Journal of Applied Psychology
1999, Vol. 84, No. 1, 58-74

0021-9010/99/$3.00

Effects and Timing of Developmental Peer Appraisals in Self-Managing Work Groups

Vanessa Urch Druskat
Case Western Reserve University

Steven B. Wolff
Boston University

This study used a repeated measures time-series design to examine the immediate and longer term impact of a structured, face-to-face developmental peer appraisal on 294 undergraduates in 44 self-managing work groups (SMWGs) and 217 MBA students in 36 SMWGs. Results revealed an immediate positive impact on perceptions of open communication, task motivation, social loafing, group viability, cohesion, and satisfaction. Also, the effects of the peer appraisal were not dependent on the perceived ratio of positive to negative feedback, and the enduring impact of the appraisal was influenced by its timing relative to task deadline. Overall, results emphasized that peer appraisals can have a positive effect on relationships and task focus, are influenced by temporal context, and have great potential for work teams.

Organizations that have implemented self-managing work groups (SMWGs) and eliminated a layer of supervisory management are now facing difficult decisions about how to manage their self-managing groups (Saavedra & Kwun, 1993). One of the most complex of these decisions is, in the absence of a manager or supervisor, how to appraise group member performance and facilitate employee development. Many are concluding that peer appraisals have great potential to help fulfill this need and are a logical and even necessary addition to the structure of group self-management (Mohrman, Resnick-West, & Lawler, 1989; Peiperl & Abelson, 1995; Rodgers, 1995). Peer appraisals can increase group members' ability to be a proactive source of their own feedback—a necessity for SMWGs that monitor and manage their own performance. Also, peers have been called the most accurate and informed judges of their coworkers' behavior (Kane & Lawler, 1978; Lewin & Zwany, 1976; Murphy & Cleveland, 1991; Wexley & Klimoski, 1984), have been found to be better than supervisors at evaluating skills that lead to improved performance (Yammarino & Waldman, 1993), and have been found to be accurate at predicting future job performance (Kane & Lawler, 1978; Reilly & Chao, 1982) and advancement (Shore, Shore & Thornton, 1992).

Despite their potential, peer appraisals in SMWGs have met stiff opposition in organizational settings (Cederblom & Lounsbury, 1980; DeNisi & Mitchell, 1978; Kane & Lawler, 1978; Love, 1981; McEvoy & Buller, 1987). Organizations are concerned about bias and appraisals turning into popularity contests (Kane & Lawler, 1978; Love, 1981) that can harm relationships and, thus, impair a group's ability to accomplish its task. Indeed, research has indicated that peer appraisals are subject to rater biases (Fox, Ben-Nahum, & Yinon, 1989; Saavedra & Kwun, 1993) and can impact member perceptions of the group (DeNisi, Randolph, & Blencoe, 1983). It is important to note that no research has addressed the issues of prime concern to organizations—how do peer appraisals impact individual members, relationships, and task focus.

We propose that if information learned from previous research is used to design a facilitative appraisal context, peer appraisals can produce the opposite of what organizations fear. They can serve to clarify expectations and improve individual satisfaction, interpersonal relationships, and task focus.

Influence of the Appraisal Context

Research on performance appraisals has consistently revealed the important influence of appraisal context (Conlon & Barr, 1989; Farh, Cannella, & Bedeian, 1991; Ilgen &

Vanessa Urch Druskat, Department of Organizational Behavior, Case Western Reserve University; Steven B. Wolff, School of Management, Boston University.

Steven B. Wolff is now at the Management Department, Marist College.

We are grateful to William A. Kahn, Philip E. Kubzansky, Michael Shwartz, and Aimin Yan for their helpful comments and suggestions. This research was supported by a grant from the Human Resource Policy Institute at Boston University.

Correspondence concerning this article should be addressed to Vanessa Urch Druskat, Department of Organizational Behavior, Weatherhead School of Management, Case Western Reserve University, 10900 Euclid Avenue, Cleveland, Ohio 44106-7235. Electronic mail may be sent to vxd@po.cwru.edu.

Feldman, 1983; Judge & Ferris, 1993; Mitchell, 1983; Saavedra & Kwun, 1993). This research suggests two aspects of the peer appraisal context that may influence its impact: the appraisal's focus on development over evaluation (Farh et al., 1991; McEvoy & Buller, 1987) and the timing of the appraisal in relation to the task deadline (Gersick, 1988, 1989; McGrath, 1991).

Peer Appraisals Focused on Development

An important factor in the peer appraisal context is its purpose (Farh et al., 1991; McEvoy & Buller, 1987). Peer appraisals are usually intended to provide feedback and facilitate employee development. They may also include an evaluative component used for such things as salary increases or future employment. Recent research has found that when peer appraisals are used for development only, several reasons for resisting their use are addressed. When strictly developmental, appraisals are significantly less lenient, less subject to halo error, more discriminating, and have higher levels of interrater reliability (Farh et al., 1991). Also, group members in field (McEvoy & Buller, 1987) and experimental studies (Farh et al., 1991) show less resistance to their use.

Peer appraisals focused on development can be comfortably less secretive and more open than those used for evaluation and can more easily involve face-to-face feedback. Thus, through performance discussions, members can build and refine feedback and communication skills. Providing peer feedback can also strengthen skills in observation, evaluation, and reinforcement—all identified as important to effective group self-management (Druskat, 1997; Manz & Sims, 1987; McIntyre & Salas, 1995).

Specifically, we propose that giving and receiving feedback in a structured face-to-face process can improve patterns of interaction and communication in a group and have a positive impact on individuals, relationships, and task focus. According to McGrath (1984), over time, task groups develop patterns around the form and content of communication. These can have an important impact on group members, relationships in the group, task performance, and future interactions (McGrath, 1984).

Several features of a structured peer appraisal process can serve to develop and improve the areas of communication discussed by McGrath (1984). Giving and receiving positive and negative feedback can improve the openness of communication (Luft, 1984). The active listening and discussion involved in a developmentally focused appraisal can improve relationships (Rogers & Farson, 1995). The act of giving and receiving feedback can also prompt examination and clarification of group norms, values, and expectations of task behavior, thus, facilitating role negotiation, task execution, and task focus (Katz & Kahn, 1978).

Peer appraisals can also serve as a temporal interruption or *trigger event* that legitimizes reflection and evaluation—necessary ingredients for change and development (Tyre, Perlow, Staudenmayer, & Wasson, 1996). Thus, peer appraisals can prevent work groups and individuals from slipping into the dysfunctional habitual routines (Gersick & Hackman, 1990) common in work groups. They can serve as important catalysts for the positive development of individuals, relationships, and task focus.

No research has examined the impact of developmental peer appraisals on individuals, group member relationships, and task focus. These are important issues because they encompass the primary reasons organizations resist peer appraisals. On the basis of the foregoing analysis, we propose the hypotheses listed below. It is important to note that our hypotheses focus on individual perceptions rather than group behavior. Group change takes time, and given our measurement period, we felt we could legitimately propose that individual perceptions, but not group behavior, would change. Moreover, the impact of the peer appraisal on individual members is critical to member relationships (e.g., cohesion and satisfaction with the group) and task focus (e.g., task motivation and social loafing). Research in social cognition also has shown that perceptions predict future behavior (Allport, 1954; Fiske & Taylor, 1991).

> *Hypothesis 1.* A developmental peer appraisal has an immediate beneficial effect on group member perceptions of (a) open communication, (b) task motivation, (c) social loafing, (d) group viability (a group's ability to continue working together effectively—a concept that involves positive member relationships and task focus), (e) cohesion, and (f) satisfaction with the group.

> *Hypothesis 2.* A developmental peer appraisal is perceived as being useful for (a) individual development, (b) increasing open communication within the group, and (c) improving the group's ability to accomplish its task.

Previous research has found that receiving a negative peer evaluation can have a negative impact on a member's ratings of perceived group performance, cohesiveness, and satisfaction, (DeNisi et al., 1983). However, that research was conducted in a laboratory context in which teams had no future together and members received bogus feedback (a rating of 2.5 out of 7) with no explanation. On the basis of our analysis, a structured, developmental peer appraisal involving discussion can clarify expectations and have a positive impact on individuals, relationships, and task focus. Thus, we propose that the immediate and enduring effect of the peer appraisal will not depend on the perceived ratio of positive to negative feedback received.

> *Hypothesis 3.* Performance appraisals that are perceived to be largely positive and performance appraisals that are perceived to contain a balance of positive and negative feedback will have the same immediate and enduring effects on perceptions of (a) group open communication, (b) task motivation, (c) social loafing, (d) viability, (e) cohesion, (f) satis-

faction with group, and ratings of the usefulness of the appraisal for (g) one's own development, (h) improving open communication, and (i) improving a group's ability to accomplish its task.

Timing of the Peer Appraisal

Theory and research have recently begun to reveal the influence of time on group development, process, and task focus (Gersick, 1988, 1989; McGrath, 1991; McGrath & O'Connor, 1996; Moreland & Levine, 1982, 1984, 1988). This research revealed that at different points in a group's history, it is focused on issues such as making choices about goals and developing initial strategies, negotiating and renegotiating expected contributions, resolving conflicting preferences, and accelerating effort when near task completion (Gersick, 1988, 1989; McGrath, 1991; McGrath & O'Connor, 1996).

In sum, this theory and research suggested that any act or intervention in a group's life takes on a different meaning depending on its timing relative to a group's task deadline. Therefore, it seems reasonable that there exists a period of time when a group and its members are most likely to accept and learn from a peer appraisal. For example, the benefits may be most enduring if the appraisal is conducted when members are settling on the most appropriate ways to carry out their tasks and negotiating their expected roles and contributions, rather than when they are getting started or focusing on task completion.

On the basis of our assessment that the timing of the peer appraisal relative to the task deadline will have an impact on the developmental or enduring effect of the appraisal, the following hypothesis is offered:

> *Hypothesis 4.* The longer term or enduring effects of the peer appraisal on perceptions of (a) group open communication, (b) task motivation, (c) social loafing, (d) viability, (e) cohesion, and (f) satisfaction with group, vary depending on when the appraisal is conducted relative to the group's task deadline.

Method

Participants

Participants were 511 students in an East Coast university working in 80 self-managing project groups in 13 sections of required introductory organizational behavior courses. Sections differed in their enrollment; eight sections were undergraduate business majors (n = 294; median age = 22 years), five were MBA graduate students (n = 217; median age = 26 years). For purposes of course standardization, all sections used a similar design. In all sections, students were required to work in SMWGs of 5–6 students responsible for completing a group research project and making a formal presentation of their results (undergraduate = 44 groups; MBA = 36 groups). In all sections, instructors assigned students to groups and made them demographically heterogeneous. Each section required groups to manage and conduct their own standardized in-class peer appraisal. The MBA students formed groups during Week 2 of a 14-week semester and conducted appraisals during Week 8. The undergraduates formed groups during Week 3 of a 14-week semester and conducted appraisals during Week 10.

The fact that we chose to test our hypotheses using students in an educational context raises the important issue of external validity. According to Locke (1986), generalizability of a laboratory study to a field setting depends on the similarity between settings on key attributes. We did our best to find a sample that simulated the key attributes of SMWGs in work settings. Groups were self-managing because they completed tasks requiring member interdependence (Saavedra & Kwun, 1993) and held full responsibility for executing their work and monitoring and managing their own processes (Hackman, 1986). Member level of involvement was high because groups worked on in-class exercises throughout the semester and the peer appraisal was conducted before project deadlines, thus simulating ongoing groups. Also, the group project accounted for 25% of the members' course grades. Finally, we studied the impact of the peer appraisal on constructs that can be considered universal in task groups (e.g., task motivation, cohesion). In sum, because key attributes in our setting were similar to those of SMWGs in work settings, we believe the study results are likely generalizable.

Peer Appraisal

Each participant received a standardized handbook (see Wolff, 1998) that included four structured team assignments; three as preparation for the peer appraisal, and the appraisal in which group members gave and received verbal and written feedback. The first assignment involved (a) developing written expectations and a performance plan (e.g., error-free work), (b) policies and procedures for addressing breaches of expectations (e.g., written warnings), and (c) a formal appraisal form.

In the second assignment, members were asked to observe team member contributions throughout the semester and keep a record of the specific behaviors exhibited by each member and their effect on the group. The handbook provided observation forms prompting a balance between positive and negative comments.

The third assignment involved preparation for the appraisal. Each team member selected one other member for whom he or she would be responsible for managing the appraisal. One class before the appraisal all team members distributed observation forms they had completed throughout the semester (Assignment 2) to each member's appraisal manager. The appraisal managers summarized all comments and completed the formal appraisal form (Assignment 1) for their appraisee. This pooling process was chosen because pooled peer ratings remove idiosyncratic biases (Saavedra & Kwun, 1993) and, thus, should increase appraisal reliability and validity. The feedback was pooled by one member rather than the group because group-determined peer ratings are highly receptive to response biases (Martell & Borg, 1993).

During the in-class appraisals, appraisal managers gave a verbal summary of the team's feedback to their appraisees while members observed. Instructions encouraged appraisees to seek clarification, ask for examples, and paraphrase. Appraisees kept the appraisal form summarizing their feedback and the observation forms used to compile the feedback. To ensure quality, a copy of

the appraisal was given to the instructor, although students knew appraisals would not affect their grades.

Research Design and Procedures

To test our hypotheses, we used a repeated measures time-series design (Campbell & Stanley, 1963). As such, a questionnaire was administered to participants four times during the semester. Time 1 was 2 weeks before the peer appraisal, Time 2 was 1 week before the peer appraisal, Time 3 was immediately after the appraisal (same class period), and Time 4 was three class periods after the peer appraisal (the midpoint of team project presentations, i.e., half of the teams had presented). The repeated measures design permitted us to assess changes in group member attitudes and ratings of the team as a result of the peer appraisal. It also allowed us to determine maturation effects occurring between Time 1 and Time 2 and then to assess whether the effects of the peer appraisal intervention at Time 3 and Time 4 were above and beyond predicted maturation. Testing for predicted maturation effects is not an ideal substitute for the use of a control group in the research design. However, we were unable to withhold the peer appraisal from some groups.

Questionnaires were administered during class time. Students were assured of confidentiality and told that participation was voluntary and would not affect their grade (instructors would see only composite summaries after the semester). Participation was 100%; however, the sample sizes in our analyses varied with class attendance on the four dates of data collection. In the undergraduate sections, questionnaires were administered by the instructors who read standardized instructions and sealed completed forms in an envelope. In the MBA sections, one of the researchers administered the questionnaires and read standardized instructions. Written and verbal instructions pointed out that students had answered the same or similar questions before, that feelings about one's group and its members may change over time, and that answers should be based on "how you feel today."

Project Deadlines and the Timing of the Peer Appraisal

Within the MBA sections, project groups had differing project deadline–presentation dates assigned at the beginning of the semester. We grouped them into three categories based on their deadlines: (a) early presenters, (b) midpresenters, and (c) late presenters. Early presenters conducted the peer appraisal one to two class periods prior to their project deadline. Midpresenters conducted the peer appraisal three class periods prior to their project deadline. The late presenters conducted the peer appraisal four to five class periods prior to their deadline.

In the undergraduate sections, students were not aware of their project deadline–presentation dates until 2 weeks before the deadline and after they had completed the peer appraisal. This ambiguity permitted them to serve as a comparison group for Hypothesis 4. The undergraduates cannot be considered a perfect comparison group as their course was not identical to the MBA course. The most important difference was that the project deadlines in the MBA sections stretched over three dates. For the undergraduates, deadlines stretched over two dates (two and three class periods after the appraisal—consistent with the timing of the early and mid-presenters in the MBA sections). Nevertheless, the two groups were similar in many ways explained above, making the undergraduates an informative comparison group.

Testing Effects

The research design required that each participant complete a similar questionnaire four times. Thus, to ensure internal validity, it was necessary to examine for testing effects (Campbell & Stanley, 1963). We used two control groups each consisting of two randomly selected undergraduate sections. These sections participated in the peer appraisal. However, one control group completed the questionnaire for the first time at Time 3 and the other for the first time at Time 4. We did not have a control group for the graduate sections; thus, we made the assumption that testing effects for the graduate group would be comparable to those of the undergraduate group.

Dependent Variables

We used previously validated scales to measure member ratings of open communication (Stokes, 1983: four items; e.g., "My group avoids saying anything which might upset someone"), task motivation (Zaccaro & McCoy, 1988: six items; e.g., "Performing well is a top priority for my team"), group viability (Hackman, 1988: six items; e.g., "As a team, this work group shows signs of falling apart" [reverse scored]), group cohesion (Stokes, 1983: five items; e.g., "If I were to participate in another group like this one, I would want it to include people who are very similar to the ones in this group"), and satisfaction with group (Hackman, 1988: two items; e.g., "Generally speaking I am very satisfied with my team"). We could find no scales measuring perceived social loafing, so we used definitions from the literature (see Harkins & Jackson, 1985) to develop a three-item scale: (a) "We have some team members that don't put much effort into their work," (b) "Every member of our team does her/his fair share of the work" (reverse scored), and (c) "There are some individuals on our team who don't do much work" (Cronbach's $\alpha = .87$). Response formats ranged from 1 (*strongly disagree*) to 7 (*strongly agree*), with some items reverse scored.

We assumed that the constructs we were measuring would be interrelated, and thus to ensure the integrity of each construct, we subjected the 511 responses collected at Time 1 to a factor analysis using principal-components analysis with oblique rotation. The eigenvalues above 1.0 and the scree plot suggested a 5-factor solution. All items for cohesion, openness, group task motivation, and social loafing loaded highest onto their respective dimensions. Two of the seven questions in the viability scale and all of the questions in the satisfaction scale loaded onto other dimensions. We decided to maintain satisfaction as a separate construct and, on the basis of the analysis, made minor modifications to increase the internal reliability of the scales measuring satisfaction and group task motivation. Table 1 presents the intercorrelations among the scales and their Cronbach's alpha internal reliabilities.

To assess team member ratings of the usefulness of the peer

Table 1
Correlations Among Variables for Entire Population at Time 3 (n = 400) and Cronbach's Alpha Reliability for Dependent Variable Scales

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9
1. Useful to group communication	5.42	1.00	—								
2. Useful to group task	5.45	0.95	.69**	—							
3. Useful to self	5.03	1.45	.52**	.52**	—						
4. Cohesion	5.07	1.17	.29**	.28**	.28**	(.84)					
5. Openness	4.80	1.05	.15**	.10*	.19**	.40**	(.56)				
6. Satisfaction	5.68	1.25	.27**	.26**	.22**	.69**	.34**	(.82)			
7. Viability	5.38	1.00	.28**	.23**	.24**	.62**	.43**	.76**	(.78)		
8. Group Task Motivation	5.43	0.74	.23**	.17**	.21**	.48**	.35**	.58**	.55**	(.68)	
9. Loafing	3.19	1.56	−.16**	−.08	−.14**	−.50**	−.32**	−.61**	−.74**	−.43**	(.87)
10. Feedback ratio	7.34	1.57	.16**	.20**	.16**	.28**	.14**	.24**	.15**	−.24**	−.07

Note. Cronbach's alpha is shown along the diagonal in parentheses.
* $p < .05$. ** $p < .01$.

appraisal, we collected data before and after the peer appraisal asking how useful the peer appraisal would be or was to them personally; how helpful or harmful they thought the peer appraisal would be or was to increasing member open and honest communication; and how helpful or harmful they thought the peer appraisal would be or was to the group's ability to accomplish its task. These items were rated on a 7-point Likert scale (1 = *extremely harmful,* 4 = *no effect,* 7 = *extremely helpful*).

Finally, to assess member perceptions of their ratio of positive to negative feedback, we included an item that asked "How positive/negative was the feedback you personally received?" This item was rated on a 10-point Likert scale (0 = *all negative,* 5 = *equal positive and negative,* 10 = *all positive*).

Results

We conducted our analysis at the individual level using a significance level (alpha) of .05 and two-tailed probabilities. Our sample sizes for the undergraduate and graduate groups were large enough to allow us to report the results for each group separately (for $\eta^2 > .05$, power $> .75$).

Immediate Effects of Peer Appraisal

Hypothesis 1 addressed the immediate impact of the peer appraisal on perceptions of (a) open communication, (b) task motivation, (c) social loafing, (d) group viability, (e) cohesion, and (f) satisfaction with the group. To test this hypothesis, we used a repeated measures analysis of variance (ANOVA) with orthogonal a priori difference contrasts. The difference contrast compares the value of a variable at a given time to the average of its values at all previous times.

We present the results in graphical format (see Figure 1) and provide descriptive statistics and results of the analysis in Tables 2 and 3. We tested for maturation between Time 1 and Time 2 by examining the contrast comparing the variable's value at Time 1 with its value at Time 2 (labeled "T2" in Tables 2 and 3). Cohesion in the undergraduate group was the only variable to show significant maturation in the same direction as the expected effect of the peer appraisal. Thus, for all variables except undergraduate cohesion, we tested for an immediate effect of the peer appraisal by examining the contrast comparing the value of the variable at Time 3 to the average of its values at all previous times (labeled "T3" in Tables 2 and 3). We tested for an immediate effect of the peer appraisal on cohesion in the undergraduate group by determining whether the change in value from Time 2 to Time 3 was greater than the change in value from Time 1 to Time 2.

Figure 1 shows that peer appraisal (Time 3) had an immediate effect on all variables. With the exception of cohesion in the undergraduate group, the effect was significant and represents improvement. For cohesion in the undergraduate group, there was an improvement at Time 3 compared to Time 2, $t(162) = 4.84$, $p < .01$; however, this increase was not different from the increase between Time 1 and Time 2, $t(128) = 1.28$, *ns.* Thus, we could not conclude that the peer appraisal produced an immediate improvement in cohesion in the undergraduate group.

We surveyed each participant four times; thus, a testing effect is a plausible alternate hypothesis for these results. Therefore, using our two undergraduate control groups, we tested each variable for testing effects at Time 3 and Time 4 using Tukey's honestly significant difference procedure (combined $\alpha = .05$) to compare the variable's mean value in the control groups (i.e., those who filled out the survey for the first time at Time 3 or Time 4, respectively) with its mean value in each equivalent treatment section. There were no cases where the control group mean was significantly different from that of a treatment section. Thus, there was not sufficient evidence to suggest that a testing effect is a viable alternate hypothesis. The above analysis supports Hypothesis 1 for the graduate group and all but Hypothesis 1e for the undergraduate group.

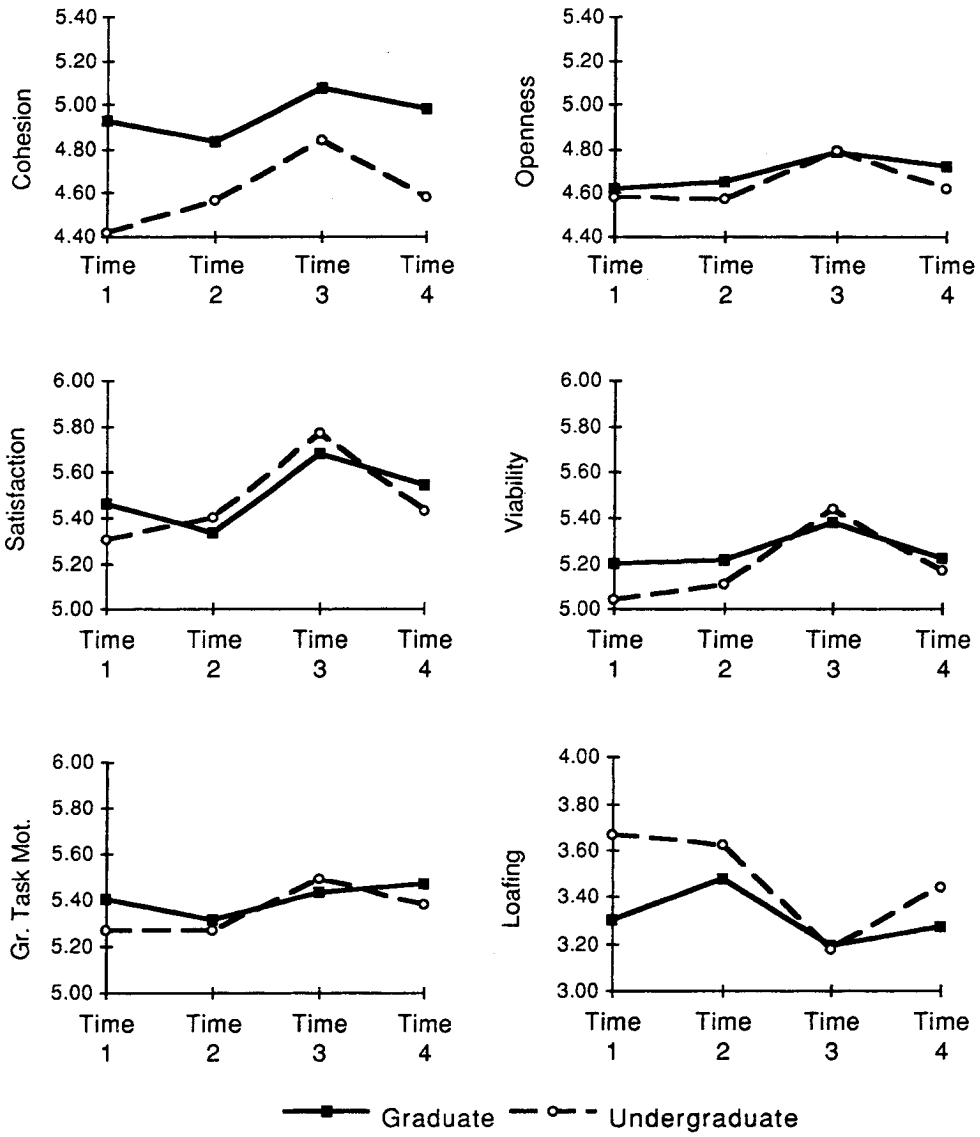

Figure 1. Graphs of dependent variables over time. There were significant effects at Time 3 (immediately after the peer review) for all variables. Time 2 was 1 week before Time 3, Time 1 was 1 week before Time 2, and Time 4 was three class sessions after Time 3. Descriptive statistics can be found in Tables 2 and 3. Gr. Task Mot. = group task motivation.

Group Member Rating of the Peer Appraisal Process

Hypothesis 2 addressed ratings of usefulness of the peer appraisal for (a) individual development, (b) open communication, and (c) ability to accomplish tasks. Tables 4 and 5 show participant ratings for the graduate and undergraduate groups.

A repeated measures ANOVA with an a priori difference contrast showed significant effects for the dimensions of open communication and ability to accomplish the task but not for individual development. Immediately following the appraisal, both graduates and undergraduates rated the peer appraisal as more helpful than expected for the group's ability to complete its task and for fostering open communication (see T3 contrast in Tables 4 and 5). For both groups, perception of the usefulness of the appraisal for individual development did not change from Time 2 to Time 3.

Additionally, for each dimension of usefulness, we performed a t test using the sample mean to determine whether the population mean score equals 4 (which was anchored by no effect). For each of the three dimensions of usefulness, we rejected the hypothesis that the population mean equals 4 ($t > 10, p < .01$) and concluded that the sample mean was

(*text continues on page 68*)

Table 2

Graduate Group (Overall and by Project Deadline; n = 138): Descriptive Statistics and Repeated Measures ANOVA A Priori Contrasts Testing for the Effect of the Peer Appraisal

	Time 1		Time 2		Time 3		Time 4		A priori contrast							
									T2[a]		T3[b]		Lin.[c]		Quad.[c]	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*	η^2	*t*	η^2	*t*	η^2	*t*	η^2
Cohesion[d]	4.92	1.08	4.83	1.25	5.07	1.17	4.98	1.30	−1.46	.02	3.50*	.08	1.99*	.03	−3.59*	.09
Early[e]			5.26	1.08	5.46	0.91	5.41	1.15			2.18*	.07	1.91	.03	3.67†	.06
Mid[f]			4.63	1.24	4.84	1.27	4.88	1.17			2.41*	.11	3.36†	.07	1.29	.03
Late[g]			4.28	1.34	4.62	1.26	4.30	1.45			2.70*	.20	0.01	.00	13.78*	.31
Openness[d]	4.62	1.00	4.65	0.99	4.80	1.05	4.72	1.03	0.45	.00	2.23*	.04	0.87	.01	−1.68	.02
Early[e]			4.88	0.99	4.95	1.08	4.93	1.03			0.52	.00	0.18	.00	0.19	.00
Mid[f]			4.41	1.00	4.64	1.06	4.56	0.99			1.77†	.06	1.18	.03	1.83	.04
Late[g]			4.55	0.92	4.73	0.97	4.53	1.01			1.07	.04	0.01	.00	1.50	.05
Satisfaction[d]	5.46	1.31	5.33	1.40	5.68	1.25	5.54	1.38	−1.55	.02	3.04*	.06	1.95*	.03	−3.09*	.07
Early[e]			5.88	1.05	6.00	1.22	5.88	1.21			0.77	.01	0.00	.00	0.90	.01
Mid[f]			4.90	1.55	5.37	1.34	5.60	1.22			2.85*	.15	13.49*	.23	0.94	.02
Late[g]			4.89	1.41	5.52	1.03	4.79	1.63			3.27*	.26	0.24	.01	17.49*	.37
Viability[d]	5.20	0.97	5.21	0.98	5.38	1.00	5.22	1.12	0.13	.00	2.54*	.05	0.13	.00	−2.88*	.06
Early[e]			5.57	0.79	5.76	0.88	5.67	0.96			1.73†	.05	0.64	.01	2.83†	.05
Mid[f]			4.78	1.14	4.94	1.12	4.87	1.13			1.34	.04	0.48	.01	1.17	.03
Late[g]			5.12	0.80	5.26	0.73	4.84	1.10			1.34	.06	3.32†	.10	6.38*	.18
Task motivation[d]	5.40	0.75	5.31	0.71	5.43	0.74	5.47	0.82	−2.10*	.03	1.53	.02	2.88*	.06	−1.07	.01
Early[e]			5.54	0.67	5.55	0.77	5.59	0.83			0.08	.00	0.27	.00	0.08	.00
Mid[f]			5.02	0.74	5.28	0.75	5.48	0.76			3.15*	.18	47.16*	.51	0.21	.00
Late[g]			5.30	0.61	5.43	0.65	5.22	0.84			1.79†	.10	0.47	.02	6.37*	.18
Loafing[d]	3.30	1.68	3.48	1.63	3.19	1.56	3.27	1.62	2.33*	.04	2.05*	.03	−1.99*	.03	2.16*	.03
Early[e]			2.86	1.34	2.57	1.30	2.58	1.23			1.85†	.05	4.03*	.06	1.46	.02
Mid[f]			4.30	1.77	3.86	1.75	3.98	1.74			2.59*	.13	3.57†	.07	2.90†	.06
Late[g]			3.49	1.43	3.41	1.30	3.56	1.62			0.43	.01	0.04	.00	0.48	.02

Note. Overall statistics (except linear [Lin.] and quadratic [Quad.] contrasts) were computed with all four data points as dependent variables in a repeated measures analysis of variance (ANOVA). Statistics by project deadline were computed with Time 2, Time 3, and Time 4 as dependent variables in a repeated measures ANOVA with Time 1 used as a covariate in the analysis.

[a] T2 reports whether the value at Time 2 was significantly different from the value at Time 1. [b] T3 reports whether the value at Time 3 was significantly different from the average of all prior values shown in the table. [c] The linear and quadratic effects were computed with data at Time 2, Time 3, and Time 4 only. The value at Time 1 was used as a covariate in the analysis. [d] Overall statistics. [e] Early presenters had a project deadline 1–2 class periods after the peer appraisal ($n = 61$). [f] Midpresenters had a project deadline 3 class periods after the peer appraisal ($n = 46$). [g] Late presenters had a project deadline 4–5 class periods after the peer appraisal ($n = 31$).

* $p < .05$. † $p < .10$.

Table 3

Undergraduate Group (Overall and by Project Deadline; n = *115): Descriptive Statistics and Repeated Measures ANOVA A Priori Contrasts Testing for the Effect of the Peer Appraisal*

	Time 1		Time 2		Time 3		Time 4		A priori contrast							
									T2[a]		T3[b]		Lin.[c]		Quad.[c]	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*	η^2	*t*	η^2	*t*	η^2	*t*	η^2
Cohesion[d]	4.41	1.36	4.56	1.22	4.84	1.28	4.58	1.24	1.99*	.03	5.46*	.21	0.25	.00	−4.16*	.13
Early[e]			4.51	1.19	4.79	1.11	4.37	1.19			1.65	.09	0.50	.02	5.91*	.17
Mid[f]			4.59	1.25	4.86	1.35	4.65	1.26			3.40*	.12	0.77	.01	10.97*	.12
Openness[d]	4.58	0.90	4.57	0.98	4.79	0.99	4.62	0.96	−0.17	.00	3.15*	.08	0.62	.00	−3.01*	.07
Early[e]			4.84	0.88	4.88	0.98	4.64	0.92			0.28	.00	3.30†	.10	1.06	.04
Mid[f]			4.47	1.01	4.76	0.99	4.62	0.99			3.63*	.14	2.24	.03	8.79*	.10
Satisfaction[d]	5.30	1.61	5.40	1.45	5.77	1.33	5.43	1.45	1.25	.01	5.61*	.22	0.33	.00	−5.78*	.23
Early[e]			5.12	1.58	5.80	1.16	5.22	1.34			3.76*	.33	0.32	.01	22.03*	.43
Mid[f]			5.50	1.40	5.76	1.40	5.50	1.50			3.21*	.11	0.00	.00	15.42*	.16
Viability[d]	5.04	1.03	5.11	1.03	5.44	0.99	5.17	1.01	1.29	.01	5.15*	.19	0.92	.01	−4.97*	.18
Early[e]			4.74	0.99	5.41	0.72	4.87	0.91			4.20*	.38	0.64	.02	32.03*	.52
Mid[f]			5.23	1.02	5.47	1.07	5.27	1.03			3.07*	.10	0.37	.00	10.34*	.11
Task motivation[d]	5.27	0.79	5.27	0.85	5.49	0.80	5.38	0.84	−0.05	.00	5.50*	.21	1.99*	.03	−4.10*	.13
Early[e]			5.10	1.11	5.37	1.08	5.21	1.06			2.93*	.23	0.89	.03	8.81*	.23
Mid[f]			5.33	0.74	5.55	0.67	5.44	0.75			4.07*	.17	2.58	.03	12.83*	.13
Loafing[d]	3.67	1.53	3.62	1.55	3.18	1.44	3.44	1.41	0.60	.00	4.27*	.14	−1.49	.02	4.03*	.12
Early[e]			3.83	1.48	3.23	1.26	3.54	1.38			2.39*	.16	1.17	.04	5.78*	.17
Mid[f]			3.55	1.59	3.15	1.51	3.41	1.44			3.31*	.12	1.16	.01	11.68*	.12

Note. Overall statistics (except linear [Lin.] and quadratic [Quad.] contrasts) were computed with all four data points as dependent variables in a repeated measures analysis of variance (ANOVA). Statistics by project deadline were computed with Time 2, Time 3, and Time 4 as dependent variables in a repeated measures ANOVA with Time 1 used as a covariate in the analysis.

[a] T2 reports whether the value at Time 2 was significantly different from the value at Time 1. [b] T3 reports whether the value at Time 3 was significantly different from the average of all prior values shown in the table. [c] The linear and quadratic effects were computed with data at Time 2, Time 3, and Time 4 only. The value at Time 1 was used as a covariate in the analysis. [d] Overall statistics. [e] Early presenters had a project deadline 1–2 class periods after the peer appraisal (n = 30). [f] Midpresenters had a project deadline 3 class periods after the peer appraisal (n = 84).

* $p < .05$. † $p < .10$.

Table 4
Graduate Group (Overall and by Project Deadline): Descriptive Statistics and Repeated Measures ANOVA A Priori Contrasts Testing for Perceived Usefulness Over Time

										A priori contrast					
	Time 2[a]		Time 3		Time 4		Within-subject effect of time			T3[b]		Lin.[c]		Quad.[c]	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	η^2	*df*	*t*	η^2	*t*	η^2	*t*	η^2
Useful to group communication	4.81	1.00	5.42	1.00	5.07	0.96	21.24**	.23	2, 144	6.04**	.20	2.34**	.05	−6.48**	.18
Early[d]	4.86	0.87	5.48	1.01	5.02	0.90	9.77**	.24	2, 62	4.16**	.22	1.26	.02	−4.45**	.24
Mid[e]	4.76	1.22	5.38	0.94	5.30	1.03	6.93**	.22	2, 48	3.67**	.22	2.50*	.11*	−3.39**	.19
Late[f]	4.79	0.91	5.35	1.10	4.82	0.90	7.13**	.31	2, 32	3.02**	.22	0.16	.00	−3.79**	.30
Useful to group task	4.85	0.99	5.45	0.95	5.10	0.93	27.50**	.28	2, 142	6.75**	.24	2.29**	.06	−7.36**	.23
Early[d]	4.92	0.85	5.46	1.01	5.09	0.86	14.47**	.32	2, 61	5.11**	.30	1.42	.03	−5.22**	.30
Mid[e]	4.71	1.08	5.53	0.79	5.26	1.04	15.54**	.40	2, 47	5.42**	.38	2.91**	.15	−5.20**	.36
Late[f]	4.91	1.11	5.32	1.06	4.85	0.89	5.00**	.24	2, 32	1.91	.10	−0.30	.00	−2.80**	.19
Useful to self	4.98	1.27	5.03	1.45	—	—	0.26	.02	1, 170	0.51	.02	—		—	
Early[d]	5.14	1.16	5.17	1.26	—	—	0.03	.00	1, 77	0.18	.00	—		—	
Mid[e]	4.77	1.58	4.85	1.60	—	—	0.12	.00	1, 52	0.34	.00	—		—	
Late[f]	4.95	1.01	5.02	1.58	—	—	0.10	.00	1, 41	0.31	.00	—		—	

Note. Dashes indicate that data were not collected. ANOVA = analysis of variance.
[a] Time 2 represents the respondent's estimate of how useful the peer appraisal would be. [b] T3 reports whether the value at Time 3 was significantly different from the value at Time 2. [c] Calculation of linear (Lin.) and quadratic (Quad.) effects used value at Time 1 as a covariate. [d] Early presenters had a project deadline 1–2 class periods after the peer appraisal. [e] Midpresenters had a project deadline 3 class periods after the peer appraisal. [f] Late presenters had a project deadline 4–5 class periods after the peer appraisal.
* $p < .05$. ** $p < .01$.

Table 5
Undergraduate Group (Overall and by Project Deadline): Descriptive Statistics and Repeated Measures ANOVA A Priori Contrasts Testing for Perceived Usefulness Over Time

	Time 2[a]		Time 3		Time 4		Within-subject effect of time			A priori contrast					
										T3[b]		Lin.[c]		Quad.[c]	
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*F*	η^2	*df*	*t*	η^2	*t*	η^2	*t*	η^2
Useful to group communication	4.84	1.26	5.36	1.07	5.10	0.96	9.08**	.12	2, 137	4.20**	.11	3.04**	.03	−3.31**	.08
Early[d]	4.53	1.18	5.06	1.22	5.11	0.92	4.42**	.21	2, 34	2.29*	.13	2.85**	.18	−1.18	.04
Mid[e]	4.95	1.14	5.47	1.00	5.10	0.98	10.17**	.17	2, 102	4.22**	.15	1.18	.01	−4.49**	.16
Useful to group task	5.12	1.08	5.39	0.99	4.96	0.89	6.18**	.08	2, 137	2.37*	.04	−0.80	.03	−3.33**	.06
Early[d]	4.92	1.25	5.19	1.09	4.97	0.97	0.81	.04	2, 34	1.24	.04	0.25	.00	−1.25	.04
Mid[e]	5.19	1.01	5.45	0.95	4.96	0.87	14.19**	.22	2, 102	2.39*	.05	−2.23*	.05	−4.36**	.16
Useful to self	5.16	1.33	5.23	1.26	—	—	0.12	.00	1, 158	0.34	.00		—		—
Early[d]	5.02	1.35	5.02	1.26	—	—	0.00	.00	1, 41	0.00	.00		—		—
Mid[e]	5.20	1.32	5.30	1.26	—	—	0.43	.00	1, 117	−0.65	.00		—		—

Note. Dashes indicate that data were not collected. ANOVA = analysis of variance.
[a] Time 2 represents the respondent's estimate of how useful the peer appraisal would be. [b] T3 reports whether the value at Time 3 was significantly different from the value at Time 2. [c] Calculation of linear (Lin.) and quadratic (Quad.) effects used value at Time 1 as a covariate. [d] Early presenters had a project deadline 1–2 class periods after the peer appraisal. [e] Midpresenters had a project deadline 3 class periods after the peer appraisal.
* $p < .05$. ** $p < .01$.

more favorable than a score of 4.0. The above analyses support Hypotheses 2b and 2c.

The Effect of Perceived Ratio of Positive to Negative Feedback

Hypothesis 3 stated that the immediate and enduring effects of the peer appraisal on the following variables do not depend on the perceived ratio of positive to negative feedback received: (a) group open communication, (b) task motivation, (c) social loafing, (d) viability, (e) cohesion, and (f) satisfaction with group, and ratings of the usefulness of the appraisal for (g) one's own development, (h) improving open communication, and (i) improving a group's ability to accomplish its task. We used a K-means cluster analysis (Norusis, 1994, p. 333) using participants' responses to the question about the ratio of positive to negative feedback to empirically identify two clusters whose means on this variable were maximally separated (see Tables 6 and 7 for descriptive statistics and comparisons of the two clusters). The graduate-group clusters centered around 8.1 and 5.2; the undergraduate-group clusters centered around 8.1 and 5.0 (scale of 10 = *all positive* to 0 = *all negative*). We refer to the two clusters in each group as the positive feedback and balanced feedback clusters, respectively.

To test whether the immediate effect of the peer appraisal was the same for positive and balanced feedback clusters, we performed the same repeated measures ANOVA tests with a priori difference contrasts that were used to test Hypotheses 1 and 2; however, we included cluster membership as a between-subjects variable. If the effect of the peer appraisal is the same for individuals in both clusters, we would expect the interaction between the effect of the peer appraisal and cluster membership to be nonsignificant. The effect of the peer appraisal (T3 contrast) only showed a significant interaction with cluster membership for undergraduate group task motivation, $F(1, 113) = 4.70, p < .05$, $\eta^2 = .03$, and rating of usefulness of the peer appraisal for completing the group task, $F(1, 139) = 12.79, p < .01$, $\eta^2 = .10$. In both cases, the effect of the peer appraisal was more favorable for the positive feedback cluster (see Table 7). Eta squared for interactions involving viability and loafing in the undergraduate group and group task motivation in the graduate group was .02. Eta squared for all other interactions was .01 or less.

To test whether the enduring effect of the peer appraisal (i.e., Time 4) was the same for the positive and balanced feedback clusters, we performed repeated measures ANOVAs with orthogonal a priori contrasts that compared Time 4 (the enduring effect of the peer appraisal) with Time 3 (the immediate effect), and Time 4 (the enduring effect) with Time 2 (prefeedback). Significant interaction between the enduring effect of the peer appraisal and cluster membership was found only for the contrasts comparing Time 4 to Time 3 for cohesion, $F(1, 145) = 6.08, p < .01, \eta^2 = .04$, and openness, $F(1, 145) = 5.19, p < .05, \eta^2 = .035$, in the graduate group. In both of these cases, the value of the variable in the balanced feedback cluster increased from Time 3 to Time 4 whereas the positive feedback cluster decreased (see Table 6). Eta squared for all other interactions involving Time 4 in both graduate and undergraduate groups was .01 or less. This analysis provides support for Hypothesis 3 with the exception of Hypotheses 3b and 3i in the undergraduate group and 3a and 3e in the graduate group.

Table 6
Graduate Group: Descriptive Statistics and t *Tests Comparing Positive and Balanced Feedback Clusters*

	Time 2[a] (n = 173)					Time 3 (n = 193)					Time 4 (n = 160)				
	Positive (n = 130)		Balanced (n = 43)			Positive (n = 146)		Balanced (n = 47)			Positive (n = 123)		Balanced (n = 37)		
Variable	*M*	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*
Useful to group communication	4.88	0.97	4.43	1.11	2.56**	5.46	0.98	5.13	0.87	2.06*	5.16	0.94	4.81	0.94	1.99*
Useful to group task	4.94	0.94	4.49	0.93	2.73**	5.49	0.96	5.19	0.85	1.92	5.13	0.91	4.92	0.89	1.24
Useful to self	5.11	1.29	4.60	1.14	2.27*	5.16	1.39	4.60	1.39	2.40*	—	—	—	—	—
Cohesion	5.03	1.15	4.56	1.18	2.28*	5.24	1.12	4.62	1.10	3.33**	5.06	1.36	4.77	0.97	1.24
Openness	4.80	0.97	4.39	0.93	2.43*	4.98	0.96	4.45	1.06	3.22**	4.80	1.04	4.54	0.90	1.38
Satisfaction	5.54	1.35	5.20	1.31	1.41	5.85	1.24	5.48	1.07	1.84	5.61	1.43	5.39	1.12	0.85
Viability	5.39	0.96	5.07	0.96	1.89	5.49	1.00	5.42	0.84	0.44	5.30	1.14	5.14	1.04	0.78
Group Task Motivation	5.37	0.70	5.21	0.63	1.29	5.53	0.72	5.25	0.74	2.35**	5.51	0.86	5.32	0.63	1.24
Loafing	3.27	1.63	3.39	1.45	−0.42	3.04	1.58	3.11	1.28	−0.30	3.21	1.63	3.21	1.57	0.00

Note. Dashes indicate that data were not collected.
[a] These numbers represent differences between the two groups prior to the peer appraisal.
* $p < .05$. ** $p < .01$.

Table 7
Undergraduate Group: Descriptive Statistics and t *Tests Comparing Positive and Balanced Feedback Clusters*

	Time 2[a] (n = 163)					Time 3 (n = 204)					Time 4 (n = 163)				
	Positive (n = 121)		Balanced (n = 42)			Positive (n = 151)		Balanced (n = 53)			Positive (n = 116)		Balanced (n = 47)		
Variable	*M*	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*	*M*	*SD*	*M*	*SD*	*t*
Useful to group communication	4.81	1.24	5.05	0.91	−1.14	5.42	1.10	5.28	1.18	0.75	5.11	0.93	5.15	1.12	−0.22
Useful to group task	5.04	1.11	5.31	1.09	−1.35	5.50	1.03	5.19	1.07	1.87	5.05	0.87	4.91	1.04	0.86
Useful to self	5.15	1.38	5.21	1.16	−0.27	5.24	1.24	5.21	1.42	0.16	—	—	—	—	—
Cohesion	4.66	1.13	4.32	1.23	1.63	5.01	1.15	4.51	1.49	2.51*	4.64	1.13	4.24	1.44	1.89
Openness	4.49	0.98	4.77	0.93	−1.59	4.81	0.99	4.95	1.02	−0.88	4.59	0.96	4.90	1.01	−1.80
Satisfaction	5.58	1.35	5.09	1.51	1.95*	5.87	1.23	5.42	1.50	2.17*	5.52	1.30	5.12	1.42	1.71
Viability	5.18	1.04	5.14	0.92	0.22	5.54	0.92	5.35	0.96	1.23	5.15	0.95	5.10	1.06	0.34
Group task motivation	5.32	0.77	5.15	0.93	1.15	5.58	0.71	5.25	0.92	2.71**	5.47	0.72	5.30	0.94	1.29
Loafing	3.43	1.56	3.68	1.31	−0.93	3.03	1.36	3.21	1.28	−0.85	3.47	1.46	3.49	1.36	−0.06

Note. Dashes indicate that data were not collected.
[a] These numbers represent differences between the two groups before the peer appraisal.
* $p < .05$. ** $p < .01$.

Effects of Timing on Peer Appraisal

Hypothesis 4 suggested that the enduring effect of the peer appraisal on perceptions of (a) group open communication, (b) task motivation, (c) social loafing, (d) viability, (e) cohesion, and (f) satisfaction with group, vary depending on timing of the project deadline. To test this hypothesis, we performed a repeated measures ANOVA analysis using a polynomial contrast (see Tables 2 and 3). The polynomial contrast provides an indication of the linear and quadratic trends of a variable over time.

The strongest indication of a lasting effect is a significant linear effect and a nonsignificant quadratic effect (strong, lasting effect). The second strongest lasting effect is both significant linear and quadratic effects (moderate lasting effect). Because these two effects could indicate either a significant improvement or worsening of a variable, the mean at Time 4 must be compared to the mean at Time 2 to determine the direction of the effect. A nonsignificant linear effect and a significant quadratic effect indicates a nonlasting effect. If both effects are nonsignificant, then there is no effect.

For this analysis, because we have already ruled out maturation effects, we did not need the Time 1 measurement. Because the measures of any variable were highly intercorrelated over time (typically > .70), we used the value of the variable at Time 1 as a covariate in the ANOVA model (see Freed, Ryan, & Hess, 1991).

Figure 2 shows graphs of each variable by project deadline (see Tables 2 and 3 for descriptive statistics). Only the graduate group was included in this analysis because the undergraduate group did not know the project deadline at the time of the peer appraisal (Time 3). However, because of that fact, the undergraduate group served as a comparison group that was not expected to show timing effects. With the exception of openness, there were no timing related differences in the strength of effects for undergraduates.

For all variables in the graduate group except openness, the strength of the effect of the peer appraisal depended on timing. With the exception of loafing, significant (or approaching significant, $p < .10$) linear effects coupled with nonsignificant quadratic effects, indicating a strong, lasting effect from the peer appraisal, only occurred with the midpresenters.

The effect of timing was further supported by group member perception of the usefulness of the peer appraisal. A moderately lasting perception that the peer appraisal was more helpful than originally expected was seen only in the midpresenters (see Table 4). Immediately after the peer appraisal (Time 3), the early presenters' and late-presenters' evaluation of its usefulness increased, but by Time 4 their evaluation did not remain significantly higher than initial expectations. This analysis supports Hypotheses 4b through 4f.

Discussion

The primary purpose of this research was to examine the immediate and enduring impact of a developmental peer appraisal on SMWGs. The major contributions of the study fall into three categories. First, the study reveals that a structured, face-to-face developmental peer appraisal can have an immediate positive impact on group member perceptions of open communication, group task focus, group viability, and member relationships. Second, the study reveals the importance of the timing of a peer appraisal in determining its lasting impact. As such, it supports and extends current knowledge about the role of temporal context in group process and development. Third, the study builds on previous research to further knowledge about

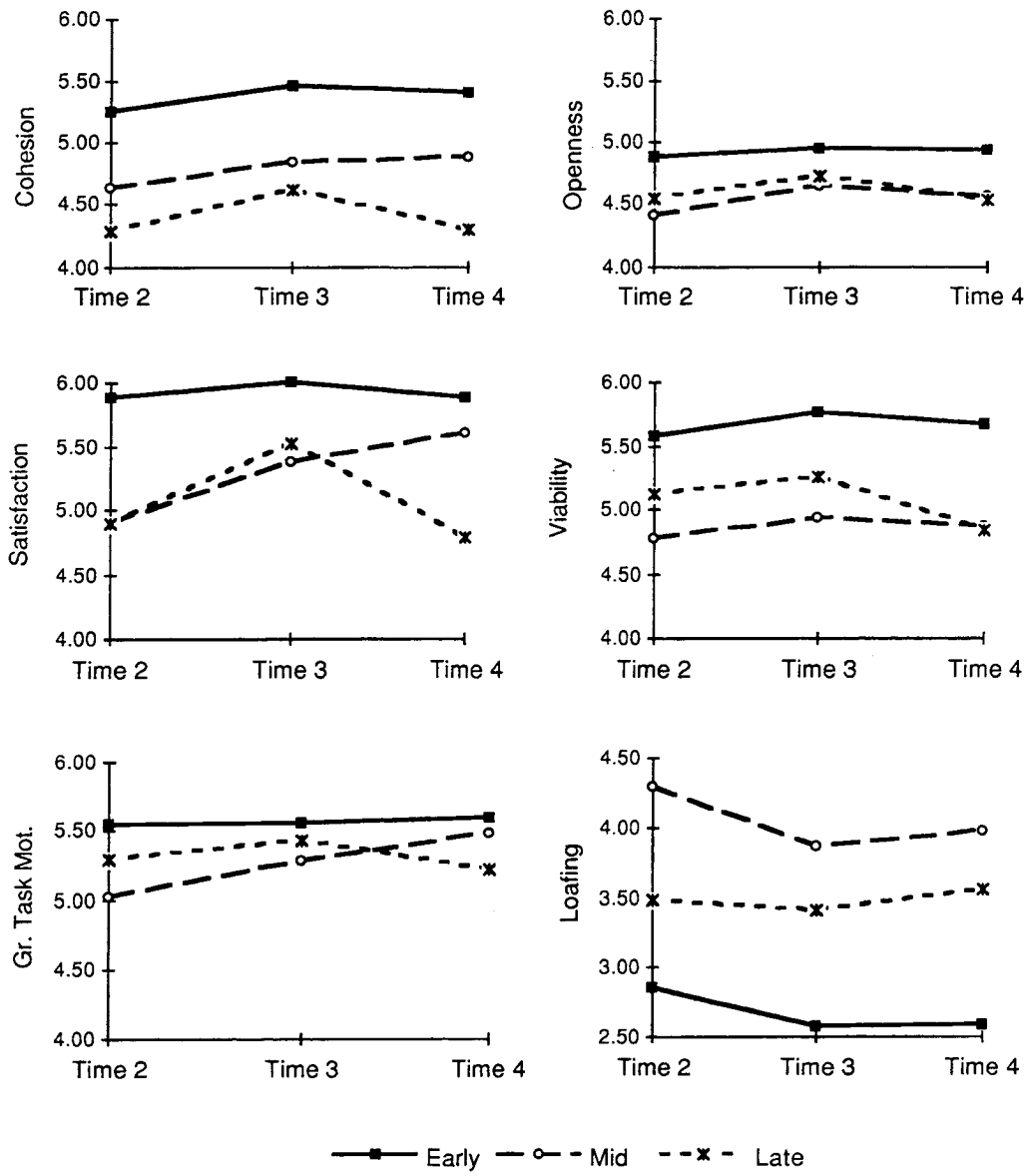

Figure 2. Graphs of dependent variables over time for graduate group, broken down by project deadline. Early presenters represent groups that presented the class session before Time 4. Mid-presenters represent groups that presented at Time 4. Late presenters represent groups that presented the class session after Time 4. Time 2 was 1 week before Time 3 (immediately after the peer review). Time 4 was three class sessions after Time 3. Descriptive statistics can be found in Tables 2 and 3. Gr. Task Mot. = group task motivation.

designing group peer appraisal systems that have a long-term developmental impact on SMWGs.

Immediate Impact of the Peer Appraisal

The present study provides evidence to suggest that peer appraisals can have an immediate positive effect on group members and on their perceptions of open group communication, group task focus, and group viability. The strength of this finding is supported by the fact that it occurred in both the undergraduate and graduate-level groups. The only construct that did not increase above predicted maturation levels after the peer appraisal was cohesion in the undergraduate group. The present results contradict the idea (Kane & Lawler, 1978; Love, 1981) that peer appraisals will harm relationships and impair group task focus and functioning. They also contradict previous research that found even positive peer rating feedback did not significantly increase ratings of member satisfaction or cohesiveness

(DeNisi et al., 1983). It is important to note that the present study used a developmental appraisal system in which group members were given control over defining behavioral expectations. Also, the full-group, face-to-face feedback process permitted the opportunity to seek clarification and discuss feedback.

The present findings add to our knowledge about the impact of individual-level feedback on groups and their members. In his review of the literature on the effects of feedback, Nadler (1979) concluded that only positive group-level feedback administered by an agent external to a team could result in increased attraction to the team and increased task motivation. The present study reveals that this result can be obtained through the use of a developmental peer appraisal in which the full group is engaged in the feedback. This finding has implications for SMWGs responsible for managing their own performance—it reveals a positive effect when group members provide their own feedback and reinforcement. Indeed, research indicates that the most effective SMWGs are skilled at their own observation, evaluation, and reinforcement (Druskat, 1997; Manz & Sims, 1987; McIntyre & Salas, 1995).

The finding that peer appraisals can have a positive impact on a group and its members is strengthened by the results of Hypotheses 2 and 3. Here, group members rated the peer appraisal as having been useful to their own development, useful for increasing open communication within the group, and useful to the group's ability to accomplish its task. A week prior to the peer appraisal, members had rated their expectations on these dimensions. Interestingly, although ratings after the peer appraisal indicated that it had been as useful to individual personal development as expected, ratings of its anticipated usefulness in fostering group open communication and in improving the group's ability to accomplish its task were significantly higher than expectations. This finding emphasizes the value of the peer appraisal for group and individual development and reinforces that its impact on group development is not intuitive.

The investigation into Hypothesis 3 revealed important information about group members who received different perceived ratios of positive and negative feedback. First, it revealed that overall feedback tended to be perceived as positive. Second, the time series design permitted us to look at differences between the balanced and positive groups prior to the appraisal, immediately after the appraisal, and three class periods after the appraisal. Most research on the effects of feedback measures it one time (see Cusella, 1987; DeNisi et al., 1983; Nadler, 1979). Results showed that some differences existed between the members of the two groups prior to the appraisal. Well before the appraisal, members of the balanced feedback graduate group felt lower levels of attraction to their groups, perceived less open communication, and were less optimistic about the upcoming appraisal. Thus, the logic that receiving negative comments in a peer appraisal may harm a member's attitude toward his or her group may be misplaced. Peer appraisals may do more good than harm by providing a structured way for groups to address their disengaged members and also for those members to voice their feelings and views. Indeed, the present research shows that although those members were not feeling much better directly after the peer appraisal, three class periods later their ratings of cohesion and open communication in their groups had improved significantly more than those of the individuals in the positive feedback group. In essence, their feelings about their group and its processes seemed to "catch up" with those of the positive feedback group. This may have been facilitated by a combination of participating in the feedback process and the change in their own or others' behavior as a result of receiving the feedback.

Three conclusions are warranted. First, an important aspect of the peer appraisal system used in this study may have been the opportunity for the appraisee to discuss and question the feedback. Second, when a developmental peer appraisal system is used, assessing the longer term impact of the appraisal may be more informative than assessing its immediate effect. Third, groups lower in open communication, task focus, viability, and cohesion may generate more negative feedback during a peer appraisal; however, bringing this out through a peer appraisal may be a useful way to help the group address its problems.

Temporal Context and the Impact of the Peer Appraisal

This study was designed to examine whether the effect of peer appraisal feedback endured over time. Most research on peer appraisals has not been designed to assess the potential of such appraisals as a feedback tool, but rather, has focused on their accuracy (Farh et al., 1991; Fox et al., 1989; Love, 1981; Saavedra & Kwun, 1993), or acceptability (Cederblom & Lounsbury, 1980; Farh et al., 1991; Love, 1981; McEvoy & Buller, 1987). As such, research designs have (a) made peer ratings confidential and private so that only researchers see the ratings (Fox et al., 1989; Saavedra & Kwun, 1993), (b) involved distributing bogus ratings to team members (DeNisi et al., 1983; Stone, 1971), and (c) distributed peer appraisals at the end of a group's tenure, when feedback could have no influence on group processes (e.g., Farh et al., 1991; Fox et al., 1989; Saavedra & Kwun, 1993).

As was hypothesized, the present study found that enduring development as a result of the appraisal depends on the timing of the peer appraisal relative to a group's project deadline. Results indicated very different outcomes for groups whose deadlines were three classes after the peer appraisal (midpresenters) than groups whose deadlines were one to two classes after the appraisal (early presenters) or

four to five classes after the appraisal (late presenters). The midpresenters experienced enduring effects of the peer appraisal, above and beyond predicted maturation levels, for four of the six constructs studied.

These findings suggest the importance of temporal context on the effect of peer appraisals. They support the idea that there exists a period of time during a task group's life when the group and its members are most likely to develop, learn from, and change as a result of an externally imposed intervention like a peer appraisal. It appears that groups may receive the most enduring benefits from a peer appraisal if it is conducted before the group is engaged in project execution, yet not so far in advance of this period that the project is not yet a priority. Groups that are already engaged in executing their task, as perhaps the early presenters were at the time of the peer appraisal, may already have laid out their task plan and negotiated expected contributions making the appraisal less valuable. Indeed, the early presenters did not experience an enduring effect from the peer appraisals. On the other hand, the late presenters may have been too far from their deadline to have been ready to use the appraisal most effectively. Given the multiple demands on the attention of MBA students, a project deadline four to five classes away may not yet have been a high priority for the late presenters. Results indicated that when the peer appraisal was conducted three classes before the project deadline, the group derived the greatest benefit from the appraisal. According to Tyre et al. (1996), temporal interruptions, such as that provided by the peer appraisal, permit the time for reflection necessary for engaging in change. The midpresenters may have had both the time for reflection and the sense of urgency necessary to motivate change.

A striking and unexpected finding was that project deadline, which had been assigned by course instructors, not only made a difference in the enduring impact of the peer appraisal, it also had a consistent effect on the absolute magnitude of member ratings of all constructs studied (see Figure 2). Early presenters consistently had the highest ratings in satisfaction, cohesion, task motivation, and group viability and lowest ratings in social loafing. This finding underscores the importance of time as a contextual factor in groups. Gersick (1989) found that the sharper a group's time constraints, the more frequently members paid explicit attention to time and pacing. The present results imply that perceived sharpness of time constraints can also speed up and facilitate the development of task motivation and focus, group viability, and member relationships.

Limitations and Future Research

This research has limitations that should be noted. First, the context in which we conducted our field study did not allow us to create or find equivalent control groups. The use of control groups for Hypothesis 4 would increase our confidence that it was the timing of the peer appraisal rather than maturation or goal proximity that influenced the outcomes. The undergraduate sample that did not know their task deadline at the time of the appraisal provided a level of confidence in our findings; however, they were not an equivalent control group.

The present research examined groups who experienced the peer appraisal at three different periods relative to their task deadlines. Although differences were found between the three groups, the research did not attempt to investigate the exact process or task issues that the groups with different deadlines were experiencing. Future research that examines the impact of timing on peer appraisals should obtain a better understanding of the issues most salient for the groups that receive the most lasting benefit from the peer appraisal. This would permit the tailoring of the appraisal to most directly meet those needs.

Future research should also go a step further by longitudinally following the behavioral and performance changes that result from a peer appraisal process. Research that follows groups for an extended period of time measuring behaviors and performance at multiple points before and after peer appraisals would be most useful for answering this question. Such research could more closely assess the immediate and enduring benefits of the peer appraisal and how they impact group performance.

Our study, like several others examining peer appraisals before us (e.g., DeNisi et al., 1983; Farh et al., 1991; Saavedra & Kwun, 1993; Stone, 1971) was conducted with students in an educational context. We believe our results are generalizable to SMWGs in work settings because of similarities between the two contexts on key attributes such as the high level of member responsibility, interdependence, and task involvement required in both settings. However, the external validity of the findings reported here must be tested in work settings.

The purpose of the peer appraisal in the present study was purely developmental. Clearly, there are reasons why an organization may need to evaluate its team members. Future research should assess the impact of a similar structured peer appraisal process used for evaluation. It would be informative to determine the aspects of the process and outcome most affected by the developmental or evaluative nature of a peer appraisal; for example, what would the impact be on the selection of behaviors to evaluate and on the face-to-face appraisal discussion?

Conclusion

Two decades ago, DeNisi and Mitchell (1978) claimed that advocates of peer assessments were too willing to ignore their problems. The results of the present study suggest that, because prior studies have ignored context, they have underestimated the potential benefits of peer

appraisals. By using the knowledge learned from research, we can begin designing group peer appraisal systems that have a positive long-term developmental effect on SMWGs.

The present study used a developmental purpose and features such as prompting appraisers for both negative and positive feedback, pooling feedback from multiple appraisers, and face-to-face discussion of the feedback to increase its reliability, validity, and user acceptance. More important, this study shows that peer appraisals can be designed to have a positive impact on group open communication, task focus, and member relations. The study also suggests that timing of the appraisal is important for ensuring a longer term developmental consequence. As research continues to expose the influence of performance appraisal context and design, it becomes increasingly clear that structured peer appraisal processes that take context into account are desirable and necessary.

References

Allport, G. W. (1954). *The nature of prejudice.* Reading, MA: Addison-Wesley.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Boston: Houghton Mifflin.

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

Conlon, E. J., & Barr, S. H. (1989). A framework for understanding group feedback. In E. J. Lawler & B. Markovsky (Eds.), *Advances in group processes* (Vol. 6, pp. 27–48). Greenwich, CT: JAI Press.

Cusella, L. P. (1987). Feedback, motivation and performance. In F. M. Jablin, L. L. Putnam, K. H. Roberts, & L. W. Porter (Eds.), *Handbook of organizational communication: An interdisciplinary perspective* (pp. 624–678). Newbury Park, CA: Sage.

DeNisi, A. S., & Mitchell, J. L. (1978). An analysis of peer ratings as predictors and criterion measures and a proposed new application. *Academy of Management Review, 3,* 369–374.

DeNisi, A. S., Randolph, W. A., Blencoe, A. G. (1983). Potential problems with peer ratings. *Academy of Management Journal, 26,* 457–464.

Druskat, V. U. (1997). *Team-level competencies in self-managing work teams: Behaviors, norms and processes that make a difference.* Unpublished Working Paper No. 97-2(2a). Department of Organizational Behavior, Case Western Reserve University, Cleveland, Ohio.

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.

Fiske, S., & Taylor, S. (1991). *Social cognition.* New York: McGraw-Hill.

Fox, S., Ben-Nahum, Z., & Yinon, Y. (1989). Perceived similarity and accuracy of peer ratings. *Journal of Applied Psychology, 74,* 781–786.

Freed, M. N., Ryan, J. M., & Hess, R. K. (1991). *Handbook of statistical procedures and their computer applications to education and the behavioral sciences.* New York: Macmillan.

Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. *Academy of Management Journal, 31,* 9–41.

Gersick, C. J. G. (1989). Marking time: Predictable transitions in task groups. *Academy of Management Journal, 32,* 274–309.

Gersick, C. J. G., & Hackman, J. R. (1990). Habitual routines in task-performing groups. *Organizational Behavior and Human Decision Processes, 47,* 65–97.

Hackman, J. R. (1986). The psychology of self-management in organizations. In M. S. Pallack & R. O. Perloff (Eds.), *Psychology and work: Productivity, change, and employment* (pp. 315–342). Englewood Cliffs, NJ: Prentice Hall.

Hackman, J. R. (1988). *Flight Crew Questionnaire.* Cambridge, MA: Author.

Harkins, S. G., & Jackson, J. J. (1985). The role of evaluation in eliminating social loafing. *Personality and Social Psychology Bulletin, 11,* 457–465.

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

Judge, T. A., & Ferris, G. R. (1993). Social context of performance evaluation decisions. *Academy of Management Journal, 36,* 80–105.

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

Katz, D., & Kahn, R. L. (1978). *The social psychology of organizations* (2nd ed.). New York: Wiley.

Lewin, A. Y., & Zwany, A. (1976). Peer nominations: A model, literature critique and a paradigm for research. *Personnel Psychology, 29,* 423–447.

Locke, E. A. (1986). *Generalizing from laboratory to field settings.* Lexington, MA: Lexington Books.

Love, K. G. (1981). Comparison of peer assessment methods: Reliability, validity, friendship bias, and user reaction. *Journal of Applied Psychology, 66,* 451–457.

Luft, J. (1984). *Group processes: An introduction to group dynamics* (3rd ed.). Mountain View, CA: Mayfield.

Manz, C. C., & Sims, H. P., Jr. (1987). Leading workers to lead themselves: The external leadership of self-managed work teams. *Administrative Science Quarterly, 32,* 106–128.

Martell, R. F., & Borg, M. R. (1993). A comparison of the behavioral rating accuracy of groups and individuals. *Journal of Applied Psychology, 78,* 43–50.

McEvoy, G. M., & Buller, P. F. (1987). User acceptance of peer appraisals in an industrial setting. *Personnel Psychology, 40,* 785–797.

McGrath, J. E. (1984). *Groups: Interaction and performance.* Englewood Cliffs, NJ: Prentice Hall.

McGrath, J. E. (1991). Time, interaction, and performance (TIP): A theory of groups. *Small Group Research, 22,* 147–174.

McGrath, J. E., & O'Connor, K. M. (1996). Temporal issues in work groups. In M. A. West (Ed.), *Handbook of work group psychology* (pp. 25–52). New York: Wiley.

McIntyre, R. M., & Salas, E. (1995). Measuring and managing for team performance: Lessons from complex environments. In

R. A. Guzzo & E. Salas (Eds.), *Team effectiveness and decision making in organizations* (pp. 9–45). San Francisco: Jossey-Bass.

Mitchell, T. R. (1983). The effects of social, task, and situational factors on motivation, performance, and appraisal. In F. Landy, S. Zedeck, & J. Cleveland (Eds.), *Performance measurement and theory* (pp. 39–59). Hillsdale, NJ: Erlbaum.

Mohrman, A. M., Jr., Resnick-West, S. M., & Lawler, E. E. (1989). *Designing performance appraisal systems: Aligning appraisals and organizational realities.* San Francisco: Jossey-Bass.

Moreland, R. L., & Levine, J. L. (1982). Socialization in small groups: Temporal changes in individual–group relations. *Advances in Experimental Social Psychology, 15,* 137–192.

Moreland, R. L., & Levine, J. L. (1984). Role transitions in small groups. In V. L. Allen & E. van de Vliert (Eds.), *Role transitions: Explorations and explanations* (pp. 181–195). New York: Plenum.

Moreland, R. L., & Levine, J. L. (1988). Group dynamics over time: Development and socialization in small groups. In J. McGrath (Ed.), *The social psychology of time: New perspectives* (pp. 151–181). Newbury Park, CA: Sage.

Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal.* Needham Heights, MA: Allyn & Bacon.

Nadler, D. A. (1979). The effects of feedback on task group behavior: A review of the experimental research. *Organizational Behavior and Human Performance, 23,* 309–338.

Norusis, M. J. (1994). *SPSS professional statistics 6.1.* Chicago: SPSS.

Peiperl, M., & Abelson, M. (1995, August). *Making peer evaluation work in practice: A field experiment.* Paper presented at the 55th annual meeting of the Academy of Management, Vancouver, British Columbia, Canada.

Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology, 78,* 113–123.

Rodgers, T. (1995, August). *Evaluating performance in self-managed work teams: An assessment of peer appraisal.* Paper presented at the 55th annual meeting of the Academy of Management, Vancouver, British Columbia, Canada.

Rogers, C. R., & Farson, R. E. (1995). Active listening. In D. A. Kolb, J. S. Osland, & I. M. Rubin (Eds.), *The organizational behavior reader* (6th ed., pp. 203–214). New York: Wiley.

Saavedra, R., & Kwun, S. K. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology, 78,* 450–462.

Shore, T. H., Shore, L. M., & Thornton, G. C. (1992). Construct validity of self and peer evaluations of performance dimensions in an assessment center. *Journal of Applied Psychology, 77,* 42–54.

Stokes, J. P. (1983). Components of group cohesion: Intermember attraction, instrumental value, and risk taking. *Small Group Behavior, 14,* 163–173.

Stone, T. H. (1971). Effects of mode of organization and feedback level on creative task groups. *Journal of Applied Psychology, 55,* 324–330.

Tyre, M., Perlow, L., Staudenmayer, N., & Wasson, C. (1996, August). *Time as a trigger for organizational change.* Paper presented at the 56th annual meeting of the Academy of Management, Cincinnati, OH.

Wexley, K. N., & Klimoski, R. (1984). Performance appraisal: An update. In K. M. Rowland & G. R. Ferris (Eds.), *Research in personnel and human resources management* (Vol. 2, pp. 35–79). Greenwich, CT: JAI Press.

Wolff, S. B. (1998). Managing group performance: Peer review and feedback. In J. W. Wholberg, G. Gilmore, & S. B. Wolff (Eds.), *OB in action: Cases and exercises* (Vol. 5, pp. 285–299). Boston: Houghton Mifflin.

Yammarino, F. J., & Waldman, D. A. (1993). Performance in relation to job skill importance: A consideration of rater source. *Journal of Applied Psychology, 78,* 242–249.

Zaccaro, S. J., & McCoy, M. C. (1988). The effects of task and interpersonal cohesiveness on performance of a disjunctive group task. *Journal of Applied Social Psychology, 18,* 837–851.

Received June 2, 1997
Revision received June 1, 1998
Accepted June 3, 1998 ■