & Haas, 1996; Greguras et al., 2001; Michaelsen, Knight, & Fink, 2004). This type of peer evaluation is less useful for developmental purposes because it comes at the end of the project and the student does not have the opportunity to take corrective action. In many cases, students never learn what their peer scores were (e.g., Bacon et al., 1999), eliminating any potential effect on future behavioral change.

*Combining developmental and evaluative purposes.* An instructor can choose to use peer assessment methods for both developmental and evaluative purposes. For example, a structured peer appraisal like that recommended by Druskat and Wolff (1999) can be used during the middle of a course or group project, giving students the chance to change their behavior and improve their performance in response to the feedback they receive. At the end of the course or project, peer assessment can be conducted once again, this time for the purpose of assigning individual grades. Ideally, final peer scores should be made available to students who wish to see them. Although people tend to be more defensive about feedback when it is linked to rewards (Meyer, 1991), students may develop greater self-awareness by learning how they were perceived by their teammates. If developmental feedback is offered during the middle of the course, there should be fewer surprises at the end of the course when peer scores are used for evaluative purposes.

Whether peer evaluation in the classroom is used for developmental or evaluative purposes, the instrument should be both practical and valid. The instrument should be practical for the instructor in that it is easily distributed, completed, and tabulated. The information obtained must also be accurate if it is to be useful for facilitating individual and group development or perceived as fair when making administrative decisions. It is conceivable that the type of instrument used or the behaviors measured may differ depending on whether the instrument is used for developmental or evaluative reasons. If the instrument is used for development, information about various behaviors important for group success must be discussed. This kind of detail may not be necessary for evaluation purposes if the item or items on the instrument can capture enough information to accurately reflect overall contribution to the group's performance.

## TYPES OF ASSESSMENTS

Instructors have used a variety of methods to assess peer performance in small groups. Most peer evaluation instruments that are described in the

literature use graphic rating scales (see Beatty & Haas, 1996; Chalupa, Chen, & Sormunen-Jones, 2000; Erez et al., 2002; Greguras et al., 2001; Gueldenzoph & May, 2002; Halfhill & Nielsen, 2007; Haas et al., 1998; Johnson & Smith, 1997; Paswan & Gollakota, 2004; Persons, 1998; Rafiq & Fullerton, 1996; Strom & Strom, 2002). Another popular method is to have peers allocate points based on overall contributions to the group (Drexler, Beehr, & Stetz, 2001; Michaelsen et al., 2004; Saavedra & Kwun, 1993). Other approaches include a set of paired comparisons (Johnson & Smith, 1997), a sociogram (Cooke, Drennan, & Drennan, 1997), a nomination method (Kane & Lawler, 1978), and a project diary (Rafiq & Fullerton, 1996). Some instruments also include a comments section. Each approach is briefly discussed later.

*Rating scales.* Rating scales are used to assess a variety of behaviors and can provide more detailed information about the ratee than other methods (Kane & Lawler, 1978). The peer rating instruments found in this review varied from 1 to 35 items on 5- to 100-point scales and were built around one or more of eight basic behavioral components:

1. Attended group meetings: was available to meet, came to meetings, was on time, did not leave early (e.g., Brooks & Ammons, 2003; Chalupa et al., 2000; Gatfield, 1999; Gueldenzoph & May, 2002; Haas et al., 1998; Lejk & Wyvill, 2001).
2. Was dependable: met deadlines, kept his or her word (e.g., Beatty & Haas, 1996; Brooks & Ammons, 2003; Chalupa et al., 2000; Clark, 1989; Gatfield, 1999; Gueldenzoph & May, 2002; Haas et al., 1998; Lejk & Wyvill, 2001; Rafiq & Fullerton, 1996).
3. Submitted quality work: contributions were of high quality (e.g., Beatty & Haas, 1996; Chalupa et al., 2000; Clark, 1989; Greguras, Robie, & Born, 2001).
4. Exerted effort and/or extra effort: did his or her share of the work or more than fair share, took an active role in getting tasks done (sometimes the specific tasks necessary to complete the project were expressly stated in the instrument), volunteered for tasks (e.g., Beatty & Haas, 1996; Brooks & Ammons, 2003; Chalupa et al., 2000; Cheng & Warren, 2000; Clark, 1989; Greguras et al., 2001; Gueldenzoph & May, 2002; Johnson & Smith, 1997; Johnston & Miles, 2004; Lejk & Wyvill, 2001).
5. Cooperated/communicated with other members: got along with others, communicated well with group members, shared information, listened (e.g., Beatty & Haas, 1996; Chalupa et al., 2000; Greguras et al., 2001; Haas et al., 1998; Halfhill & Nielsen, 2007; Johnson & Smith, 1997; Lejk & Wyvill, 2001; Strom & Strom, 2002).
6. Managed group conflict: helped resolve interpersonal or group conflict, helped create an environment that minimized destructive group conflict (e.g., Chalupa et al., 2000; Gueldenzoph & May, 2002; Halfhill & Nielsen, 2007; Rafiq & Fullerton, 1996).
7. Made cognitive contributions: possessed and applied the necessary knowledge and skills to accomplish group goals (e.g., Chalupa et al., 2000; Cheng &

Warren, 2000; Cohen & Ledford, 1994; Gatfield, 1999; Greguras et al., 2001; Gueldenzoph & May, 2002; Johnson & Smith, 1997; Lejk & Wyvill, 2001; Rafiq & Fullerton, 1996; Strom & Strom, 2002).

8. Provided structure for goal achievement: established group goals, identified/assigned tasks, monitored progress (Chalupa et al., 2000; Halfhill & Nielsen, 2007).

In addition, numerous authors included one or more items assessing a member's total contribution to the group ("overall evaluation," "desirability as a future coworker," "committed to group goal"; see Beatty & Haas, 1996; Brooks & Ammons, 2003; Chalupa et al., 2000; Clark, 1989; Haas et al., 1998; Johnson & Smith, 1997; Rafiq & Fullerton, 1996). According to most authors, items were selected based on literature reviews, knowledge gained from previous group experience, and/or suggestions generated by the groups that would be using them.

One of the more rigorous attempts to develop a valid peer rating instrument was made by Paswan and Gollakota (2004), who created a 35-item form, using a 5-point Likert-type scale. They used principal component analysis and tests for internal consistency, convergent, and discriminant validity to reduce redundancy, ambiguity, and lack of fit. The principal components analysis conducted by Paswan and Gollakota revealed five factors: competence, task and maintenance orientation, domineering behavior, dependability, and free-riding behavior. These factors appear to be related to five of the basic behavioral components noted earlier: specifically, made cognitive contributions, cooperated/communicated with other members, managed group conflict, attended group meetings, and exerted effort. In Paswan and Gollakota's study, dependability was strictly a function of attending meetings, whereas this component was more broadly defined by instruments in other studies as meeting deadlines, following through, and so forth. Paswan and Gollakota's instrument does not address the quality of work submitted by group members.

A particular type of rating scale, behaviorally anchored rating scales (BARS), should theoretically be more valid than other rating instruments because each point on the rating scale is associated with specific, observable behaviors that are critical to successful group performance, thereby reducing ambiguity. However, empirical studies to date have frequently failed to demonstrate that BARS have an advantage over graphic rating scales (Kingstrom & Bass, 1981; Solomon & Hoffman, 1991; Tziner, Joanis, & Murphy, 2000). When evaluating teachers, for example, Solomon and Hoffman (1991) found that BARS resulted in fewer leniency and halo errors, but differences were not enough to offset the costs associated with the development and implementation of BARS. Kingstrom and Bass (1981) noted that many of the findings in comparison studies were inconclusive because of methodological problems, such as small sample sizes, differences in scale anchor points, and differences in dimension names. Studies indicate that BARS are potentially valid instruments,



and they continue to be used effectively in a variety of settings (e.g., Harrell & Wright, 1990; Hedge, Borman, Bruskiwicz, & Bourne, 2004; Pounder, 2002; Ramus & Steger, 2000).

*Allocation of points.* Another common peer assessment method used in small group settings involves an allocation of points based on overall contribution to group performance. In an approach recommended by Michaelsen et al. (2004), the number of points to be allocated is determined by multiplying the number of rateres by 10. Thus, if a group member had to assess four peers, he or she would assign a total of 40 points ( $4 \times 10$ ) based on the contribution of each. Drexler et al. (2001) asked each group member to rate others in the group on a scale from 80% to 120%, with the stipulation that the average rating for the group total was 100%. Several researchers had peers allocate a total of 100 points among group members based on contributions (e.g., Lejk & Wyvill, 2001; Saavedra & Kwun, 1993). Some instructors do not allow students to assign all their teammates the same score, requiring that they give at least one person a higher or lower score than other members. Forcing students to engage in the challenging work of making distinctions among peers encourages them to pay attention and practice the art of giving feedback.

*Peer comparisons.* Another set of methods involved comparing group members to each other. For example, Johnson and Smith (1997) had members make a series of paired comparisons on each of five dimensions: effort, cooperation, initiative, technical knowledge/expertise, and overall contribution. Cooke et al. (1997) recommended a method they called a sociogram. In this approach, each group member identifies the group member who was most outstanding on one or more performance dimensions (e.g., “the most cooperative,” “most responsible in developing the proposal,” and “most task oriented”). Their sample forms included 20 dimensions. Points are assigned based on the number of times a student is listed by his or her peers for each dimension. This appears to be similar to the nomination method described by Kane and Lawler (1978), in which members name a designated number of peers who were best on one or more performance dimensions. Kane and Lawler also noted that members could be asked to identify the worst performers in one or more categories.

*Project diaries.* Rafiq and Fullerton (1996) used project diaries to assess member contributions made at various stages in the group project. Critical tasks and behaviors such as “planning the project,” “suggesting ideas,” and “writing a report” were listed on a form. At various checkpoints during the semester, peers were asked to list the names of the group members who had performed those specific tasks. At the end of the semester, the instructors counted how often each student was mentioned and compared that to the

maximum number of times that a student could be mentioned. This approach served to clarify performance expectations, ensure accountability, and reduce the effect of memory deterioration on peer assessment.

### GRADING ISSUES

When peer assessment is used for evaluation, the instructor must decide how to score the results of the appraisal method. How should peer assessment translate into grade decisions? Each approach has its own problems. If nominations are used, several group members are not likely to be mentioned at all, leaving no information for developmental feedback or administrative decisions (Kane & Lawler, 1978). If ranking is used, there is no information about the extent to which members differ from one another. Ranking or nominations may be useful for groups that have members whose contributions are easily differentiated, but what if the performances of two or more members are perceived as similar?

For example, consider two groups of five students. In the first group, two members do all the work, while the other three members do nothing. In the second group, all five members work very hard to achieve group goals. If peer comparisons are used to determine participation grades (e.g., Cooke et al., 1997; Drexler et al., 2001; Michaelsen et al., 2004; Saavedra & Kwun, 1993), the two hard-working members of the first group will probably receive the vast majority of nominations or points from their peers because they obviously contributed far more to group performance than the three loafers, who are unlikely to receive many nominations or points at all. In contrast, perceptions about who gave the most effort in the second group will vary because differences in contributions are minor; nominations or points will be more evenly divided among the members and all scores will fall below that of the top two performers in the first group. The person ranked third in the first group did nothing, but the person ranked third in the second group worked very hard. How does one then assign grades?

Rating systems seem easier than ranking systems to convert into grades because they are based on absolute, as opposed to relative, standards. A notable problem, however, is that peer ratings tend to be inflated because of leniency effects; raters tend to use the upper end of the scale only (Greguras et al., 2001; Johnson & Smith, 1997; Paswan & Gollakota, 2004). To determine a grade from ratings, some instructors divide an individual's score by the total points possible (e.g., Beatty & Haas, 1996). For example, if an instrument contained 10 items on a 5-point scale (1 representing poor performance, 5 representing high performance), a student could receive as many as 50 points ( $10 \times 5$ ) multiplied by the number of raters. A traditional grading scale could then be used; students who received 90% or more of the total points possible would achieve an "A," 80% or more of the total points

possible would achieve a “B,” and so forth. Alternatively, the person who received the most number of points in the group could set the standard, receiving a score of 100% (e.g., Johnson & Smith, 1997). Grades for other members would then be calculated by dividing each member’s total points by the number of points received by the highest performer in the group. Paswan and Gollakota (2004) dropped the lowest single peer evaluation before determining the grade. Regardless of whether the lowest grade is dropped, leniency effects that typically occur with peer evaluation systems may lead to grade inflation.

A key concern when using peer evaluations for grading purposes is the extent to which group members accept the ratings of their peers, questioning the freedom, for example, from bias (Fedor et al., 1999). Based on studies from the workplace, doubts about rating accuracy may be more likely when peer assessment is used for evaluative purposes (Fedor et al., 1999; McEvoy & Buller, 1987). When peer feedback affects the course grade, student concerns about accuracy are legitimate and they deserve a process that provides valid information.

### A COMPARISON OF THREE METHODS

The question is: Which, if any, peer assessment method or form is better when used for developmental and/or evaluative purposes? Are the differences enough to matter? Ideally, in an instructional setting, a peer evaluation form should be easy to implement, easy to score, provide good feedback to the learner, motivate higher levels of positive behaviors, be perceived as fair, and be valid and reliable. Is it too much to ask that a performance evaluation method achieve all these results? Using these criteria, the remainder of the article describes three different peer assessment tools that have been tested and compared in several management classes and explains the advantages and disadvantages of each for development and evaluation.

*Context.* The courses used to assess the peer evaluation tools were designed using a Team Learning model of instruction, an approach that relies heavily on group activities to meet learning objectives (see Michaelsen et al., 2004). Groups of five to seven members were assigned at the beginning of the semester and remained intact to accomplish assignments of various duration and complexity. Graded assignments included five to six quizzes that were taken first by individuals and then again as a group. These quizzes came at the beginning of each unit and primarily tested students’ understanding of textbook concepts. Students received feedback about their individual and group performance on the quizzes during the same class period in which they took the quizzes. In all classes, students completed a peer evaluation form at the end of the course that assessed the contributions of

**TABLE 1**  
**Comparison of Peer Evaluation Forms**

	<i>Short Form</i> ( <i>n</i> = 128)	<i>Long Form</i> ( <i>n</i> = 151)	<i>Points Allocation</i> <sup>a</sup> ( <i>n</i> = 191)
Time required to complete	5 to 10 min	15 to 20 min	5 to 10 min
Interrater reliability	.80	.82	.82
Item reliability	.91	.94	—
Correlation with quiz scores	.42 ( <i>p</i> < .001)	.45 ( <i>p</i> < .001)	.42 ( <i>p</i> < .001)
Gender differences	Male mean: 3.36 Female mean: 3.57 <i>p</i> = .017	Male mean: 3.30 Female mean: 3.45 <i>p</i> = .046	Male mean: 9.77 Female mean: 10.66 <i>p</i> = .002
Average peer grade	89%	90%	83%
Behaviors assessed	9 items: preparation, cognitive ability, effort, commitment to performance, communication, initiative, conflict resolution, overall contribution to task, overall contribution to relationship	4 items: preparation, communication, commitment to performance, cooperation	Overall contribution only

a. At times, the points allocation method was administered together with the long form (*n* = 52) or the short form (*n* = 98).

their group members. Scores from the last peer evaluation were used to determine the students' participation grade, which varied in weight from 10% to 15% of the final course grade.

*Instruments.* To identify the advantages and disadvantages of various peer assessment methods, two different peer evaluation rating instruments were developed and tested in the classroom: one based on a behaviorally anchored model ("long form") and a short, graphic rating form ("short form"). In addition, several classes used the points allocation method developed by Michaelsen (Michaelsen et al., 2004), often in conjunction with one of the rating forms. These methods are described later, and the rating forms are provided in Appendix A and Appendix B. Item ratings and overall scores were collected from 320 juniors and seniors (169 males, 151 females) in 13 classes, across 4 different courses ("Organizational Behavior," "Survey of

Management,” “Managerial Ethics,” and “Human Resource Management”) during a period of 5 years. The number of students evaluated by each method is listed in Table 1.

The long form (Appendix A) consists of seven items that are critical to group success and two items that summarize overall contributions to the group: preparation for the quiz, understanding of quiz concepts, effort on group activities, commitment to high group performance, facilitating group discussion, leadership in initiating and organizing tasks, role in conflict resolution, overall contribution to group task, and overall contribution to group member relationships. The items were selected because they were important given the context of the course (e.g., preparation for the quiz) or because an extensive literature review identified them as among the most critical for team success (Ancona, 1990; Bettenhausen, 1991; Buchholz, 1987; Buller, 1986; Cohen & Ledford, 1994; Cooper, 1975; Cummings, 1981; Davis & Hinsz, 1982; Dyer, 1984; Gist, Locke, & Taylor, 1987; Gladstein, 1984; Goodman, Devadas, & Hughson, 1988; Goodman, Ravlin, & Schminke, 1987; Greenbaum, Kaplan, & Damiano, 1991; Guzzo & Salas, 1995; Hackman, 1987, 1990; Hackman & Morris, 1975; Hare, 1976; Hirokawa, 1980; Kaplan & Greenbaum, 1989; Levine & Moreland, 1990; Manz & Sims, 1987; McGrath & Kravitz, 1982; Shaw, 1981; Shea & Guzzo, 1987; Sundstrom, De Meuse, & Futrell, 1990; Tannenbaum, Beard, & Salas, 1992; Watson, Michaelsen, & Sharp, 1991; Woodman & Sherwood, 1980). Each point on the scale of the first seven items is described using specific behaviors. Descriptions were largely developed intuitively, based on the literature about and experience with groups in the classroom. The last two items are graphic rating scales. All items are assessed using a 4-point scale.

The short form (Appendix B) assesses four items: member's preparation, participation and communication, commitment to high group performance (helps group excel), and cooperation. Each item is described and raters use a 4-point rating scale (*more than 90% of the time, more often than not, less than half of the time, and never or once in a great while*). This form was developed with help from students in human resource management classes. For the past several years, groups of students in human resource management classes were assigned the task of creating a peer assessment form. Each group had to identify basic competencies that were important for their group's success. They described specific behaviors indicating various levels of performance for each competency. The first time this task was assigned, four basic categories emerged from five groups of undergraduate students. These categories were used to create the first draft of the short form. After assigning the same task the following semester to four groups of graduate students in a human resource management class, only minor revisions were made to the original form because the students expressed the need for the

same basic competencies and behaviors. No changes have been made to the form since that revision, because across groups and over time, behaviors related to these four categories consistently emerge. Once the form was finalized, data were collected to assess its effectiveness.

The form for the points allocation method can be found in *Team-Based Learning: A Transformative Use of Small Groups in College Teaching*, by Michaelsen et al. (2004, p. 230). The total number of points allocated to group members is equal to 10 times the number of people being evaluated. Students are told that they have to differentiate among group members based on "the extent to which the other members of your team contributed to your learning and/or your team's performance" (Michaelsen et al., 2004, p. 230). In addition, students cannot simply assign everyone in their group 10 points; at least one member has to be assigned a score of 11 or higher, and one has to be assigned a score of 9 or lower. Typically, average performers in the group or those who perform satisfactorily receive 10 points from each group member. The highest performers in the group usually receive 11 to 15 points from each rater. Poor performers usually receive 7 to 9 points from their peers. For example, in a group of five people, a rater may assign two members 10 points each, the poor performer 8 points, and the high performer 12 points, for a total number of 40 points allocated.

*Which method is best for evaluation?* Lejk and Wyvill (2001) claimed that no matter which rating form one uses for peer evaluation, the outcomes will be the same. Indeed, many of the results from all three instruments were similar. Table 1 provides a summary of the key statistics. Analyses indicate that all three assessment tools had high interrater reliability and were related to an individual's knowledge of course material as measured by the individual quizzes. Leniency effects occurred on both rating forms, with most students using the upper half of the rating scale. Gender differences occurred; on average, women received higher ratings than men on all three methods ( $p < .05$ ). One reason women received higher peer scores for this sample of students may have been because their average quiz scores were 3 points higher than those of the men ( $t = 2.63, p = .009$ ).

An important issue is how peer scores are converted into grades and how grades may differ depending on the peer evaluation method used. As noted earlier, a common approach to calculating a grade is to divide a student's average peer score by the highest average peer score received by a member in his or her group. When this calculation was made for the 98 students who were rated using both the short form and points allocation method, the average score for the short form was 88.8% ( $SD = 13.72$ ), and the average score for the points allocation method was 82.6% ( $SD = 17.10$ ). When calculating a letter grade based on the traditional grading scale, 90% is an "A," 80% is

a "B," and so on, 58 of the participants would have received the same letter grade under either method. If the short form had been used alone, without the points allocation method, 2 participants would have received a letter grade lower, 26 would have received a letter grade higher, 8 would have received two letter grades higher, and 4 would have received three letter grades higher.

A comparison was also calculated to assess grade outcome differences for the long form and points allocation methods ( $n = 52$ ), and the results were similar. The average score for the long form was 89.5% ( $SD = 11.62$ ), and the average for the points allocation method was 82.9% ( $SD = 16.69$ ). Out of those participants who were evaluated with both instruments, 23 would have received the same letter grade. If the long form had been used alone instead of the points allocation method, 3 would have received a letter grade lower, 20 would have received a letter grade higher, 5 would have received two letter grades higher, and 1 would have received three letter grades higher.

As Table 1 reveals, there are few meaningful differences between the short and long rating forms with respect to reliability, relationship to individual performance, and grade outcomes. The two rating forms assess essentially the same type of behaviors (i.e., interpersonal skills, knowledge, and effort; evidently the amount of detail has little influence on average rater responses). If peer assessment is conducted for evaluation purposes only, the statistical evidence indicates that the shorter form is a better choice than the long form simply because it is easier for both students and instructors to use.

When used for evaluation, is a rating form a better choice than the points allocation method? Lejk and Wyvill (2001) preferred their single-item assessment method to a rating instrument for grading because they claimed it supported the goal of working together as a team; individuals are supposed to contribute the best they can in whatever way they can to complete the group project. Some categories listed in a rating form may be less important than others to group performance, and it may be unrealistic to expect group members to be strong in all the categories that are evaluated. Overall contributions may therefore be more important than the ability to get high ratings in every single category.

On the other hand, a points allocation method that forces students to divide a fixed number of points among their teammates may reduce cooperation. If members know they will be competing for points, it reduces their incentive to encourage everyone to fully participate (Bacon et al., 1999). For example, why should members exert any effort following up on a member who misses a group meeting? If one group member fails to participate, there are potentially more points available to be divided among the other members.

Fortunately, there are usually other factors at play that encourage members to work together, such as the amount of work required for a project or the common goal to achieve a good grade. In some cases, students may see

a points allocation method for peer assessment as a threat to their cohesiveness and plot to make sure all members receive the same score. Lejk and Wyvill (2001) suspected that some collusion had occurred among their students when they used a points allocation method.

When students are asked to give an overall score for their teammates' contributions, what member characteristics, behaviors, and outcomes are they thinking about when they make their decision? A correlation analysis revealed that all items in the long form were related with the points allocation score, except the item measuring conflict management ( $r = .03$ ; see Table 2). In addition, the correlation between the points allocation score and an item on the long form, "Facilitating Discussion," was a relatively low  $.37$  ( $p < .01$ ); all other items on the long form were more strongly correlated with the points allocation score including "effort on group activities" ( $r = .78$ ,  $p < .001$ ), "preparation" ( $r = .76$ ), and "leadership" ( $r = .70$ ,  $p < .001$ ). These statistics provide some, albeit weak, evidence that students rely on perceptions of effort, preparation, and more demonstrative components of leadership to assess a group member's contributions. Although effort, preparation, and leadership are important, the ability to help the group work through conflict and facilitate effective group discussion is needed to minimize process losses and increase potential gains from group synergy; interpersonal skills are needed to ensure the strategic use of members' task skills (Hackman, 1987).

Practically speaking, a lack of emphasis on a particular interpersonal skill or two is likely to have little overall impact on most students' grades. Occasionally, however, a student who made a valuable contribution to the group by using effective communication and/or negotiation skills may not get the credit he or she deserves when a single-item assessment is used. When an overall measure is used to determine a student's group participation grade, it may be worth noting in the instructions, at least for the sake of learning, what kind of contributions raters should consider when determining a teammate's score.

*Which method is best for development?* Both the long and short forms can be used for development. If used for developmental purposes only, the longer form includes more detail about the specific behaviors that support and detract from group performance. This form offers feedback on both task and relationship roles and serves to educate students about the behaviors that are important to group success. It can be used during the middle of the semester or before the group project is complete to clarify expectations about appropriate member behavior and to help students identify their strengths and weaknesses as group members.

The short form may also be used for developmental purposes but will not yield as much detail as the long form regarding relationship roles. In particular, the long form provides information about communication and



TABLE 2  
Correlation Matrix for Long Form<sup>a</sup>

	1	2	3	4	5	6	7	8	9	10	11
1. Individual quiz	.51***										
2. Preparation for quizzes	.48***	.75***									
3. Understanding of quiz concepts	.37***	.82***	.66***								
4. Effort on group activities	.50***	.83***	.74***	.83***							
5. Commitment to high performance	.21**	.60***	.50***	.51***	.55***						
6. Facilitating group discussion	.44***	.78***	.72***	.85***	.84***	.56***					
7. Leadership, initiative	.06	.34***	.20*	.24**	.25**	.54***	.21*				
8. Role in resolving group conflict	.48***	.85***	.76***	.83***	.80***	.65***	.84***	.40***			
9. Overall contribution to task	.30***	.68***	.61***	.64***	.64***	.76***	.68***	.59***	.81***		
10. Overall contribution to process	.45***	.90***	.80***	.86***	.88***	.77***	.88***	.49***	.94***	.87***	
11. Long form mean	.42***	.76***	.63***	.78***	.68***	.37**	.70***	.03	.79***	.63***	.78***
12. Points allocation mean											

a.  $n = 151$ , except for points allocation mean (Line 12), when  $n = 52$ .

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

**TABLE 3**  
**Correlation Matrix for Short Form<sup>a</sup>**

	1	2	3	4	5	6
1. Individual quiz						
2. Cooperation	.19**					
3. Commitment to high group performance	.44***	.71***				
4. Preparation and effort	.47***	.66***	.82***			
5. Participation and communication	.36***	.68***	.75***	.72***		
6. Short form mean	.42***	.84***	.93***	.91***	.89***	
7. Points allocation mean	.42***	.71***	.85***	.82***	.81***	.88***

a.  $n = 128$  for all items except points allocation mean (Line 7), when  $n = 98$ .

\* $p < .05$ . \*\* $p < .01$ . \*\*\* $p < .001$ .

conflict resolution that may help those students who tend to dominate discussion or disregard the ideas of others, behaviors that can be extremely frustrating for group members and detrimental to group effectiveness. Although the short form includes an item on communication, it lacks specificity; a low rating could be the result of any number of communication problems. The correlation matrix for the short form indicates a stronger halo effect than for the long form, which also contributes to the ambiguity about which behaviors led to low ratings (see Table 3).

In defense of the short form as a teaching tool, it is worth reiterating that earlier versions of this form were developed with the help of groups of students who were assigned the task of designing a peer assessment instrument. This assignment encouraged students to reflect on the behaviors that are important for effective group performance.

Some have argued that students should develop and use their own peer assessment instruments (see Lejk & Wyvill, 2001). Willcoxson (2006), for example, had groups compose a list of guidelines for behavior before they began work on the group project. At the end of the project, students evaluated each other based on those guidelines. Past experience shows that the guidelines, competencies, and/or behaviors identified by some student groups are incomplete or unclear. If groups are to use their own forms, instructors will need to encourage a round or two of revisions. The process of creating one's own instrument is a useful one, however, because it clarifies for each member what is expected and required. To save time, instructors may provide students with a standard feedback form at the beginning of the semester and encourage groups to discuss changes that they think would improve the form for their own learning goals. This approach clarifies expectations and requirements while building ownership in the peer assessment process.

The points allocation method examined in this study provides students with information about how their peers perceive their contributions relative to others in the group. This method provides little information about specific behaviors. There is no way to know exactly which behaviors or work products were considered when the rater assigned scores to group members. To some degree, this problem may be addressed by encouraging students to provide comments about each member's performance. If an overall score is combined with comments, this method provides information that may be useful for development.

## RECOMMENDATIONS ABOUT IMPLEMENTATION

When peer evaluations are used for development or to inform grading decisions, instructors have an obligation to ensure fairness. No peer evaluation system is free of error; different approaches will yield different results, and there will be variance between raters. There are steps that an instructor can take that may encourage students to view the evaluation process seriously and reduce the effect that bias may have on a student's grade.

Leniency error is exacerbated when raters give all their teammates the same peer scores. For example, Johnston and Miles (2004) found that 26% of student raters did not differentiate ratings among their peers. To avoid this problem, raters should not be allowed to give everyone the same score. Differences may be minor, but in the majority of groups, performance differs between group members and ratings should reflect that.

To offset possible damage done by various forms of bias, the lowest and/or highest ratings can be thrown out before calculating grades. Throwing out the lowest rating for each student protects the student from receiving unfairly low marks that are based on grudges, disputes, personality conflicts, or other irrelevant factors. Unfortunately, the lowest rating may be an accurate reflection of a student's actual performance, and ignoring this bit of data leads to grade inflation. To maintain balance and reduce the effect of an unfairly high rating, the top rating may also be thrown out. Alternatively, the median score could be used instead of the average.

To remind students of the importance of their rating decisions, instructors could place an honor pledge at the bottom of the evaluation instrument that states, "To the best of my recollection and ability, the above ratings accurately reflect the performance of my peers." This pledge hopefully encourages students to take the assessment seriously.

Whether peer assessment is used for development or evaluation reasons, it is important to inform students early in the semester about how they will be assessed. As in the workplace, expectations should be made clear from the start of any group project so that students can be more intentional about their behavior and to remove the anxiety or frustration associated with a surprise evaluation at the end.

## Summary and Conclusion

Given the importance of team skills in almost any organizational setting, it is a good first step for instructors to provide group learning experiences in the classroom. We cannot assume, however, that students will learn how to become better group members simply by participating in group activities. If instructors of management are serious about helping students improve their team skills, feedback about individual behaviors in a group setting is needed. Through peer assessment, instructors can collect information about group member performance and use that information for the purpose of development, evaluation, or both. A review of the literature indicated that numerous methods are currently being used to provide students with feedback and instructors with data for grading decisions. The intent of this article was to examine how instructors have used various peer assessment tools and to highlight the key issues associated with selecting and implementing a peer evaluation process.

Two of the most common methods of peer assessment found in the literature are graphic rating forms and single score methods. The competencies assessed on rating forms vary but generally include items related to attendance, dependability, quality of work, effort, cooperation, managing group conflict, cognitive contributions, and structuring group work. Single score methods typically require students to allocate a fixed number of points to each group member based on his or her overall contribution to the group product. When tested in the classroom, the short rating form, long rating form, and points allocation method demonstrated acceptable levels of reliability and were related to individual performance measures. The rating forms suffered from leniency bias and resulted in higher grades than the points allocation method. The single score method implicitly emphasized the final result; the totality of one's contributions was primary. In contrast, the rating instruments focused on the means to the end; they provided more information about the specific task and relationship behaviors that are needed to maximize group performance.

An instructor has many choices with respect to peer assessment tools and processes. To increase learning and ensure fair grading, decisions about peer assessment should be made intentionally, with a clear understanding of the goals of the course and the objectives of group assignments.

## Appendix A Long Form

### Peer Evaluation Score Sheet (Long Form)

Team # \_\_\_\_\_

Write the name of each group member in the space provided. Refer to the *Categories and Behaviors* handout for a list of the categories that you will use to evaluate each group member. Under each category, 4 sets of behaviors are described. For each group member, you must decide which set of behaviors under each category is most consistent with the behaviors that the member displayed during class. Circle the corresponding letter on the form below. Circle only one letter per category for each member. I recommend that you evaluate all group members on the first category, and then evaluate all group members on the second category, etc., until you have evaluated everyone on all categories. Do not rate yourself. **If you fill this form out correctly, you will receive 5 points on your final.** Please carefully consider your ratings and be honest. Ratings cannot be identical for all members (there must be at least one different rating).

#### *Categories*

<i>Group member name</i>	<i>1</i>	<i>2</i>	<i>3</i>	<i>4</i>	<i>5</i>	<i>6</i>	<i>7</i>	<i>8</i>	<i>9</i>
_____	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d
_____	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d
_____	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d
_____	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d
_____	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d	a b c d

Comments:

*I have carefully considered the above ratings and believe that they fairly reflect each of my team member's contribution to team performance.*

Signature \_\_\_\_\_

### Peer Evaluation: Categories and Behaviors

Use the Peer Evaluation Score Sheet to rate each group member in each. Please consider each category carefully. You will receive 5 points on your final exam if you fill this form out correctly. Please be honest. Do not write on this sheet!

*(continued)*

## Appendix A (continued)

---

### 1. QUIZ PREPARATION. How well prepared was this person for the quizzes?

#### Focus on his/her effort to understand the material.

- a Member was well prepared for all of the quizzes. S/he did the readings and studied the objectives. S/he did everything the group asked. S/he was highly dependable. S/he met group deadlines.
- b Member was prepared for most of the quizzes. S/he did the readings and studied the objectives for most of the quizzes. If your group assigned objectives, s/he usually did the work and got them to everyone in time. There was only one or two times that s/he wasn't fully prepared. S/he met most, but not all group deadlines.
- c Member was well prepared for a couple of the quizzes. Sometimes s/he was well prepared, but not always. If your group assigned objectives, s/he sometimes did a decent job, but not always. You couldn't really depend upon him/her, but some times s/he came through. Sometimes s/he came through late, though.
- d Member was rarely, if ever, prepared for the quizzes. If you assigned objectives, s/he didn't do a good job, and/or handed them in too late to be helpful. S/he may have read a little of the chapters, but if s/he did, s/he had only skimmed them.

### 2. QUIZ UNDERSTANDING. How well did this person understand the material covered on each quiz? Focus on his/her grasp of the material.

- a Member clearly understood most concepts. S/he had insight about the terms and how they applied. Even if s/he didn't get a question right, you could tell s/he had thought about the issues.
- b Member understood a lot of the concepts, but not all. S/he misunderstood a couple of things on every quiz. Although it appeared that s/he had read the material, some of it hadn't really sunk in.
- c Member had trouble understanding a lot of the material. During group discussions, others in the group had to explain some of the important issues to him/her. On a lot of the questions, it was difficult for him/her to contribute to the discussion, because s/he didn't know that much. Or when s/he did contribute, s/he didn't really understand the concepts. Sometimes s/he knew the material, but s/he was "fuzzy" about a lot of the material.
- d Member did not understand most of the material. Based on his/her comments, s/he knew very little of the material. Maybe s/he read it, maybe s/he didn't, but s/he demonstrated very little understanding.

### 3. EFFORT ON GROUP ACTIVITIES. How much effort did the group member exert on behalf of the group on activities other than the quizzes (this includes the appeals process for the quizzes)? Focus on effort.

- a Member helped the group understand what they were supposed to do. S/he dug into his/her textbook and looked through class notes in order to help the group figure out how to answer the questions on each activity or appeal.
- b Member would share some good ideas to help the group with the activity or appeal. S/he didn't always bring his/her book or look through it, although s/he could be coaxed by others to do so. On some days, s/he demonstrated real effort, but not always.
- c Member would often discuss the questions on the activity or appeal with group members but was distracted easily. Sometimes, s/he let the others do whatever they wanted and didn't show much interest in the activity. On some days, s/he would float in and out of the conversation, paying attention only when something struck a chord. S/he relied on the others to do most of the work.

- d Member was usually only remotely interested in the activity or appeal. S/he usually seemed bored with the whole process. Maybe s/he talked with other disinterested group members about unrelated topics, left the room for various reasons, worked on other projects while the group worked on the activity, etc. S/he pretty much let the other group members work on the activity and only occasionally contributed.

**4. COMMITMENT TO PERFORMANCE. How committed was the group member to high performance?**

- a Member set high standards for self and encouraged others to set high standards. When others were willing to settle for mediocre work, s/he encouraged them to push a little harder. His/her work was excellent and s/he met the deadlines group members set. S/he modeled high performance and encouraged it from others, too.
- b Member set high standards for self, but didn't encourage others to set high standards. S/he was willing to go along with whatever standard the group chose. His/her work was on time and of high quality.
- c Member neither encouraged nor discouraged the group to set high standards. His/her work met what was minimally required and s/he didn't push others to do any more or any less. "Whatever" was his/her motto.
- d Member didn't expend much effort in support of group performance. S/he did not do what was asked of him/her or only did the work with a lot of prodding. His/her behavior may have actually encouraged members to accept *lower* levels of performance.

**5. FACILITATING DISCUSSION. How helpful was the group member in facilitating group discussion?**

- a Member made suggestions, shared ideas, asked questions, summarized what others had to say, etc. S/he didn't dominate the discussion, but s/he wasn't silent either. S/he showed genuine interest in what others had to say. S/he was willing to share her ideas, but s/he didn't force them on anyone.
- b Member made good contributions to the group discussion but didn't actively seek to engage everyone in the discussion. S/he was good at sharing ideas or she was good at listening to ideas. Perhaps s/he showed more interest in what certain members had to say and paid less attention to what others had to say. S/he wasn't rude, but perhaps talked a bit too much and/or interrupted others when they talked.
- c Member rarely said anything at all. S/he seemed interested in the discussion and paid attention, but rarely spoke. S/he didn't encourage others to speak either.
- d Member tended to dominate the discussion or was rude and disrespectful. Member had trouble really hearing what others said. S/he was too quick to discount others' ideas. S/he rarely asked about what others were thinking. S/he didn't seem that interested in the opinions of others. Sometimes, s/he had something good to say. However, his/her harsh words were frequently discouraging and harmful to the group.

**6. LEADERSHIP. To what extent was the member a team leader?**

- a Member initiated tasks and made suggestions as to how to proceed. S/he helped resolve disputes within the group. If the group drifted off task, s/he would help steer members back on track. S/he checked on absent or nonperforming members to offer support, encouragement and feedback. S/he cheered the group on when morale was low.
- b Member usually didn't initiate tasks or suggest how to proceed but was a good role model for other members in that s/he worked hard and met his/her responsibilities to the group. Member kept his/her focus on the task and was rarely the cause for the group to get off-track. S/he had a positive effect on group morale.

(continued)

## Appendix A (continued)

---

- c Member had to be asked to do tasks but was usually willing to help the group. S/he was a follower but made some meaningful contributions to the group in this role. Sometimes, s/he was distracted from the group task or hesitant to meet or complete tasks as suggested by other group members.
- d Member was uncooperative and/or apathetic. S/he showed very little interest in group activities or tasks. S/he had to be asked and prodded to do anything.

### **7. CONFLICT RESOLUTION. What role did the member have in creating and resolving group conflict?**

- a Member used excellent communication skills to reduce the likelihood of conflict. When conflict occurred in the group, member helped the members who were in conflict work out the problem. Member helped make conflict productive. When member disagreed with someone, s/he listened carefully to both sides of the argument and recommended ways to resolve differences. S/he worked toward consensus formation and collaboration.
- b Member sometimes got involved in disagreements with other members but tried to remain open to the opinions of others. Sometimes s/he appeared a little agitated with others and may have occasionally pushed his/her ideas a little too hard. With the encouragement of others, s/he eventually would agree to a compromise or talk through his/her concerns until a consensus could be reached.
- c Member had a hard time compromising and/or reaching consensus with those who disagreed with him/her. S/he would frequently become agitated or, alternatively, withdraw from the discussion altogether. Member didn't seek to understand other viewpoints. In some instances, s/he may have made consensus impossible and the most the group could achieve was a compromise. Still, the conflict generally remained friendly and the group was able to use it to clarify the issues involved.
- d Member initiated conflict that was destructive. His/her disagreements escalated into destructive group conflict. Member would not listen to other viewpoints and refused to compromise. The conflict sometimes became personalized and the member made harmful remarks to other members or about other members.

### **8. OVERALL CONTRIBUTION TO GROUP TASK. To what extent do you agree with this statement: This group member consistently made meaningful contributions to group tests and activities?**

- a Strongly agree
- b Agree somewhat
- c Disagree somewhat
- d Strongly disagree

### **9. OVERALL CONTRIBUTION TO GROUP PROCESS. To what extent do you agree with this statement: This group member was important in building group cohesion, maintaining group morale and resolving group conflict?**

- a Strongly agree
  - b Agree somewhat
  - c Disagree somewhat
  - d Strongly disagree
-



## Appendix B

### Peer Evaluation Short Form

Evaluate each member by circling the number that best reflects the extent to which he/she participated, prepared, helped the group excel and was a team player. Use the following ratings:

- 4 Usually (over 90% of the time)      2 Sometimes (less than half the time)  
 3 Frequently (more often than not)    1 Rarely (never or once in a great while)

<b>Preparation</b> Prepared for team meetings; has read course material and understands the issues and subject matter; completes team assignments on time; attends and is on time to team meetings				
<b>Participation &amp; Communication</b> Articulates ideas effectively when speaking or writing; submits papers without grammatical errors; listens to others; encourages others to talk; persuasive when appropriate			↓	
<b>Helps Group Excel</b> Expresses great interest in group success by evaluating ideas and suggestions; initiates problem solving; influences and encourages others to set high standards; doesn't accept just any idea but looks for the best ideas; stays motivated from beginning to end of projects			↓	
<b>Team Player (Cooperation)</b> Knows when to be a leader and a follower; keeps an open mind; compromises when appropriate; can take criticism; respects others		↓		
	↓	↓		
<b>MEMBER NAME</b>	<b>Team Player</b>	<b>Helps Group Excel</b>	<b>Communication</b>	<b>Preparation</b>
	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely
	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely
	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely
	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely	4 usually 3 frequently 2 sometimes 1 rarely

**Honor Pledge:** To the best of my recollection and ability, the above ratings accurately reflect the performance of my peers.

Signature: \_\_\_\_\_

## References

- Ancona, D. G. (1990). Groups in organizations: Extending laboratory models. In C. Hendrick (Ed.), *Annual review of personality and social psychology: Group and intergroup processes* (pp. 207-231). Beverly Hills, CA: Sage.
- Bacon, D. R., Stewart, K. A., & Silver, W. S. (1999). Lessons from the best and worst student team experiences: How a teacher can make the difference. *Journal of Management Education*, 23, 467-488.
- Beatty, J. R., & Haas, R. W. (1996). Using peer evaluations to assess individual performances in group class projects. *Journal of Marketing Education*, 18(2), 17-28.
- Bettenhausen, K. L. (1991). Five years of group research: What we have learned and what needs to be addressed. *Journal of Management*, 17, 345-382.
- Brooks, C. M., & Ammons, J. L. (2003). Free riding in group projects and the effects of timing, frequency and specificity of criteria on peer assessments. *Journal of Education for Business*, 78, 268-272.
- Buchholz, S. (1987). *Creating the high-performance team*. New York: John Wiley & Sons.
- Buller, P. F. (1986). The team building-task performance relation: Some conceptual and methodological refinements. *Group & Organization Studies*, 11, 147-168.
- Cawley, B. D., Keeping, L. M., & Levy, P. E. (1998). Participation in the performance appraisal process and employee reactions: A meta-analytic review of field investigations. *Journal of Applied Psychology*, 83, 615-633.
- Chalupa, M. R., Chen, C. S., & Sormunen-Jones, C. (2000). Reliability and validity of the group member rating form. *The Delta Pi Epsilon Journal*, 42(4), 83-88.
- Chapman, K. J., & Van Auken, S. (2001). Creating positive group project experiences: An examination of the role of the instructor on students' perceptions of group projects. *Journal of Marketing Education*, 23(2), 117-127.
- Cheng, W., & Warren, M. (2000). Making a difference: Using peers to assess individual students' contributions to a group project. *Teaching in Higher Education*, 5, 243-255.
- Clark, G. L. (1989). Peer evaluations: An empirical test of their validity and reliability. *Journal of Marketing Education*, 11(3), 41-58.
- Cohen, S. G., & Ledford, G. E., Jr. (1994). The effectiveness of self-managing teams: A quasi-experiment. *Human Relations*, 47, 13-43.
- Cooke, J. C., Drennan, J. D., & Drennan, P. (1997). Peer evaluation as a real-life learning tool. *The Technology Teacher*, 57(2), 23-27.
- Cooper, C. L. (Ed.). (1975). *Theories of group processes*. London: Wiley.
- Cummings, T. G. (1981). Designing effective work groups. In P. C. Nystrom & W. H. Starbuck (Eds.), *Handbook of organizational design* (Vol. 2, pp. 250-271). London: Oxford University Press.
- Davis, J. H., & Hinsz, V. B. (1982). Current research problems in group performance and group dynamics. In H. Brandstatter, J. H. Davis, & G. Stocker Kreichgauer (Eds.), *Group decision making* (pp. 1-20). London: Academic Press.
- Drexler, J., Beehr, T. A., & Stetz, T. A. (2001). Peer appraisals: Differentiation of individual performance on group tasks. *Human Resource Management*, 40, 333-345.
- Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of development peer appraisals in self-managing work groups. *Journal of Applied Psychology*, 84, 58-74.
- Dyer, J. (1984). Team research and team training: A state-of-the-art review. In F. A. Muckler (Ed.), *Human factors review* (pp. 285-323). Santa Monica, CA: Human Factors Society.
- 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.
- Farh, J., Cannella, A. A., Jr., & Bedeian, A. G. (1991). Peer ratings: The impact of purpose on rating quality and user acceptance. *Group and Organization Studies*, 16, 367-386.

- Fedor, D. B., Bettenhausen, K. L., & Davis, W. (1999). Peer reviews: Employees' dual roles as raters and recipients. *Group and Organization Management, 24*, 92-120.
- Fox, S. B., Zeev, B., & Yinon, Y. (1989). Perceived similarity and accuracy of peer ratings. *Journal of Applied Psychology, 74*, 781-786.
- Gatfield, T. (1999). Examining student satisfaction with group projects and peer assessment. *Assessment and Evaluation in Higher Education, 24*, 365-377.
- Gist, M. E., Locke, E. A., & Taylor, M. S. (1987). Organizational behavior: Group structure, process, and effectiveness. *Journal of Management, 13*, 237-257.
- Gladstein, D. L. (1984). Groups in context: A model of task group effectiveness. *Administrative Science Quarterly, 29*, 499-517.
- Goodman, P. S., Devadas, R., & Hughson, T. L. G. (1988). Groups and productivity: Analyzing the effectiveness of self-managing teams. In J. P. Campbell & R. J. Campbell (Eds.), *Productivity in organizations* (pp. 295-327). San Francisco: Jossey-Bass.
- Goodman, P. S., Ravlin, E. C., & Schminke, M. (1987). Understanding groups in organizations. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior* (Vol. 9, pp. 121-173). Greenwich, CT: JAI Press.
- Greenbaum, H. H., Kaplan, I. T., & Damiano, R. (1991). Organizational group measurement instruments: An integrated survey. *Management Communication Quarterly, 5*, 126-148.
- Greguras, G. J., Robie, C., & Born, M. P. (2001). Applying the social relations model to self and peer evaluations. *Journal of Management Development, 20*, 508-525.
- Gueldenzoph, L. E., & May, G. L. (2002). Collaborative peer evaluation: Best practices for group member assessments. *Business Communication Quarterly, 65*(1), 9-20.
- Guzzo, R. A., & Salas, E. (Eds.). (1995). *Team effectiveness and decision making in organizations*. San Francisco: Jossey-Bass.
- Haas, A. L., Haas, R. W., & Wotruba, T. R. (1998). The use of self-ratings and peer ratings to evaluate performances of student group members. *Journal of Marketing Education, 20*, 200-209.
- Hackman, J. R. (1987). The design of work teams. In J. W. Lorsch (Ed.), *Handbook of organizational behavior* (pp. 315-342). Englewood Cliffs, NJ: Prentice Hall.
- Hackman, J.R. (Ed.). (1990). *Groups that work (and those that don't): Creating conditions for effective teamwork*. San Francisco: Jossey-Bass.
- Hackman, J. R., & Morris, C. G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (pp. 45-99). New York: Academic Press.
- Halfhill, T. R., & Nielsen, T. M. (2007). Quantifying the "softer side" of management education: An example using teamwork competencies. *Journal of Management Education, 31*, 64-80.
- Hare, A. P. (1976). *Handbook of small group research* (2nd ed.). New York: Free Press.
- Harrell, A., & Wright, A. (1990). Empirical evidence on the validity and reliability of behaviorally anchored rating scales for auditors. *Auditing, 9*, 134-149.
- Hedge, J. W., Borman, W. C., Bruskiwicz, K. T., & Bourne, M. J. (2004). The development of an integrated performance category system for supervisory jobs in the U.S. Navy. *Military Psychology, 16*, 231-244.
- Hirokawa, R. Y. (1980). A comparative analysis of communication patterns within effective and ineffective decision-making groups. *Communication Monographs, 47*, 312-321.
- Johnson, C. B., & Smith, F. I. (1997). Assessment of a complex peer evaluation instrument for team learning and group processes. *Accounting Education, 2*(1), 21-41.
- Johnston, L., & Miles, L. (2004). Assessing contributions to group assignments. *Assessment and Evaluation in Higher Education, 29*, 751-768.
- Kane, J. S., & Lawler, E. E. (1978). Methods of peer assessment. *Psychological Bulletin, 85*, 555-586.
- Kaplan, I. T., & Greenbaum, H. H. (1989). Measuring work-group effectiveness: A comparison of 3 instruments. *Management Communication Quarterly, 2*, 424-448.
- Keeping, L. M. (2003). Self-ratings and appraisal reactions: The effects of context and individual differences. *Academy of Management Proceedings*, 11-16.

- Kingstrom, P. O., & Bass, A. R. (1981). A critical analysis of studies comparing behaviorally anchored rating scales (BARS) and other rating formats. *Personnel Psychology*, *34*, 263-289.
- 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*, 61-72.
- Levine, J. M., & Moreland, R. L. (1990). Progress in small group research. *Annual Review of Psychology*, *41*, 585-634.
- Manz, S. (1987). Leading workers to lead themselves. *Administrative Science Quarterly*, *32*, 106-129.
- McEvoy, G. M., & Buller, P. F. (1987). User acceptance of peer appraisals in an industrial setting. *Personnel Psychology*, *40*, 785-789.
- McGrath, J. E., & Kravitz, D. A. (1982). Group research. *Annual Review of Psychology*, *33*, 195-230.
- Meyer, H. H. (1991). A solution to the performance appraisal feedback enigma. *Academy of Management Executive*, *5*, 68-76.
- Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2004). *Team-based learning: A transformation use of small groups in college teaching*. Sterling, VA: Stylus Publishing.
- Mumford, M. D. (1983). Social comparison theory and the evaluation of peer evaluations: A review and some applied implications. *Personnel Psychology*, *36*, 867-881.
- Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational and goals-based perspectives*. Thousand Oaks, CA: Sage.
- Paswan, A. K., & Gollakota, K. (2004). Dimensions of peer evaluation, overall satisfaction, and overall evaluation: An investigation in a group task environment. *Journal of Education for Business*, *79*, 225-231.
- Persons, O. S. (1998). Factors influencing students' peer evaluation in cooperative learning. *Journal of Education for Business*, *73*, 225-229.
- Pounder, J. S. (2002). Public accountability in Hong Kong higher education. *International Journal of Public Sector Management*, *15*, 458-475.
- Rafiq, Y., & Fullerton, H. (1996). Peer assessment of group projects in civil engineering. *Assessment and Evaluation in Higher Education*, *21*(1), 69-82.
- Ramus, C. A., & Steger, U. (2000). The roles of supervisory support and behaviors and environmental policy in employee "Ecoinitiatives" at leading-edge European companies. *Academy of Management Journal*, *43*, 605-627.
- Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, *78*, 450-462.
- Shaw, M. E. (1981). *Group dynamics*. New York: McGraw-Hill Book Company.
- Shea, G. P., & Guzzo, R. A. (1987). Groups as human resources. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 5, pp. 323-356). Greenwich, CT: JAI Press.
- Solomon, R. J., & Hoffman, R. C. (1991). Evaluating the teaching of business administration: A comparison of two methods. *Journal of Education for Business*, *66*, 360-365.
- Strom, P. S., & Strom, R. D. (2002). Overcoming limitations of cooperative learning among community college students. *Community College Journal of Research and Practice*, *26*, 315-331.
- Sundstrom, E., De Meuse, K. P., & Futrell, D. (1990). Work teams: Applications and effectiveness. *American Psychologist*, *45*, 120-133.
- Tannenbaum, S. I., Beard, R. I., & Salas, E. (1992). Team building and its influence on team effectiveness: An examination of conceptual and empirical developments. In K. Kelley (Ed.), *Issues, theory, and research in industrial/organizational psychology* (pp. 117-153). Amsterdam, the Netherlands: Elsevier.
- Tziner, A., Joanis, C., & Murphy, K. R. (2000). A comparison of three methods of performance appraisal with regard to goal properties, goal perception, and rate satisfaction. *Group & Organization Management*, *25*, 175-191.

- Watson, W., Michaelsen, L. K., & Sharp, W. (1991). Member competence, group interaction and group decision making: A longitudinal study. *Journal of Applied Psychology, 76*, 803-809.
- Whetten, D. A., & Cameron, K. S. (2002). *Developing management skills* (5th ed.). Upper Saddle River, NJ: Prentice Hall.
- Willcoxson, L. E. (2006). It's not fair! Assessing the dynamics and resourcing of teamwork. *Journal of Management Education, 30*, 798-808.
- Woodman, R. W., & Sherwood, J. J. (1980). The role of team development in organizational effectiveness: A critical review. *Psychological Bulletin, 88*, 166-186.