JOURNAL OF APPLIED SOCIAL PSYCHOLOGY

Elements Influencing Peer Evaluation: An Examination of Individual Characteristics, Academic Performance, and Collaborative Processes

Warren E. Watson[1] and Anat BarNir
Department of Management
University of North Texas

Robert Pavur
Department of Information Technology and Decision Sciences
University of North Texas

Collaborative learning has existed for some time, and instructional methods have been developed, such as cooperative learning, team learning, and problem-based learning. The purpose of this study is to observe students in learning teams and how their observations of their team and team members are reflected in peer evaluations. We examine pre- and post-team history elements, such as basic demographics, personality factors, exam scores, and observations of team process and team goal attainment. Pre- and post-team history factors correlate with peer evaluations and are discussed. In our analyses, we take into account intraclass correlations and individual and group effects. Suggestions are given for future research.

Collaborative learning is widely applied in academia and business and crosses many content disciplines (Brooks & Ammons, 2003; Sivan, 2000). During the last 25 years, approaches to team instructional methods have been developed, such as cooperative learning (Chasnoff, 1979; Johnson & Johnson, 1994), team learning (Watson, Michaelsen, & Sharp, 1991), and problem-based learning (Savin-Baden, 2003). The use of team projects in the classroom and in business has long been popular, but over the last two decades, the application of a more organized team-learning format across the entire semester or training period has emerged. No longer do we have students and trainees only participate in teams for interpersonal exercises or temporary projects, but we increasingly apply the team format throughout the semester or training program, with participants' contributions being assessed to a significant degree by how they produce on team projects and team exams. In these instructional methodologies, versus the more traditional individual orientation, participants are required to work together regularly for a longer period of time and produce evaluated team outcomes that have an impact on each individual's final assessment (Michaelsen & Watson, 1993).

[1]Correspondence concerning this article should be addressed to Warren E. Watson, Department of Management, University of North Texas, P. O. Box 305429, Denton, TX 76203. E-mail: watson@unt.edu

Journal of Applied Social Psychology, 2010, **40**, 12, pp. 2995–3019.

In collaborative education, students and employees learn to communicate better and develop team skills that are important in most organizations. Content acquisition is enhanced by team application of course objectives, combined with individual work (Watson et al., 1991). It is becoming difficult today to find classes and training programs that do not use collaborative learning to a significant degree. Individual and team performance are combined to assess individuals' progress in courses and training programs. Team performance on exams, cases, and projects become a percentage of each individual's assessment. Along with this emphasis on team activities (e.g., team exams, projects) comes the task of evaluating team member contributions to the team. A commonly applied assessment of team member effort is the peer evaluation (Blackmore, 2005; Falchikov & Goldfinch, 2000), and the thrust of the present research is an examination of elements that impact peer evaluations that team members receive.

Peer Evaluation

The application of peer evaluation is a technique for assessing differential member contributions to a team. The practice of 360-degree feedback has been used for some time in business, and certainly peer feedback is accepted as an important information source in academia. *Peer assessment* involves the observations of peers who are in the closest contact with and the most familiar with team members' actions (Barrett, 1996; Cederblom & Lounsbury, 1980). Peers often make finer distinctions about team member performance than do instructors and supervisors (DeNisi, Randolph, & Blencoe, 1982). Another benefit of peer assessment is that when team members know that they will be evaluated by peers, they tend to perform better than teams who will not receive such an evaluation (Ward, 2005). The other side of this coin is the issue of evaluation accuracy in peer evaluations.

One concern with peer evaluation is inflated ratings (Ilgen & Feldman, 1983; Landy & Farr, 1983). A team member may give an evaluation that is intentionally better or worse than the actual perceived contribution of the evaluated individual. This tendency stems from peer pressure and the fact that grades are affected by the evaluation (Murphy & Cleveland, 1991). If the peer evaluation is administered confidentially and does not include self-evaluations, the inflationary effect is mitigated and tends to be an accurate assessment (Lejk & Wyvill, 2001). Research on peer evaluation has examined the reliability and validity of peer ratings (Beatty, Hass, & Seiglimpaglia, 1996; Morahan-Martin, 1996), the bias of peer ratings (Ghorpade & Lackritz, 2001), and the development of an evaluation instrument (Johnson &

Smith, 1997; Levi & Cadiz, 1998). Very little research has examined what team member characteristics and what team activities affect peer evaluations.

Peer assessment has examined validity and reliability, instrument development, and evaluation bias. Ante- and post-team experience were time elements used to examine characteristics that may affect an individual's peer evaluation (Persons, 1998). Persons classified ante factors as those that exist before the implementation of collaborative learning, such as gender and ethnicity. Post factors are those that are occurring across the implementation of collaborative learning, such as exam scores and participation in teams. Students' grade point average (GPA) and team participation correlated in their peer evaluations. This is one of the few studies to address what we term *pre-team* and *post-team history factors*. We follow that investigation with similar and additional variables to understand better what influences peer evaluations.

The Peer Rating Process

Conditions that are conducive to using peer assessments are those in which team members have unique views of each other's behaviors, team members are capable of making judgments about each other, and the issue of team effectiveness is relevant (Kane & Lawler, 1978). Three common types of peer assessment are peer nomination, peer rating, and peer ranking. Kane and Lawler explained that peer nomination is best for noting the best and least performing members, peer ranking is best for discriminating all members of the group from one another, and peer ratings work best for assessing the contribution of each member. The peer rating procedure is our choice for examining team member reports of each other's contributions.

Peer evaluations tend to be more accurate when the teams are involved in feedback regarding individuals' performance across time, with multiple observations and frequent process communication (Brooks & Ammons, 2003). Utilizing peer feedback early on familiarizes team members more with the process and enhances accuracy (Dominick, Reilly, & McGourty, 1997). In addition, providing feedback to team members at multiple points allows members the opportunity to improve. When the final peer evaluation is completed at the end of the review period, team members are acquainted with the process and more clearly report observations of peers' contributions to the team. Once understanding and practice with peer feedback has been established, a peer assessment that requires a global judgment based on well described criteria works best (Falchikov & Goldfinch, 2000).

Little research has addressed factors that affect peer evaluations, which include factors before the team collaboration and across that collaboration.

We anticipate that peer evaluations accurately reflect member contributions to the team during the review period. Certainly, one issue is validity, but also important is the identification of individual elements that influence the ratings. Peer ratings often are used in determining student grades and employee performance reviews. This research will examine pre- and post-team history elements that may affect peer evaluation.

Research Focus

Teams are prevalent in today's organizations (Devine, Clayton, Philips, Dunford, & Melner, 1999), both in the educational environment and in product and service producing operations (Watson, BarNir, & Pavur, 2005). The use of teams can be a mixed blessing, with synergistic output resulting in the whole being greater and better than the sum of the individuals, and on the other hand at times demonstrating process loss and social loafing (Strong & Anderson, 1990). How a team actually blends individual member inputs into team-level outputs is unclear (Stewart, Fulmer, & Barrick, 2005). Nevertheless, a collective action comes from individual actions (Kozlowski & Klein, 2000), and individual contributions to the team are one interest we have regarding peer evaluations.

We are interested in individual characteristics that team members possess when they join a team and begin working as team members, such as gender, ethnicity, age, and personality traits. We refer to those characteristics as *pre-team history*. We are also interested in the characteristics that are developed during the collaboration, such as exam scores (individual and team) and member observations of team process and goal attainment. We refer to those factors as *post-team history*. These factors are important for examining the overall peer-evaluation picture and for better understanding what influences team members' evaluations of their peers. Even though many reliability and validity questions about peer evaluations have been researched, very little research has explored what factors affect a member's peer evaluation.

Hypotheses

Pre-Team History

Individual elements that members bring to a team are gender, ethnicity, age, and personality traits. We refer to these as *pre-team history* elements since they existed before the team was formed but are important characteristics influencing an individual's views. Research on gender bias in peer evaluations has primarily focused on inter-gender evaluations and has shown

no main effects (Correll, 2004; Falchikov & Magin, 1997). Ethnicity may be a different issue. When teams are comprised of members from different ethnic backgrounds, different processes may occur that lead to peer evaluations being affected by the ethnic background of the member.

Specifically, social psychological explanations suggest that team members belonging to the ethnic group that is the majority in the team may end up with a higher evaluation than individuals from the ethnic groups that are the minority. This may happen for several reasons. First, it may be associated with latent racial bias. Such a bias will reflect favorable attitudes toward majority groups, which manifest into behaviors that favor the majority group (Chen & Bargh, 1997; Dovidio, Kawakami, & Gaertner, 2002; Fazio, 1990). Second, it may reflect the *other-race effect*, which argues that recognition and memorization are better for faces of one's own race than for a different race (Bothwell, Brigham, & Malpass, 1989; Lindsay, Jack, & Christian, 1991). Finally, to the extent that a larger proportion of the team members are Caucasians, an overall more favorable rating may be given to those individuals simply because their actions were more representative (Schneider, Hastorf, & Ellsworth, 1979), as compared to non-Whites. We propose the following:

> *Hypothesis 1*. Individuals belonging to the ethnic majority in a team will receive higher peer evaluations than will individuals of other ethnicities.

Personality traits affect one's overall behavior and personal style, and our focus in the present work is on those traits that likely have direct influence on interpersonal interaction and task behaviors in teams. One such trait is communal orientation. *Communal traits* indicate characteristics that describe an individual's tendency to be helpful toward others, which is expected to be beneficial within a team. In order to achieve the synergy for a team to be better than the sum of its individuals, a high level of communal traits can be important. A communal orientation is associated with increased attention to each other's needs, and increased responsiveness to others (Clark, 1986). As such, individuals with greater communal orientation not only will help others more on general needs and tasks, but will also be more attentive to each other's emotions (Clark, Ouellette, Powell, & Milberg, 1987). We expect that team members will value individuals with high communal tendencies and reward them through positive peer evaluation.

Internal locus of control is a second trait that is expected to influence behavior in teams, as it presents the can-do attitude and represents a style of individual contribution to the team through encouragement of goal pursuit and willingness to make personal effort to move the team forward (Spector,

Cooper, Sanchez, & O'Driscoll, 2002). Research has shown the importance of not only perceptions of control in a work environment, but also of a person's more general beliefs about control. Levels of locus of control are the degree to which individuals believe in personal control in life (i.e., internality), rather than in control by outside forces or individuals (i.e., externality). It has been noted that internal control beliefs also are an important component of emotional adjustment and stress management (e.g., Kobasa, Maddi, & Kahn, 1982; Spector, 1982). Internal locus of control would be evident when individuals contribute their input (e.g., for delegated tasks). We hypothesize that individuals with high levels of internal locus of control will have a greater sense of ownership of the work situation (Spector, 1988), will display more of a can-do attitude, will tend to take on more tasks, and will have a general positive impact on the overall team spirit, and thus will be valued and recognized by other members and will receive higher evaluations. We offer the following hypotheses:

> *Hypothesis 2a*. Team members with higher levels of communality will receive higher peer evaluations.
>
> *Hypothesis 2b*. Team members with higher levels of internal locus of control will receive higher peer evaluations.

Post-Team History

Post-team history elements come from the collaboration and results achieved across a team's life cycle. Individual scores may influence the attributions of peers. Those receiving higher scores may receive higher peer evaluations as a result of their presumed greater preparation. Individual abilities have been studied in the context of the effect on team performance, on engaging in novel and complex tasks, and on adaptation (Harrigan & Wigdor, 1989; Hunter, 1986; LePine, 2003; LePine, Colquitt, & Erez, 2000). Interestingly, the majority of work has focused on team members' responses to low-ability individuals, rather than high-ability individuals. For the most part, studies have focused on the reasons behind team members' responses to poor performers, finding those reasons to vary from the personality of the poor performers, attributions regarding the causal stability of the poor performance, or the perceived controllability of poor performance (LePine & van Dyne, 2001; Struthers, Weiner, & Alfred, 1998; Taggart & Neubert, 2004; Weiner, 2000). Our assumption is that students with better individual performance scores are able to transfer their abilities to the team effort, and the team acknowledges this contribution.

Hypothesis 3. Team members with higher scores on individual exams will receive higher peer evaluations.

Also, the extent to which individual team members observe their team working well together may influence the evaluations they give. We argue that individuals who perceive and report their teams to work well together will end up with higher evaluations from peers on their teams. Therefore, it may be that team members who report that their teams work well together show a more positive attitude toward teamwork and have a higher level of social or practical intelligence (Sternberg & Wagner, 1986; Thorndike, 1920), which demonstrates an improved ability to understand and empathize with others, to manage people, and to react to social situations in a manner that is appreciated by others. In effect, it may be that those members with a higher level of relational skills or emotional intelligence tend to view the team context in a positive manner; to behave in a manner that enhances cohesiveness and positive attitudes among all members; and, consequently, to earn the appreciation of other members manifested through improved evaluations. Support for this logic was provided by Offermann, Bailey, Vasilopoulos, Seal, and Sass (2004), who found a positive relationship between the level of emotional intelligence of individual members and general attitudes toward the team. The same may be said for individuals who report that the team allows them to achieve their personal goals.

Hypothesis 4a. Team members reporting higher levels of positive teamwork will receive higher peer evaluations.

Individuals reporting that their teams are working well together may experience a greater degree of personal goal attainment. In other words, it may be that those individuals consider the team goals and their individual goals to be complementary; thus, they perceive the effort they make in collaborating to also promote their personal goals. In such instances, the members will have greater motivation to contribute to the team—as compared to members whose personal goals are inconsistent with the team goals—because this contribution is perceived to serve both the team and the personal goal. Greater contributions can be manifested in a variety of ways, ranging from performance on specific tasks, commitment to the team (which is perceived as contributing to the attainment of the personal goals), or providing support through building of cohesion and team spirit. Regardless of the contribution, the result will be more positive peer evaluations to such individuals. Therefore, we propose the following:

Hypothesis 4b. Team members reporting greater goal attainment with their team will receive higher peer evaluations.

Method

Participants

The sample consisted of 287 participants (148 males, 139 females) who were enrolled in principles of management courses at a large university in the southwestern United States. Participants' mean age was 23.4 years ($SD = 4.1$), and their mean full-time work experience was 4.2 years ($SD = 5.0$). The ethnicity breakdown was 75% Caucasian, 8% African American, 10% Asian, and 5% Hispanic. The remaining ethnicities were a small number of American Indians and other ethnicities.

Each participant was a member of one of 71 learning teams. Each team consisted of 4 or 5 members, and membership remained constant throughout the semester. During this time, the team members had frequent opportunities to interact as they engaged in a wide variety of team activities.

Procedure

All participants completed a demographic information form and a personality survey during the first class period. The demographic information gathered was used to divide participants into small teams balanced by gender. The participants were then assigned to their particular teams during the second class period. Data were collected from three individual and team exams that were spaced at 5-week intervals throughout the semester.

The exams were administered first to each individual, which we consider to be indicative of their academic ability. The individual answer sheets were collected, and the same exam was administered to each team, which they answered by having face-to-face discussions. After each exam period in which individuals and then teams completed the exams, the teams would record the scores on a spreadsheet provided to them, which they maintained for the semester. The instructor would give a folder with a spreadsheet to each team at the beginning of the class, and the team would return the folder at the end of each class. Therefore, after each exam, team members had access to the team folder during class time and observed all team members' scores and their team score. Only the members of a specific team and the instructor had access to the folder.

To be successful, teams had to work face to face and integrate their results. The exams were 50 multiple-choice items taken from the test bank of a well known management textbook. The grades received on the three individual exams contributed 70% to each member's grade for the course, and the three team exams accounted for another 20% of a student's grade. Therefore,

teamwork and preparation for the team were very important to all students. Instruments that were administered regarding team issues were given after the team assignments were turned in and prior to receiving evaluation for each segment.

Feedback on performance and processing team interaction was performed throughout the semester. During the formation of the teams during the first week of class, communication from each team member with each other was organized within each team. During the second week of class, the instructor led each team in having each member describe what is an effective team and what is an ineffective team. Then, each team decided the same issues. These findings were discussed with emphasis on how each team should use this information to enhance their performance. Common issues were communication, preparation, and leadership. During the next week of class, a list of traditional team-effectiveness characteristics were presented and discussed. Teams compared the traditional elements with what they had discovered earlier from their own observations. The instructor led team discussion of these commonalities and the new ideas from the traditional items.

After every team exam and team activity, team performance and team interaction were processed regarding what they did well and what they could do better. Two times during the semester each team member completed the Group Style Instrument (Watson, Johnson, & Merritt, 1998), which described specific synergy behaviors and specific individually oriented behaviors. Team members would then discuss their observations within the team and how they may improve. At the beginning of the semester, the instructor also discussed the value of leadership in teams and how it could manifest differently for each team. This was repeated on at least two occasions during the semester. Therefore, there were regular formal and conversational discussions and descriptions regarding the evaluation of the team and team members' contributions and what could be improved. This is suggested when the global type of final peer evaluation that we used is applied (Falchikov & Goldfinch, 2000). The confidential peer evaluation on which members rated each other but not themselves was administered to each member near the end of the semester. The average of these evaluations for each member within a team was a member's peer score. This peer score accounted for 10% of a member's total grade.

Measures

The dependent variable is an individual's peer evaluation averaged from team members in their respective teams. Ipsitive and other ranking

procedures have shown validity problems (Ward, 2005). A numerical rating format has been used on a standard scale of several behaviorally descriptive items, either a Likert format or behaviorally anchored ratings. A problem with this approach is that on a 1-to-5 scale, everyone could be evaluated a 5 or a similar evaluation bias.

In our review of peer-evaluation approaches, we discussed how frequent and regular feedback sessions that described teams' performance and the facets of effective interaction would best be followed by a singular, global evaluation of peers (Falchikov & Goldfinch, 2000). Therefore, we used a format that uses a set number of points based on team size, so there are only so many points to go around. Each member has the same fixed number of points that they distribute among their members according to their observations of the members' contributions to the team. As described previously, several discussions and team activities focused on member quality contributions to team collaboration. The formula for determining a team's peer points is

$$(N-1)\times 10$$

For instance, if there are 5 members on a team, the team peer points would be 40. Team members do not evaluate themselves, which is the reason for $N-1$. By this method, each team has the same standard number of points with which to rate each other. This evaluation is one of the rating type approaches described by Kane and Lawler (1978) and is a widely used peer-evaluation method (Michaelsen, Knight, & Fink, 2004). On the last class day, a confidential peer evaluation was administered to each team. Team members did not rate themselves. These scores were then averaged within each team to give each member his or her peer-evaluation score (Michaelsen & Watson, 1993).

Ethnicity was measured by team members indicating their ethnic background on a demographic questionnaire. The ethnic breakdown for the classes was 79% Caucasian, 7% African American, 7% Mexican American, and 7% Asian American. The ethnic variable in the cultural model was coded as 1 for Caucasians and 0 for the combined non-Caucasians. Gender and age data were also gathered on the demographic form.

We used personality measures of internal locus of control and communality. We feel that these two measures are descriptive of key behaviors in a team setting. Internal locus of control is the can-do perspective that is important for individual contributions. Locus of control was operationalized using the Work Locus of Control Scale (WLCS) developed by Spector (1988). The WLCS is a 16-item instrument with one half of the statements worded for internal control and one half worded for external control. The items were

rated on a 6-point Likert-type scale. The external control items were recoded in order for all items to be an internal orientation.

Spector (1988) reported on the validation of his instrument based on six different samples. Coefficient alpha measures ranged from .75 to .85. Evidence of the scale's convergent validity was demonstrated by correlating the WLCS with other general locus of control scales. *Communality* is the effort given to assist others, which is critical for a team to be successful. Our communal orientation instrument measures the tendency for an individual to help others. This instrument was the Communal Orientation Scale (Clark et al., 1987), which is a 10-item, 5-point Likert-type scale that has shown a reliability of .80.

Reports of goal attainment satisfaction were measured with four items that were taken from a cohesiveness instrument (Watson et al., 1991) describing the extent to which the individual would like to be on another team like this, whether the team enables a person to reach his or her goals, whether the team emphasizes goals, and whether the team members exchange ideas about goals. The scale demonstrated a reliability of .84.

Effective team process behaviors were measured (Watson et al., 1998) with three items describing team members' actions regarding planning effectively, inclusion of every member in the team's work, and following through on responsibilities. The scale demonstrated a reliability of .85. These scores were obtained following each of the three individual and team exams. Individual exams and team exams were described previously. Scores for these exams were used in the post-history analysis.

Results

Analyses

Descriptive statistics and intercorrelations for ordinal and continuous data variables are presented in Table 1 for individual-level data. The unit of analysis is an individual team member. Since individuals in the same group will be more similar (or dissimilar) than will individuals in different groups, a problem of nonindependence of observations may arise. The nonindependence of the individual data values, which can be measured by computing an intraclass correlation (ICC), will affect the significance of predictor variables if a model does not account for this violation of the usual assumption of independent outcomes (Kashy & Kenny, 2000; Kenny, Mannetti, Pierro, Livi, & Kashy, 2002). We used Kenny et al.'s (2002) revised multilevel approach. Kenny et al. recommend using a mixed-model procedure in SAS to incorporate the APIM into multilevel modeling (MLM). We implemented

Table 1

Descriptive Statistics for Ordinal Variables

Variable	*M*	*SD*	1	2	3	4	5	6
1. Age	23.41	4.38	—					
2. Internal locus	66.79	8.33	.06	—				
3. Communality	39.59	5.28	.09	.35	—			
4. Team process	12.62	2.54	.09	.13	.14	—		
5. Goal alignment	16.71	3.44	.08	.14	.17	.68	—	
6. Individual exams	69.27	8.82	.17	−.01	.03	.18	−.12	—
7. Peer evaluations	10.13	1.44	.19	.15	.03	.23	−.07	.12

this MLM strategy using the SAS code that Kenny et al. provided. This MLM approach incorporates positive as well as negative nonindependence, thus taking into account the ICC within groups, and models the group effects as the APIM partner effects.

Several steps were taken to verify the accuracy of our findings, given the possible nonindependence problem. First, the ICC for the dependent variable (peer evaluation) was computed ($r = -.096$). In addition, we calculated the ICCs for the individual variables measuring pre-team history: age, $r = .11$; gender, $r = .07$; ethnicity, $r = .08$; internal locus of control, $r = .01$; and communality, $r = .05$. Fisher's Z transformation is one approach mentioned by Kenny et al. (2002) to test whether the ICC differs significantly from 0. The ICC for peer evaluation was significant at $p < .01$, and the ICC for age, gender, and ethnicity were significant at $p < .05$. According to Kenny et al., a significant correlation for the dependent variable is evidence of nonindependence. Kenny et al. stated that a positive or negative sign of the ICC of an independent variable indicates whether the independent variable is more similar to a between-group variable or to a within-group variable, respectively. An ICC near 0 indicates a relative balance of the between- and within-group aspects of an independent variable.

Second, we used Kenny et al.'s (2002) revised MLM strategy using PROC MIXED within SAS that accounts for nonindependence and also accounts for the effect of predictor variables, which is referred to as the *actor effect*, and the average of those predictor variables excluding the individual's value for that individual whose outcomes are in the model, which is referred to as the *partner effect*. Consistent with the notation in Kenny et al., the term *mean prime* of a variable refers to the group average of a variable, excluding the

Table 2

Summary of Maximum Likelihood Regression Results for Effects of Pre-Team History Variables

Variable	Estimate	*SE*	*t*	*p*[a]
Intercept	8.98	1.42	6.30	.0001
Age	.06	.02	3.05	.0026
Ethnicity	.29	.21	1.40	.16
Gender	.07	.17	0.40	.69
Internal locus	.02	.01	2.51	.01
Communality	−.02	.02	−1.37	.17
Age mean prime	−.03	.03	−1.11	.27
Gender mean prime	.002	.26	0.01	.99
Ethnicity mean prime	.14	.30	0.48	.63
Communality mean prime	−.003	.03	−0.11	.91
Internal locus mean prime	−.002	.01	−0.14	.89

Note. $\chi^2(10, N = 270) = 33.72, p < .01$.
[a]Two-tailed significance.

individual's value that is included in the model as a predictor variable. The mean prime values are entered as control variables to denote the partner effect (Tables 2 and 3) in the individual-level analyses. The pre-team history, individual-level data hypotheses were tested using this model, with the dependent variable being individual peer evaluations and the independent variables being ethnicity, communality, and internal locus of control. The control variables were age, gender, and the mean prime variables.

The analysis was conducted using maximum likelihood estimation for the mixed model specified in PROC MIXED within SAS to accommodate nonindependence by allowing for the specification of a compound symmetry covariance structure for the group data, thus accounting for ICC (Kenny et al., 2002). This multilevel modeling approach, described by Kenny et al. as "a simpler strategy that accommodates negative nonindependence and models the group effect as the APIM partner effect" (p. 133), treats the individual scores as repeated measures in a group with the covariance of pairs of group members being equal. The coefficients of the predictor variables and their significance levels are presented in Table 2. None of the mean prime variables were significant for this model.

Table 3

Summary of Maximum Likelihood Regression Results for Effects of Post-Team History Variables

Variable	Estimate	*SE*	*t*	p^a
Intercept	12.59	1.66	10.80	.001
Individual exams	.02	.09	1.68	.09
Goal alignment	.02	.03	0.60	.55
Team process	–.11	.04	–2.49	.01
Individual exams mean prime	–.03	.01	–2.48	.01
Goal alignment mean prime	.09	.04	2.18	.03
Team process mean prime	–.14	.06	–2.34	.02

Note. $\chi^2(6, N = 225) = 265.00, p < .01$.
[a]Two-tailed significance.

Kenny, Kashy, and Cook (2006) described a chi-square difference test that can be used to test for the difference between two nested models fitted by maximum likelihood estimation. To test the significance of an overall model fitted by a maximum likelihood approach, a chi-square statistic equal to the difference in the -2LogLikelihood statistic for the null and complete model was used. This difference yielded a significant overall model, $\chi^2(10, N = 270) = 33.72, p < .001$. The results of the PROC MIXED analysis revealed support for the proposed hypotheses. Age was positively correlated with peer evaluation, $t(194) = 3.05, p < .01$. Gender showed no significant relationship.

Hypothesis 1, stating that the ethnic majority in a team (Caucasian students) would receive higher peer evaluations, was supported at the 10% significance level, $t(194) = 1.40, p = .08$. For Hypothesis 2a, there was support for students with lower levels of communality receiving higher peer evaluations at the 10% significance level, $t(194) = –1.37, p = .08$, which is significant, but in the opposite direction of the prediction. Hypothesis 2b was supported since students with higher levels of internal locus of control received higher peer evaluations, $t(194) = 2.51, p < .01$. None of the mean prime control variables were significant.

The same multilevel modeling strategy that was used to model the individual pre-team history variables was used in modeling the individual post-team history variables to predict peer evaluations. Kenny et al.'s (2002) mixed model approach using PROC MIXED within SAS including both actor and partner variables was used to test the post-team history,

Table 4

Summary of Maximum Likelihood Regression Results for Effects of Post-team History Individual Level Variables

Variable	Estimate	Standard Error	t-Value	Sig.*
Intercept	12.59	1.66	10.8	.001
Team Process Mean Prime	–.14	.06	–2.34	.02
Goal Alignment Mean Prime	.09	.04	2.18	.03
Individual Exams Mean Prime	–.03	.01	–2.48	.01
Individual Exams	.02	.09	1.68	.09
Goal Alignment	.02	.03	.6	.55
Team Process	–.11	.04	–2.49	.01

$\chi^2 = 265.00$, $p < .01$.
*Two-tailed significance.

individual-level data hypotheses. These hypotheses tested a relationship between the dependent variable individual peer evaluations and the independent variables; namely, individual exam scores and member observations of team process effectiveness and of satisfaction with goal attainment. Since the overall model was fitted using a maximum likelihood approach, a chi-square statistic equal to the difference in the -2LogLikelihood statistic (Kenny et al., 2006) for the null and complete model was computed, and the overall model presented in Table 4 was significant, $\chi^2(6,\ N = 225) = 263.37$, $p < .001$. The post-team variables and their ICCs were as follows: individual exams, $r = .17$; team process, $r = .30$; and goal attainment, $r = .19$. All were significant at the 5% significance level. In contrast to the pre-team analysis, the ICCs and mean prime variables (partner effects) were significant for the post-team analysis. These results appear reasonable since teams are not expected to influence the pre-team variables, but are expected to have an influence on the post-team variables and, consequently, the mean prime variables for the post-team analysis affect the prediction of peer evaluations. These mean prime variables act as significant predictor control variables that allow for a more accurate interpretation of the hypotheses.

As illustrated in Table 3, individual exam scores were positively correlated with higher levels of peer evaluations, $t(152) = 1.68$, $p < .05$, one-tailed, providing support for Hypothesis 3. Satisfaction with goal attainment showed no correlation with peer evaluations (Hypothesis 4b). The results show a significant relationship between student observations of team process

effectiveness and peer evaluations, but in the opposite direction of what we proposed in Hypothesis 4a. Those students reporting more effective team processes in working together also received lower significant peer evaluations, $t(152) = -2.49$, $p < .01$.

Post hoc analyses. Our results regarding team process were opposite what we predicted, so we decided to examine the findings further. If team members rated each other equally, every team member's score would be 10, and we classified the team members' peer evaluations into three groups. Group 1 consisted of students who received an average of 10; that is, students whose contribution was evaluated as comparable to that of the other members in the group, suggesting that those individuals' peers perceived them as contributing equitably to other team members. Group 2 consisted of those receiving less than 10 on average (contribution rated as below an even share or contributing less than most team members). Finally, Group 3 consisted of those receiving more than 10 on average (contribution rated as beyond most other members' contributions). We ran a one-way ANOVA with these three groups using team process scores as the dependent variable, $F(2, 222) = 27.18$, $p < .01$.

The individuals in Group 1 (i.e., those who were evaluated as equitably participating) reported the highest team process effectiveness ($M = 13.48$, $SD = 2.08$, $N = 133$), the individuals in Group 2 (receiving lower evaluations than most other team members) reported the next highest team process effectiveness ($M = 12.50$, $SD = 2.45$, $N = 26$), and the individuals in Group 3 (receiving the highest evaluations) reported the lowest team process effectiveness ($M = 10.93$, $SD = 2.60$, $N = 66$). The average of Group 1 scores was significantly higher than the average of Group 3 scores ($M_{\text{difference}} = 2.37$, $p < .01$) when calculated with Dunnett's multiple comparison procedure.

Discussion

Collaborative, cooperative, problem-based, and team learning instructional technologies are increasing in popularity. These instructional approaches are more student-focused and team-oriented, as compared to the traditional instructor-focused models in education and training. In cooperative educational teams, students and employees learn to communicate better and develop team skills that are important in most organizations and content and concept acquisition is enhanced with the combination of team and individual work. Most of the collaborative approaches involve team performance on exams or projects that affect participants' final assessment to some degree. Since team members work closely over a period of time, they are more familiar with each other's contributions than are observers who are not part of the team. Because of the team history and familiarity, peer assessment

often is used as part of the evaluation of individuals on the team, such that members rate each other on contributions to the team, which is then applied to their final individual assessment. Because of the prevalence and importance of peer evaluation, we focused on this activity.

Extensive research has been directed toward peer evaluation with regard to reliability, validity, rater bias, inflated ratings, and methodology, but very little research has examined individual and team characteristics that have a relationship to peer assessment and that take into account time. In this investigation, we looked at a team member's characteristics prior to beginning the team experience (e.g., demographics, personality). For the post-team history factors, we examined individual performance and member observations of team process and goal attainment for any relationship with member's peer results. Little research has examined pre- and post-team history models and has taken into account individual influences, team influences, and ICCs on the analyses.

Pre-Team History

Our findings show that older team members, Caucasian members, and those reporting higher levels of internal locus of control were evaluated higher by their peers, while individuals with higher levels of communality tended to be evaluated lower by their peers. Older students most likely were more mature in their behavior and more experienced and, thus, were more respected. Additionally, older students may have full- or part-time work experience that they can leverage as a resource for team work, and, as a result of their maturity, may find it easier to focus on their class work. The fact that Caucasians received higher evaluations than did non-Caucasians may be explained by the large proportion of Caucasian team members (75%) in the sample. It may be that because of the large numbers, the behaviors and attitudes of Whites were perceived more as the norm and as desirable or acceptable by the majority, resulting in improved evaluations. Alternatively, it may be that communication was easier among Whites, as compared to non-Whites, since they are a large majority and because minorities were present in much smaller numbers.

Team members who were observed to have higher characteristics of communality received lower peer evaluations. The communality trait indicates an individual's tendency to help others, to be concerned about others, and to be attentive to their emotions. In this collaborative learning environment in which the focus was on assignments and academic tasks, higher levels of communality might not be welcome behaviors regarding the completion of these types of team requirements. This variable should be included in other

research to clarify its relationship with group activities. The internal locus of control characteristic manifests in a can-do attitude, which in a team setting could be comforting to other team members who wonder if their group can produce effectively. Such attitudes suggest that individuals with high internal locus of control tend to believe in themselves and in the positive consequences of their working hard for the team. Individuals with high internal locus of control may obtain higher evaluations because they may be assuming more responsibilities and accomplishing more tasks for the team.

Post-Team History

Our second model consisted of metrics gathered across the teams' history. A positive relationship exists between individual members' individual exam scores and their evaluations, and a negative relationship exists between how individuals rate the effectiveness of the team process and their peer evaluations. Team members with higher individual scores probably shared this information on team assignments, since members with the highest individual exam scores were on teams with the highest team scores. This preparation and sharing of information was rewarded by these individuals receiving higher peer-evaluation scores. We also assumed that members who reported more effective team process would be more positive about the team and more involved with the team, which will be rewarded by other members through higher peer evaluations. We did find a significant relation opposite the direction we predicted: Members who reported more effective team process received lower peer evaluations. We decided to examine this finding further.

Our post hoc analyses show that group members receiving average evaluations and those receiving less than average evaluations reported greater team process effectiveness than did members who received higher than average peer evaluations. That is, those who were evaluated as contributing more to the team than others reported lower team process effectiveness. An explanation could be that team members with below-average peer evaluations received more from the team (e.g., in team exam points) than they contributed to the team, so they had the disposition to feel better about the team process. Members with greater than average peer scores also had higher individual exam scores than other members, so they may have felt that they prepared more and gave more to the team, thus resulting in their dissatisfaction with the team and their lower team process scores.

The instructor communicated several times that the final peer evaluation would be confidential. To be fair to everyone, if a team member contributed more or less than did others on their team, this should be accounted for in the evaluations. Nevertheless, a sociality effect is likely. We observed that many

teams bonded together well, and we used no measures of team friendship or social closeness. Our feeling is that this did influence a number of members to give equal ratings on their teams, which inflated the reports of team process. This notion should be part of future peer-evaluation research.

Applications for Peer Evaluation

In this study, we attempted to provide a contribution to research on peer evaluations through better understanding of the factors that may influence peer evaluations at the individual and team levels. Of particular interest are the implications of the findings for the application and use of evaluation as a pedagogical tool and as a factor in student evaluation. For example, since age often correlates with experience and maturity, assigning older participants to each team, to the extent possible, seems to add those that are held in higher regard. The older member can be a resource, but in this situation, sufficient communication should be directed toward diversity and individual differences, from the point of older students understanding younger students and vice versa. The same focus should be applied especially when the class environment is largely comprised of one ethnicity. Our situation was predominantly Caucasian, and the peer ratings show a preference for this group. Considerable discussion and application of diversity exercises and techniques should be conducted when one ethnicity is the majority.

This also occurred for individuals demonstrating internal locus of control characteristics. The assumption is that students scoring higher levels of internal locus of control do exhibit these behaviors, and this can-do attitude influences higher peer evaluations. Trait research has shown that individuals with higher levels of this trait do take on more responsibility, so this could be rooted in positive team behaviors. Regular communication describing behaviors and activities needed for team synergy and performance should be emphasized over traits, age, and ethnicity. Once regular processing of these team behaviors is conducted, demographic elements tend to decline in influence (Watson et al., 1998).

Study Limitations

This was initial research on individual and collaborative factors that may influence peer evaluations, and we consider this only a starting point for this type of investigation. Validity research is very important for peer evaluations, and we assert that examination of individual and team elements affecting these assessments is also vital. We used one type of peer evaluation. From

many years of working with team learning, we feel our approach is very solid, but it is only one approach. Other methods of assessment should be used in this context.

Our sample was predominantly Caucasian and produced a bias on that variable. Similar studies with other majority ethnicities should be conducted. In addition, we had regular feedback sessions across the semester regarding team issues, and we explained team factors and individual behaviors necessary for high team performance. This should be examined further. Different types of feedback mechanisms should be applied to assess which ones may be better. We did not measure participant observations of our processing, which should also be included in future research.

Other personality measures should be applied. We felt the two that we used were very appropriate for the present study, but there are many other relevant instruments to help us better understand the mechanisms at play (e.g., Big Five personality model). Even though we felt that our analyses were proper, with larger *N* sizes and from other learning environments, more extensive models could be developed. We do think this examination of elements that may influence peer evaluations is a contribution to the current literature, but much more is needed in this direction.

With increases in the use of collaborative learning formats, the importance of effective peer assessment also increases. Many studies have examined validity, reliability, and type-of-assessment issues. Very little work has addressed pre-team history and post-team history variables that correlate with peer assessment. We found no research that has focused on pre- and post-team history elements and that has taken into account ICCs and individual and team influences on the analyses (e.g., Kenny et al., 2002).

We found that the demographics of age and ethnicity, the trait of internal locus of control, and reports of team process significantly correlated with levels of peer evaluation and the individual level of analysis. Older Caucasians and those reporting internal locus of control were evaluated more highly by their peers regarding their contributions to the team. Members who reported higher levels of team process were more often evaluated as doing their full share of teamwork.

If collaborative team instructional formats are used in a classroom setting, then regardless of the content, feedback and processing individual and team behaviors should be part of the instructional technology. Communication and instruction should be directed toward individual differences in terms of basic demographics and personality traits. Discussion should take place to explain how biases easily arise from this, but that resources could be derived from the differences (Watson, Kumar, & Michaelsen, 1993). Critical team process effectiveness activities should be targeted because the behavior of helping other team members is important as is individually performing one's own tasks.

References

Barrett, R. S. (1996). *Performance rating*. Chicago: Science Research Associates.

Beatty, J. R., Hass, R. W., & Seiglimpaglia, I. (1996). Using peer evaluations to assess individual performances in group class projects. *Journal of Marketing Education*, *18*, 17–27.

Blackmore, J. A. (2005). A critical evaluation of peer review via teaching observation within higher education. *International Journal of Educational Management*, *19*, 218–232.

Bothwell, R. K., Brigham, J. C., & Malpass, R. S. (1989). Cross-racial identification. *Personality and Social Psychology Bulletin*, *15*, 19–25.

Brooks, C. M., & Ammons, J. L. (2003). Free riding in group projects and the effects of timing, frequency, and specificity of criteria in peer assignments. *Journal of Education for Business*, *78*, 268–272.

Cederblom, D., & Lounsbury, J. W. (1980). An investigation of user acceptance of peer evaluations. *Personnel Psychology*, *33*, 567–579.

Chasnoff, R. (Ed.). (1979). *Structuring cooperative learning: The 1979 handbook*. New Brighton, MN: Interaction Book Company.

Chen, M., & Bargh, J. (1997). Nonconscious behavioral confirmation processes: The self-fulfilling consequences of automatic stereotype activation. *Journal of Experimental Social Psychology*, *33*, 541–560.

Clark, M. S. (1986). Evidence for the effectiveness of manipulations of desire for communal and exchange relationships. *Personality and Social Psychology Bulletin*, *12*, 414–425.

Clark, M. S., Ouellette, R., Powell, M. C., & Milberg, S. (1987). Recipient's mood, relationship type, and helping. *Journal of Personality and Social Psychology*, *53*, 94–103.

Correll, S. J. (2004). Constraints into preferences: Gender, status, and emerging career aspirations. *American Sociological Review*, *69*, 93–113.

DeNisi, A. S., Randolph, W. A., & Blencoe, A. (1982). Level and source of feedback as determinants of feedback effectiveness. *Proceedings of the Academy of Management Annual Meeting*, 175–179.

Devine, D. J., Clayton, L. D., Philips, J. L., Dunford, B. B., & Melner, S. B. (1999). Teams in organizations: Prevalence, characteristics, and effectiveness. *Small Group Research*, *30*, 678–711.

Dominick, P. G., Reilly, R. R., & McGourty, J. W. (1997). The effects of peer feedback on team member behavior. *Group and Organization Management*, *22*, 508–520.

Dovidio, J. F., Kawakami, K., & Gaertner, S. L. (2002). Implicit and explicit prejudice and interracial interaction. *Journal of Personality and Social Psychology*, *82*, 62–68.

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

Falchikov, N., & Magin, D. (1997). Detecting gender bias in peer marking of students' group process work. *Assessment and Evaluation in Higher Education*, *22*, 393–404.

Fazio, R. H. (1990). Multiple attitudes by which attitudes guide behavior: The MODE model as an integrative framework. In M. P. Zanna (Ed.), *Advances in experimental social psychology* (Vol. *23*, pp. 75–109). Orlando, FL: Academic Press.

Ghorpade, J., & Lackritz, J. R. (2001). Peer evaluation in the classroom: A check for sex and race/ethnicity effects. *Journal of Education for Business*, *76*, 274–281.

Harrigan, J. A., & Wigdor, A. K. (1989). *Fairness in employment testing: Validity, generalization, minority issues, and the General Aptitude Test Battery*. Washington, DC: National Academy Press.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior*, *29*, 340–362.

Ilgen, D. R., & Feldman, J. M. (1983). Performance appraisal: A process focus. In L. Cummings & B. Staw (Eds.), *Research in organizational behavior* (Vol. *5*, pp. 141–197). Greenwich, CT: JAI Press.

Johnson, C. B., & Smith, F. I. (1997). Assessment of a complex peer evaluation instrument for team learning and group processes. *Accounting Education*, *2*, 21–40.

Johnson, D. W., & Johnson, F. P. (1994). *Joining together: Group theory and group skills*. Boston: Allyn & Bacon.

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

Kashy, D. A., & Kenny, D. A. (2000). The analysis of data from dyads and groups. In H. Reis & C. M. Judd (Eds.), *Handbook of research methods in social psychology* (pp. 451–477). New York: Cambridge University Press.

Kenny, D. A., Kashy, D. A., & Cook, W. L. (2006). *Dyadic data analysis*. New York: Guilford.

Kenny, D. A., Mannetti, L., Pierro, A., Livi, S., & Kashy, D. A. (2002). The statistical analysis of data from small groups. *Journal of Social Psychology and Personality*, *83*, 126–137.

Kobasa, S. C., Maddi, S. R., & Kahn, S. (1982). Hardiness and health: A prospective study. *Journal of Personality and Social Psychology*, *42*, 168–177.

Kozlowski, S. W. J., & Klein, K. J. (2000). A multilevel approach to theory and research in organizations: Contextual, temporal, and emergent processes. In K. J. Klein & S. W. J. Kozlowski (Eds.), *Multilevel theory,*

research and methods in organizations: Foundations, extensions, and new directions (pp. 3–90). San Francisco: Jossey-Bass.

Landy, F. J., & Farr, J. L. (1983). *The measurement of work performance: Methods, theory, and applications*. New York: Academic Press.

Lejk, M., & Wyvill, M. (2001). The effect of the inclusion of self-assessment with peer assessment of contributions to a group project: A quantitative study of secret and agreed assessments. *Assessment and Evaluation in Higher Education*, *26*, 551–561.

LePine, J. A. (2003). Team adaptation and post change performance: Effects of team composition in terms of members' cognitive ability and personality. *Journal of Applied Psychology*, *88*, 27–39.

LePine, J. A., Colquitt, J. A., & Erez, A. (2000). Adaptability to changing task contexts: Effects of general cognitive ability, conscientiousness, and openness to experience. *Personnel Psychology*, *53*, 563–593.

LePine, J. A., & van Dyne, L. (2001). Peer responses to low performers: An attributional model of helping in the context of groups. *Academy of Management Review*, *26*, 67–84.

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.

Lindsay, D. S., Jack, P. S., & Christian, M. A. (1991). Other-race face perception. *Journal of Applied Psychology*, *76*, 587–589.

Michaelsen, L. K., Knight, A. B., & Fink, L. D. (2004). *Team-based learning: A transformative use of small groups in college teaching*. Sterling, VA: Stylus Publishing.

Michaelsen, L. K., & Watson, W. E. (1993). Beyond groups and cooperation: Building high performance learning teams. In D. L. Wright & J. P. Lunde (Eds.), *To improve the academy* (Vol. *12*, pp. 127–146). Stillwater, OK: New Forums Press.

Morahan-Martin, J. (1996). Should peers' evaluations be used in class projects? Questions regarding reliability, leniency, and acceptance. *Psychological Reports*, *78*, 1243–1250.

Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal: An organizational perspective*. Boston: Allyn & Bacon.

Offermann, L. R., Bailey, J. R., Vasilopoulos, N. L., Seal, C., & Sass, M. (2002). The relative contribution of emotional competence and cognitive ability to individual and team performance. *Human Performance*, *17*, 219–243.

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

Savin-Baden, M. (2003). *Facilitating problem-based training*. Philadelphia: McGraw-Hill.

Schneider, D. J., Hastorf, A. H., & Ellsworth, P. C. (1979). *Person perception*. Reading, MA: Addison-Wesley.

Sivan, A. (2000). The implementation of peer assessment: An action research approach. *Assessment in Education*, *7*, 193–213.

Spector, P. E. (1982). Behavior in organizations as a function of employee's locus of control. *Psychological Bulletin*, *91*, 482–497.

Spector, P. E. (1988). Development of the Work Locus of Control Scale. *Journal of Occupational Psychology*, *61*, 335–340.

Spector, P. E., Cooper, C. L., Sanchez, J. I., & O'Driscoll, M. (2002). Locus of control and well-being at work: How generalizable are Western findings? *Academy of Management Journal*, *45*, 453–466.

Sternberg, R. J., & Wagner, R. K. (Eds.). (1986). *Practical intelligence: Nature and origins of competence in the everyday world*. New York: Cambridge University Press.

Stewart, G. L., Fulmer, I. S., & Barrick, M. R. (2005). An exploration of member roles as a multilevel linking mechanism for individual traits and team outcomes. *Personnel Psychology*, *58*, 343–365.

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

Struthers, C. W., Weiner, B., & Alfred, K. (1998). Effects of causal attributions on personnel decisions: A social motivation perspective. *Basic and Applied Social Psychology*, *20*, 155–166.

Taggart, S., & Neubert, M. (2004). The impact of poor performers on team outcomes: An empirical examination of attribution theory. *Personnel Psychology*, *57*, 935–968.

Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine*, *140*, 227–235.

Ward, M. A. (2005). An empirical assessment of peer review reliability: The ipsitive instrument paradox. *Journal of Education for Business*, *81*, 35–40.

Watson, W. E., BarNir, A., & Pavur, R. (2005). Cultural diversity and learning teams: The impact on desired academic team processes. *International Journal of Intercultural Relations*, *29*, 449–467.

Watson, W. E., Johnson, L., & Merritt, D. (1998). Team orientation, self-orientation, and diversity in task groups: Their connection to team performance over time. *Journal of Group and Organization Management*, *23*, 161–188.

Watson, W. E., Kumar, K., & Michaelsen, L. K. (1993). Cultural diversity's impact on interaction process and performance: Comparing homogeneous and diverse task groups. *Academy of Management Journal*, *36*, 590–602.

Watson, W. E., Michaelsen, L. K., & Sharp, W. (1991). Member competence, group interaction, and group decision-making: A longitudinal study. *Journal of Applied Psychology*, *76*, 803–810.

Weiner, B. (2000). Attributional thoughts about consumer behavior. *Journal of Consumer Research*, *27*, 382–387.