PERSONNEL PSYCHOLOGY
2002, 55

EFFECTS OF ROTATED LEADERSHIP AND PEER EVALUATION ON THE FUNCTIONING AND EFFECTIVENESS OF SELF-MANAGED TEAMS: A QUASI-EXPERIMENT

AMIR EREZ, JEFFREY A. LEPINE, HEATHER ELMS
Department of Management
Warrington College of Business Administration
University of Florida

In a quasi-experiment of 38 self-managed undergraduate teams, we examined the effects of team designs that differed with respect to the form of member evaluation and team leadership. Relative to teams that relied on external evaluations, teams with peer evaluations had higher levels of workload sharing, voice, cooperation, performance, and member satisfaction. Relative to teams that relied on leader emergence, teams that rotated leadership among members had higher levels of voice, cooperation, and performance. Overall, results of the study demonstrate the potential importance of team-design decisions in self-managed teams.

One of the most common team types found in organizations is the self-managed or empowered team (Lawler, Mohrman, & Ledford, 1995). In self-managed teams, decision-making authority concerning the specific means of accomplishing the team's work is left up to the individuals who compose the team. However, as Mohrman, Cohen, and Mohrman (1995) note, although self-management reflects the capability to determine how team goals are achieved, self-managed teams can be designed differently, and decisions regarding team designs have implications with respect to team functioning and effectiveness. Unfortunately, there has not been much research on these implications. This study attempts to address them by investigating the effects of two design elements.

The first design element we investigate concerns the manner in which team members' performance is evaluated and rewarded in the absence of an individual internal to the team (e.g., leader, supervisor, manager) whose responsibility it is to perform this function. Specifically, we examine effects of peer evaluations. On the one hand, peer evaluations

The authors thank Christine Jackson, Frank Goeddeke, Chris Purdy, and Caroline Mae for their assistance with this research.

Correspondence and requests for reprints should be addressed to Amir Erez, Department of Management, Warrington College of Business Administration, University of Florida, Gainesville, FL 32611-7165; ereza@ufl.edu.

are not well accepted in organizations because it is assumed that they are inconsistent with the goal of promoting a cooperative team climate (Kiesler & DeSalvo, 1976; Saavedra & Kwun, 1993). On the other hand, there are several theoretical and practical reasons to expect that the use of peer evaluations should promote team functioning and effectiveness (e.g., Kane & Lawler, 1978; Murphy & Cleveland, 1991; Shepperd, 1993; Wexley & Klimoski, 1984). For example, there has been research demonstrating the usefulness of peer ratings for development (Druskat & Wolff, 1999). However, effects of peer ratings for reward purposes have not yet been studied.

The second design element we investigate concerns the nature of team leadership. Although it is generally thought that empowerment through self-management includes reliance on emergent leadership and the team's ability to allocate leadership responsibilities among members consensually, there may be benefits to alternative designs. For example, designating a leader may help ensure that the critical team management functions noted above are accomplished and that members have a sense of their place on the team. Moreover, if the *leadership responsibilities are rotated* among members, a climate of shared leadership may be fostered and this should promote the overall capacity of the team to function and perform effectively (Mohrman et al., 1995, p. 150). Unfortunately, there has been no empirical research investigating the effects of rotated leadership on team functioning and effectiveness.

In summary, our purpose is to investigate effects of peer evaluations and rotated leadership in self-managed teams. We investigate the effects of these designs on three team processes (workload sharing, voice, and cooperation) and two aspects of team effectiveness (task performance and member satisfaction). Before outlining our expectations regarding effects of team designs, however, we will define the focal team processes and describe how they relate to team effectiveness.

Dimensions of Team Processes and Effectiveness

Our research is framed using the input–process–output heuristic (McGrath, 1964). According to this heuristic, team inputs (such as the team's design) influence team effectiveness (i.e., performance, member satisfaction) through team processes. Team processes reflect the nature of the team's functioning, and can be captured by constructs such as workload sharing, voice, and cooperation.

Workload sharing reflects the extent to which members of a team do a fair share of the team's work. Logically, team performance should be higher when members do not shirk (i.e., do their fair share), and empirical research has verified this (e.g., Barrick, Stewart, Neubert, & Mount,

1998; Campion, Medsker, & Higgs, 1993, Campion, Papper, & Medsker, 1996) in several different team settings. Moreover, because doing a fair share of the team's work maintains equity norms, social responsibility norms, and norms of reciprocity (Kerr, 1983), team members' satisfaction should be higher in teams where workload sharing is high.

Voice reflects the extent to which people speak up and offer constructive suggestions for change (LePine & Van Dyne, 2001; Van Dyne & LePine, 1998). Although there has not been any research linking levels of voice on a team to the team's performance, there are reasons to expect a positive relationship. Perhaps most obviously, at least some of the suggestions for change are likely to result in improvements (e.g., increased efficiency) in the way the team does its work (LePine & Van Dyne, 1998). The level of voice in a team should also be positively related to member satisfaction because people tend to be more satisfied when they are able to express ideas and suggestions (Folger, 1977; Lind & Tyler, 1988).

Cooperation refers to the quality of interaction among members of a team. There have been many studies of team cooperation, and this research generally demonstrates that cooperation promotes team performance (Campion et al., 1993, 1996; Deutsch, 1949; Leavitt, 1951). The positive relationship between cooperation and team performance is generally attributed to the notion that cooperation promotes the integration of members' task focused inputs (Hackman, 1987; LePine, Hanson, Borman, & Motowidlo, 2000). The empirical literature also suggests that cooperation and member satisfaction are positively related, albeit the relationship may be stronger in teams that do complex knowledge work than in teams that do other types of work (Campion et al., 1993; 1996). The rationale for the positive relationship between cooperation and member satisfaction is that individuals' social needs are satisfied when interpersonal relationships are positive.

Team Designs

In this section, we describe reasons to expect peer evaluations and rotated leadership to be associated with higher workload sharing, voice, cooperation, and through these effects (as described above), higher team performance and member satisfaction.

Peer evaluations. Organizations resist peer evaluations because peers are thought to be uncomfortable in the role of the rater when there are material consequences. Peer raters are believed to be unwilling to differentiate among members for fear of damaging interpersonal relationships and the team's social climate (Liden & Mitchell, 1983; Mohrman & Lawler, 1983; Naiper & Latham, 1986). However, there are several reasons to believe that peer evaluations for reward purposes can actu-

ally promote the functioning of a team relative to situations where an external manager provides the evaluations.

First, individuals tend to put forth less effort when working on a group task than when working on an individual task (Jackson & Williams, 1985; Kerr, 1983; Latané, Williams, & Harkins, 1979; Ringelman, 1913). Moreover, this "social loafing" effect appears to be greatest when individuals do not perceive that their efforts are being evaluated and rewarded (Kerr, 1983; Kerr & Bruun, 1981; Shepperd, 1993, Szymanski & Harkins, 1987). Therefore, one way to reduce social loafing, or in other words, increase workload sharing, is to use a team design that increases members' perceptions that their behavior is being monitored and that there are consequences for their behavior. Relative to evaluations from external managers, who may only have limited contact with the team, evaluations from peers would seem to be well suited to promote these perceptions.

Hypothesis 1a: The use of peer evaluations in self-managed teams will be positively associated with workload sharing.

Second, peer evaluations are a form of communication that requires members to think about and assess other members' contributions. One consequence of this evaluative process is that it may cue thoughts about alternative ways of going about the team's task. Assuming that at least some of these thoughts translate into suggestions about new ways of approaching the task (Borman & Motowidlo, 1993; Borman, Motowidlo, Rose, & Hanser, 1985), voice should be promoted (LePine & Van Dyne, 1998). In addition, individuals being evaluated by peers have an incentive to make contributions that are not only valued by peers, but also visible to the peers, and voice is a visible behavior.

Hypothesis 1b: The use of peer evaluations in a self-managed team will be positively associated with the level of voice on the team.

Third, given that peer evaluations make members accountable to one another and may have material consequences, team members should be less likely to want to appear to be disagreeable or unsupportive. Instead, peers will want to be perceived as team players, and therefore, interactions will be more likely to reflect the type of courtesy and thoughtfulness that are characteristic of cooperative teams. The linkage between the use of peer evaluations and team cooperation has been demonstrated by Druskat and Wolff (1999), who found that peer evaluations had a positive impact on team members' perceptions of open communication, group viability, and member relationships—concepts all closely related to team cooperation.

Hypothesis 1c: The use of peer evaluations in self-managed teams will be positively associated with cooperation.

As discussed in the earlier section, Dimensions of Team Performance and Effectiveness, we also expect that the positive effect of peer evaluations on workload sharing, voice, and cooperation should translate to higher levels of team effectiveness. Thus:

Hypothesis 1d: The use of peer evaluations in self-managed teams will be positively associated with team performance.

Hypothesis 1e: The use of peer evaluations in a self-managed team will be positively associated with the member satisfaction.

Rotated leadership. The design of the leadership role in a self-managed team can take several forms. For example, self-managed teams could rely on a member or members to step forward and carry out leadership functions. Reliance on this type of leader emergence is often used in self-managed teams, and is assumed to be appropriate for at least two reasons. First, it is possible that a leader who eventually emerges through some natural selection process will be the most qualified to lead and carry out leadership functions (the cream rises to the top). Second, it is possible that the members who are actually doing the work are in the best position to determine who should carry out leadership responsibilities. However, we suggest that rotated leadership has benefits in terms of promoting team functioning over teams that rely on leader emergence.

A downside of leader emergence is that the nonemergent leader members may feel less responsibility for team outcomes. Nonleader members may come to rely on the emergent leader to carry out many responsibilities that members themselves could otherwise accomplish. In addition, the emergent leader may eventually feel less responsible to an external manager for their team's performance, given that the leadership status is unofficial and unrecognized. Ultimately, feelings of reduced responsibility translate into reductions in effort toward the accomplishment of team outcomes (Shepperd, 1993). Rotated leadership among team members over the life of a team may ameliorate this tendency because members share (albeit not all at once) in the responsibilities involved in the leadership role. Rotated leadership should promote the perception that each member plays an important part in determining team effectiveness. This perception, in turn, should translate into a greater likelihood that members do their fair share of the team's work (Kerr, 1983; Shepperd, 1993).

Hypothesis 2a: The use of rotated leadership will be positively associated with levels of workload sharing.

It is also likely that rotated leadership will increase the overall level of voice on a team relative to a team that relies on emergent leadership. First, members' experience in the leadership role should increase

their overall knowledge of the team and its task, and this should translate into a greater ability to generate suggestions for change. Second, members' experience in the leadership role should increase their self-efficacy for expressing themselves (Eden, 1992), and this should translate into greater motivation to express suggestions and ideas (Van Dyne & LePine, 1998).

Hypothesis 2b: The use of rotated leadership will be positively associated with the level of voice on a team.

Finally, rotated leadership should clarify who is responsible for performing behaviors associated with specific roles (i.e., leader and member). By clarifying who is to perform specific role behaviors, members should have a better sense of which types of behavior to enact, and this should translate into fewer misunderstandings and, overall, a smoother system of interpersonal interaction (Tuckman, 1965). It is possible that assigned leadership in self-managed teams could create tension among members if some nonleaders want to play a leadership role. However, this tension should be reduced in teams where leadership is rotated and each member gets an opportunity to be the leader. Finally, members in teams with rotated leadership should be empathetic with other members because they have the shared experience of the difficulties associated with the leadership role and empirical evidence shows that shared experiences breed empathy (Baston, Sympson, Hindman, & Decruz, 1996).

Hypothesis 2c: The use of rotated leadership will be positively associated with cooperation.

Here again, as discussed in the section Dimensions of Team Performance and Effectiveness, we also expect that the positive effects of rotated leadership on workload sharing, voice, and cooperation should translate to greater team effectiveness. Thus:

Hypothesis 2d: The use of rotated leadership in self-managed teams will be positively associated with team performance.

Hypothesis 2e: The use of rotated leadership in self-managed teams will be positively associated with the member satisfaction.

Process mediation. As implied by the input–process–outcome perspective, we also expect that the team processes will mediate the relationship between the team designs and team effectiveness. However, given that we are only assessing a subset of team processes (Marks, Mathieu, & Zaccaro, 2001), we expect only partial mediation. Thus:

Hypothesis 3: The relationships between the team designs and the indices of team effectiveness will be partially mediated by the team processes.

Method

Participants and Procedure

One hundred sixteen undergraduates enrolled in three sections of a required human resources management (HRM) course were asked to participate in a study on team effectiveness. Participation was voluntary. One hundred and fourteen students elected to participate, and these students composed our sample of 38 teams. The average age of participants was 21.5 years, and 38% were female.

A single instructor (the first author) taught the three sections of the course. This instructor encountered complaints about social loafing and low cooperation by former students who were involved in team projects. The instructor decided to compare different approaches that might reduce these problems. He wanted to compare what he had done in the past (no formal structure) with teams that had a specific leader for each project and teams that used peer evaluations linked to rewards. His intention was to find out which structure would be best in terms of structuring student teams more effectively. During the first week of classes, the second and third authors suggested that it would be beneficial to collect data to augment the data naturally generated by the course, in order to assess the efficacy of the structures with respect to team processes and outcomes. Therefore, this study was not originally designed as an experiment, but as a *quasi-experiment*.

The course content, group projects, and individual exams were identical across the three sections. Students were assigned randomly to 3-person teams that were responsible for three team projects accounting for 60% of students' grade. The first two projects were cases that involved extensive analysis of HRM issues and they were due in the 5th and 8th week of the semester. The third project required students to design and teach a 45-minute training session and this was due the 11th, 12th, or 13th week of the semester (depending upon when teams were scheduled to make their presentations). There were also two exams accounting for 40% of students' grades.

Quasi-Experimental Conditions

Teams in the three sections were the units of analysis, and each section served as an experimental condition. All students were given the same general guidelines concerning how to work effectively in teams. For example, students were instructed that team members should respect, support, and cooperate with one another.

In addition to these general guidelines, students in the section with peer evaluations were told that they would evaluate one another at the end of the semester, and that these evaluations influenced 10% of their final course grade. The 10% team-based proportion of the grade is consistent with what organizations use (e.g., Caudron, 1994, Lawler, 1999). The formula for the group portion of the final course grade was:

> (Nonpeer evaluated portion of group grade, i.e., 50% of total grade × average group grade across projects, i.e., 0–100%) + (Peer evaluated portion of group grade, i.e., 10% of total grade × average group grade across projects, i.e., 0–100% × average peer evaluation score, i.e., 0–100%)

For teams with *rotated leadership*, members were instructed that they had to choose a different member to lead the team for each project. The project leader was given a higher percentage of the grade for the project for which he or she was responsible. The project team leader also needed to write a short report about his or her function as team leader. Here again, the three projects represented 60% of the course grade. However, of this 60%, 16.7% was given for each of the two projects where an individual was a nonleader, and 26.6% was given for the project where the individual was the leader.

Measures

Team performance. Team performance measures were taken from the two case studies and the mini lessons. Three graduate students who were blind to the experimental conditions graded the two case studies. Because the cases were complex, the graduate students received extensive training on how to grade the cases. They also practiced on cases submitted in previous semesters. Interrater reliability for the cases, as indexed by a 2-way random effects intraclass correlation (ICC), was .66. Accordingly, case grades were averaged across raters. Because the two cases required similar skills, they were averaged to create a single case study performance score. The internal consistency reliability for this score (based on the two cases) was .60. Two additional graduate students, who were also blind to the study conditions, rated videotapes of the teams' performance on the mini lessons. We created a scale to describe the low end and the high end of team performance. Examples of the anchors for low performance were: "Did not describe at all training objectives and assignments" and "Did not create a very positive learning environment (low end)." Examples of the anchors for high performance were "communicated ideas and information well" and "tried to help trainees with their tasks whenever possible." Internal consistency reliability for this measure was .89. Interrater agreement for mini-lesson

performance as indexed by the 2-way random effects ICC was .76, and this supported the aggregation of the scores across the two raters. Because the two measures of performance (e.g., case study, mini lesson) were highly correlated ($r = .70, p < .01$) we standardized the two measures and averaged them. Internal consistency reliability for this measure was .82.

Member satisfaction. The member satisfaction measure was based on LePine and Van Dyne's (1998) 5-item, 7-point faces scale. The items asked participants to select the face that best expressed how they felt about their work team, the members of the team, the quality of interaction among team members, the information they got from team members, and the influence they had with their team. Coefficient alpha for the individual-level data was .89. A significant one-way random effects ICC (.20) and an acceptable mean $r_{wg(j)}$ (.81) provided justification for aggregating the scores to the team level. The team level coefficient alpha (using item means) was .92.

Workload sharing. Participants rated one another on workload sharing using five items: "Adequately completed his or her responsibilities here in this team," "Did his or her fair share of the work on the team's task," "Took responsibilities for team work even though he or she could have avoided it," "Did not purposely avoid working hard on the team's task," and "Was fair in doing his or her fair share of the less pleasant team tasks," and a 7-point scale (1 = *strongly disagree* to 7 = *strongly agree*). Coefficient alpha for the individual-level data was .89. A statistically significant one-way random effects ICC (.16) and an acceptable average level of $r_{wg(j)}$ (.72) indicated support for aggregation (using an average). The team level coefficient alpha (using item means) was .91.

Voice. Participants rated one another on a 6-item voice scale adopted from Van Dyne and LePine (1998). Examples of items included "Made suggestions to others in this group about changes that might have improved the group," and "Expressed opinions about issues even when others in the group thought differently." Respondents used a 7-point scale (1 = *strongly disagree* to 7 = *strongly agree*). The coefficient alpha for the individual-level data was .86. A statistically significant one-way random effects ICC (.05) and an acceptable average level of $r_{wg(j)}$ (.76) indicated support for aggregating the scores (using an average). The team level coefficient alpha (using item means) was .80.

Cooperation. Cooperation was measured using six items and a 7-point scale (1 = *strongly disagree* to 7 = *strongly agree*). Example items included "The members of my group were cooperative with each other," and "Everyone on the team seemed to work together well." Coefficient alpha reliability estimate for the individual-level data was .83. A statistically significant one-way random effects ICC (.25) and an acceptable average

level of r_{wg} (.82) provided justification for aggregating the individual level data. Team level coefficient alpha (using items means) was .81.

Results

Participant Perceptions of Conditions

To determine whether the quasi-experimental conditions influenced perceptions in ways that were consistent with our expectations, we conducted two one-way ANOVA's with planned contrasts. The dependent variable for the first ANOVA was the item "The members of my team had significant power over my class grade." This item was rated on a 7-point scale ranging from 1 = *strongly disagree* to 7 = *strongly agree*. Results indicated significant differences between classes ($F[2, 106] = 106$, $p < .05$). Consistent with expectations, there was a significant difference between individuals in the peer evaluation class and individuals in the no formal design class (mean difference .89, $p < .05$). Likewise, and as expected, the difference between individuals in the peer evaluation class and individuals in the rotated leadership class was also statistically significant (mean difference .74, $p < .05$). There was no difference on this item for teams in the rotated leadership and the no formal design class (mean difference = .15, *ns*).

The dependent variable for the second one-way ANOVA was an item stating, "Each person in the team took a leadership role in managing the team for one specific class assignment." This question was rated on a 7-point scale ranging from 1 = *strongly disagree* to 7 = *strongly agree*. Results indicated that there were significant differences between the classes ($F[2, 106] = 4.90$, $p < .01$). As expected, there was a significant difference between the individuals in the class with rotated leadership and the individuals in the class with peer evaluations (mean differences 1.10, $p < .05$) and individuals in the class with no formal team design (mean differences 1.22, $p < .01$). There was no difference between individuals in the peer evaluation class and individuals in the no formal design class (mean difference = .12, *ns*).

Overall, participant responses to these two items indicated that individuals in the different classes perceived differences in the extent to which (a) peers had power over their grades and (b) team leadership for the projects was rotated.

Relationships Among Study Variables

As illustrated in Table 1, the relationships among the team process variables were positive and statistically significant. These correlations

TABLE 1

Descriptive Statistics and Correlations

	M	*SD*	α^a	α^b	ICC	$R_{wg(j)}$	1	2	3	4	5
1. Workload sharing	4.62	0.47	.89	.90	.16[c]	.72	–				
2. Voice	5.34	0.48	.86	.80	.05[c]	.76	.41*	–			
3. Cooperation	5.54	0.72	.83	.81	.25[c]	.82	.48*	.34	–		
4. Team performance	0.00	1.00	.82	–	–	–	.62*	.73*	.55*	–	
5. Member satisfaction	5.43	0.77	.89	.92	.20[c]	.81	.75*	.19	.51*	.38*	–

Note:ICC's and R_{wg} for team performance are reported separately for the two measures of performance (case study and mini lesson) in the measures section.

[a] Cronbach's alpha based on individual level responses.

[b] Cronbach's alpha based on item means across members.

[c] One-way random effects ICC.

$p < .05$

are either similar to, or smaller than those found in previous research. For example, the relationship between cooperation and workload sharing in this study ($r = .48$) is exactly the same as that reported in the first Campion et al. (1993) study, but much smaller than that reported in the subsequent Campion et al. (1996) study ($r = .65$). The correlation between the effectiveness criteria (team performance and member satisfaction) was also positive ($r = .38$). Finally, with the exception of the relationship between voice and satisfaction, and consistent with past research (e.g., Barrick et al., 1998; Campion et al., 1993, 1996), correlations between the process and team performance were also positive and statistically significant.

Assessing the Hypotheses

We created two dummy variables to capture relationships between the conditions (Cohen & Cohen, 1983) and the processes and outcomes. One dummy variable captured effects of peer evaluations and the other dummy variable captured effects of rotated leadership. The condition with no formal structure served as the comparison group. We then regressed each process and outcome variable on the two dummy variables. Because we used dummy variables, the *B* weight for the intercept is the mean of the relevant dependent variable for teams in the comparison group. The sum of the intercept *B* weight and the *B* weight for the peer evaluation variable is the mean of the relevant dependent variable for teams with peer evaluations. Finally, the sum of the intercept *B* weight and the *B* weight for the rotated leadership variable is the mean of the relevant dependent variable for teams with rotated leadership.

As Table 2 illustrates, the study conditions explained a statistically significant amount of variance in all the process and outcome measures (range of $R^2 = .20–.61$). In support of Hypotheses 1a–1e, the peer eval-

TABLE 2

Effects of Study Conditions on Team Processes and Outcomes

	(1) Workload sharing		(2) Voice		(3) Cooperation		(4) Team performance		(5) Member satisfaction	
	B	R^2	*B*	R^2	*B*	R^2	*B*	R^2	*B*	R^2
Study conditions		.20*		.38*		.39*		.61*		.18*
Constant	4.44*		4.97*		4.95		−.88*		5.19*	
Peer evaluation	.47*		.73*		1.10*		.83*		.72*	
Rotated leadership	.07		.32*		.67*		1.73*		.04	

Note: Constant represents performance level of the control group. *B* = unstandardized regression coefficient. *B*s for treatment groups (peer evaluation and rotated leadership) represent mean difference between treatment group and control group. Team performance is standardized (mean across all groups = 0).

*$p < .05$

TABLE 3

Regression Results

	(1) Team performance		(2) Member satisfaction	
	B	ΔR^2	*B*	ΔR^2
Step 1—Processes		.69*		.57*
Workload sharing	.57*		1.15*	
Voice	.98*		−.22	
Cooperation	.34*		.20	
Step 2—Manipulations		.09*		.04
Peer evaluation	.97*		.40	
Rotated leadership	.40		.00	

*$p < .05$

uation *B* weights when predicting the processes and outcomes were all statistically significant and positive. Workload sharing, voice, cooperation, team performance, and member satisfaction were all higher in teams using peer evaluations. In support of Hypotheses 2b–2d, the *B* weights for rotated leadership when predicting voice, cooperation, and team performance were positive and statistically significant. The rotated leadership *B* weights for the workload sharing and member satisfaction variables were not statistically significant, however. Thus, Hypotheses 2a and 2e were not supported.

Next, we examined the potential mediating role of the team processes on the relationships between the team designs and the effectiveness variables (James & Brett, 1984). We first regressed the effectiveness criteria on the team process variables. As shown in Table 3, the set of

team process variables accounted for 69% of the variance in team performance, and 57% of the variance in member satisfaction. Workload sharing ($B = .57, p < .05$), voice ($B = .98, p < .05$), and cooperation ($B = .34, p < .05$), were all significant predictors of team performance. Finally, only workload sharing predicted member satisfaction ($B = 1.15$, $p < .05$).

We then added a step in the regressions that included the two dummy variables. Given that the team designs had statistically significant effects on the three outcomes, full mediation would exist if the incremental R^2 for the step including the team designs was not statistically significant in the regressions for the outcomes when controlling for the team processes. Partial mediation would exist if the incremental R^2 for the step including the team designs was less than in the relevant regressions reported in Table 2 (i.e., the regressions where only the two dummy variables served as predictors of the outcomes).

Overall, the results for team performance supported partial mediation. Although the incremental R^2 for the second step of the regression of team performance remained significant (.09), it was substantially reduced (87%) when controlling for the processes. The results for member satisfaction, however, supported full mediation. Specifically, there was no incremental variance explained in member satisfaction when the three team processes were included in the regression ($\Delta R^2 = .04$, *ns*). Overall, the analysis supported Hypothesis 3. The amount of variance explained in the team outcomes by the team designs was either reduced substantially or eliminated when the team processes were controlled.

Discussion

The use of self-managed teams in organizations is increasing. Although decisions regarding the design of teams can be made by managers or the teams themselves, there is a lack of research on the implications of these choices. The primary purpose of this study was to investigate design decisions related to forms of member evaluation and leadership. In a quasi-experiment of 38 self-managed student teams over 14 weeks, we found that peer evaluations or rotated leadership promoted team effectiveness as indexed by team performance and member satisfaction. Moreover, these effects appeared to be at least partially mediated by three team processes: workload sharing, voice, and cooperation.

Contributions

Differences in team designs explained more than half of the variance in team performance and about one fifth of the variance in member sat-

isfaction. The sizes of these effects were fairly strong relative to other manageable team inputs such as team composition (e.g., Barrick et al., 1998; LePine, Hollenbeck, Ilgen, & Hedlund, 1997). From a broad perspective, these findings suggest that relatively simple differences in team designs may have important effects on team processes, and ultimately, team effectiveness. Thus, our results may be relevant to observations that teams are often launched in a vacuum, with little or no consideration of those structural factors that might promote or inhibit their functioning and effectiveness (e.g., Devine, Clayton, Philips, Dunford, & Melner, 1999; Dumaine, 1994). Our study also makes a number of more specific contributions, however.

First, our study examines the effects of peer ratings for evaluation and reward purposes. Although many believe that such ratings may be valuable with respect to the quality of information they provide, they are often not used in organizations because they are believed to be deleterious with respect to team functioning and effectiveness. We found evidence to the contrary. Specifically, we found that peer evaluations promoted workload sharing, voice, and cooperation, and that these effects translated into higher levels of performance and member satisfaction.

Second, although some have mentioned rotated leadership as a means of promoting a team's ability to function effectively (Mohrman et al., 1995), this idea has not, to our knowledge, been assessed. Although teams with rotated leadership did not appear to have higher levels of workload sharing or member satisfaction, these teams did have higher levels of voice and cooperation, and these relationships did appear to translate into higher levels of team performance. We are not suggesting that rotated leadership is universally preferable to emergent leadership or to a single designated leader, but we are suggesting that rotated leadership may be a viable team design option.

Limitations

As in all quasi-experiments, this study may have been exposed to threats to internal validity. Especially relevant to this study may be expectancy and selection effects. The instructor of the course was not blind to the experimental conditions and therefore may have inadvertently treated students in ways that increased probabilities of getting the desired results (Schwab, 1999). For example, although the class topics and materials were identical in all three classes, the instructor may have invested extra effort in teaching the peer evaluation structure class. If this were true, participants in the peer evaluation structure condition would have outperformed participants in the other two conditions on other types of performance in addition to the team-performance mea-

sures previously described. However, we assessed whether there were differences between the classes in individual exam grades and results of a one-way ANOVA where the combined grade of these two exams served as the dependent variable (at the team level), revealed no significant differences ($F[2, 32]= 1.00, ns$).

Another alternative explanation for our results might be that preexisting differences between the groups in ability and/or motivation caused participants in the peer appraisal condition to outperform those in the other two conditions. Accordingly, we measured participants' ability (using the Wonderlic personnel test, [Wonderlic, 1983] form 2) and conscientiousness (using the NEO-PIR; Costa & McCrea, 1995). Here again, however, there were no significant differences between the classes (cognitive ability: $F[2,31]=1.27, ns$; conscientiousness: $F[2, 32]=1.24, ns$).

Although the interrater agreement for all group-level variables as indicated by the r_{wg}'s were all above .70, the ICC values for some of the variables were low. This discrepancy could be due to relatively low levels of between groups variance on these variables because, whereas low between groups variance reduces the ICC (the ICC is essentially a ratio of between groups variance to total variance), it does not affect the r_{wg}. We do note, however, that the ICC1s were generally consistent with those found in the literature. James (1982), for example, noted that the median ICC1 in the climate literature was approximately .12. Bliese (2000) suggested that this median ICC1 value may be an overestimate and that values in applied research typically range between .05–.20. We acknowledge, however, that although the ICC1s are reasonable and justify aggregation, the ICC2 values, which indicate interrater reliability of the group mean, are low. The main reason for this is because we used very small (3-person) teams and ICC2s are a function of group size. Concerns about this issue should be reduced somewhat because the team-level variables with the lowest ICCs in our study (e.g., voice and workload sharing) are, in Chan's (1998) terminology, of the additive form. In the context of our study, this means that the items in the measures reflected the behavior of the individuals on the team, and the aggregates of these measures simply accounted for the overall level of these variables on the team. According to this model, therefore, within group heterogeneity of members' scores on workload sharing and voice does not indicate measurement error. This does not mean that interrater reliability is irrelevant, but that perhaps the more appropriate interrater reliability would be that for members' ratings of one another (i.e., where some level of agreement is expected). Indeed, these ICCs were uniformly higher. For example, although the ICC1 value for the aggregate voice measure was only .05, the average ICC1 of members' evaluation of one another was .27.

Ambiguity of causal direction may also be an issue in our study. Although this cannot be a problem associated with the design conditions, we cannot rule out the possibility that member satisfaction or knowledge of team performance drove members' perceptions of workload sharing, voice, and cooperation, and not the other way around. One factor that might reduce this concern is that we measured the processes during Week 10, before teams received feedback on the second case and before teams completed their mini lessons. We also note that because of the possibility of nonrecursive relationships between the processes and outcomes, we were careful about describing our tests of mediation in noncausal language, as recommended by James and Brett (1984).

Another limitation with our study is that we considered a limited number of team processes. Although assessing other team processes (e.g., conflict) could have led to increased insight, measures of team processes tend to be highly correlated (Barrick et al., 1998). Given our modest sample size, we needed to balance comprehensiveness with parsimony, and therefore, we assessed what we believed to be the most relevant process from each of the three general types outlined by Marks et al. (2001). Nevertheless, future research should investigate other team processes in studies of team designs.

Our study may also be limited in terms of the generalizability of our findings. Although we agree that student teams have characteristics that differ from teams in most organizational settings (e.g., members are not compensated with money), we feel there are reasons why this should not be a major concern. First, the members of the teams in our sample did identify with one another as being part of a team, and there were meaningful consequences for their collective performance on tasks for which the members were collectively held responsible. Thus, our quasi-experiment on student teams did seem to posses a large degree of both mundane and psychological realism (Berkowitz & Donnerstein, 1982; Carlsmith, Ellsworth, & Aronson, 1976). Second, as mentioned earlier, we found relationships among the team processes and outcomes that were very consistent with past research conducted in organizational settings (e.g., Barrick et al., 1998; Campion et al., 1993, 1996). Finally, there is evidence supporting the generalizability of research findings obtained in contrived settings across many psychological domains, including leadership and social loafing (Anderson, Lindsay, & Bushman, 1999, Locke, 1986). Nevertheless, scholars should replicate our findings using different subjects, settings, and time, in order to assess generalizability (Cook & Campbell, 1979; Dipboye & Flanagan, 1979).

In fact, the literature would be best served by future research conducted in settings where our hypotheses would be most likely to be disconfirmed (Platt, 1964). For example, our teams were fairly homo-

geneous in terms of race and ethnic background, and in this setting peer evaluations worked well. However, peer evaluations may not work as well in settings where teams are more diverse with respect to these characteristics. This is because raters tend to give higher ratings to members of their own race (e.g., Mount, Sytsma, Hazucha & Holt, 1997; Prewett-Livingston, Field, Veres, & Lewis, 1996), and assuming that those who receive the lower ratings perceive this bias, dissension and dissatisfaction among team members may result. Future research aimed at this issue may be very valuable in terms of promoting our knowledge of team designs.

Finally, we only considered a few of the team designs that could potentially be used by organizations. Although considering other types of team structures would have been desirable, we were constrained with what we could do with the student teams. For example, we did not consider a team structure where the leader was also responsible for providing specific evaluations of his or her subordinates. This was because the instructor did not think it was reasonable to give a single student (i.e., the leader) control over a significant portion of two other students' grades. As in all quasi-experiments, we had to balance the demands of the situation with our need to be able to draw causal inferences.

Conclusion

Despite the limitations, our study makes a number of contributions to the literature on teams. Primarily, our findings suggest that relatively simple team designs may have important impact on the teams' functioning and effectiveness. Although the findings need to be verified in other settings with other samples, the research is potentially important given the increased use of self-managed teams in organizations.

REFERENCES

Anderson CA, Lindsay JJ, Bushman BJ. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*, 3–9.

Barrick MR, Stewart GL, Neubert MJ, Mount MK. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of Applied Psychology, 83*, 377–391.

Batson CD, Sympson SC, Hindman JL, Decruz P. (1996). "I've been there, too": Effects on empathy of prior experience with need. *Personality and Social Psychology Bulletin, 22*, 474–482.

Berkowitz L, Donnerstein E. (1982). External validity is more than skin deep: Some answers to criticisms of laboratory experiments. *American Psychologist, 37*, 245–257.

Bliese PD. (2000). Within-group agreement, non-independence, and reliability: Implications for data aggregation and analysis. In Klein KJ, Kozlowski SWJ (Eds.), *Multilevel theory, research and methods in organizations* (pp. 349–381). San Francisco: Jossy-Bass.

Borman WC, Motowidlo SJ. (1993). Expanding the criterion domain to include elements of contextual performance. In Schmitt N, Borman WC (Eds.), *Personnel selection in organizations* (pp. 71–98). San Francisco: Jossey Bass.

Borman WW, Motowidlo SJ, Rose SR, Hanser LM. (1985). *Development of a model of soldier effectiveness* (Institute Report No. 95), Minneapolis, MN: Personnel Decisions Research Institutes.

Campion MA, Medsker GJ, Higgs AC. (1993). Relations between work group characteristics and effectiveness: Implications for designing effective work groups. PERSONNEL PSYCHOLOGY, *46*, 823–850.

Campion MA, Papper EM, Medsker GJ. (1996). Relations between work team characteristics and effectiveness: A replication and extension. PERSONNEL PSYCHOLOGY, *49*, 429–452.

Carlsmith JM, Ellsworth PC, Aronson E. (1976). *Methods of research in social psychology*. Reading, MA: Addison-Wesley.

Caudron S. (1994). Team staffing requires new HR role. *Personnel Journal, 73*, 88–94.

Chan D. (1998). Functional relations among constructs in the same content domain at different levels of analysis. A typology of composition models. *Journal of Applied Psychology, 83*, 234–246.

Cohen J, Cohen P. (1983). *Applied multiple regression/correlation analysis for the behavioral sciences*. New York: Erlbaum.

Cook TD, Campbell DT. (1979). *Quasi-experimentation: Design, and analysis issues for field settings*. Boston: Houghton Mifflin.

Costa PT Jr, McCrae RR. (1995). Domains and facets: Hierarchical personality assessment using the revised NEO personality inventory. *Journal of Personality Assessment, 64*, 21–50.

Deutsch M. (1949). An experimental study of the effects of cooperation and competition upon group process. *Human Relations, 2*, 199–231.

Devine DJ, Clayton LD, Philips JL, Dunford BB, Melner SB. (1999). Teams in organizations: Prevalence, characteristics, and effectiveness. *Small Group Research, 30*, 678–711.

Dipboye RL, Flanagan MF. (1979). Research settings in industrial and organizational psychology. *American Psychologist, 34*, 141–150.

Druskat VU, Wolff SB. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology, 83*, 58–74.

Dumaine B. (1994, September 5). The trouble with teams. *Fortune*, 86–92.

Eden D. (1992). Leadership and expectations: Pygmalion effects and other self-fulfilling prophecies in organizations. *Leadership Quarterly, 3*, 271–305.

Folger R. (1977). Distributive and procedural justice: Combined impact of "voice" and improvement on experienced inequity. *Journal of Personality and Social Psychology, 25*, 108–119.

Hackman JR. (1987). The design of work teams. In Lorsch JW (Ed.), *Handbook of organizational behavior* (pp. 315–342). Englewood Cliffs, NJ: Prentice-Hall.

Jackson JM, Williams KD. (1985). Social loafing on difficult tasks: Working collectively can improve performance. *Journal of Personality and Social Psychology, 49*, 937–942.

James LR. (1982). Aggregation bias in estimates of perceptual agreement. *Journal of Applied Psychology, 67*, 219–229.

James LR, Brett JM. (1984). Mediators, moderators, and tests for mediation. *Journal of Applied Psychology, 69*, 307–321.

Kane JS, Lawler EE. (1978). Methods of peer assessment. *Psychological Bulletin, 85*, 555–586.

Kerr NL. (1983). Motivation losses in small groups: A paradigm for social dilemma analysis. *Journal of Personality and Social Psychology, 45*, 819–828.

Kerr NL, Bruun SE. (1981). Ringelmann revisited: Alternative explanations for the social loafing effect. *Personality and Social Psychology Bulletin, 7*, 224–231.

Kiesler CA, DeSalvo J. (1976). The group as an influencing agent in a forced compliance paradigm. *Journal of Experimental Social Psychology, 3*, 160–171.

Latané B, Williams K, Harkins S. (1979). Many hands make light the work: The causes and consequences of social loafing. *Journal of Personality and Social Psychology, 37*, 822–832.

Lawler EE. (1999). Creating effective pay systems for teams. In Sundstrom E (Ed.), *Supporting work team effectiveness: Best management practices for fostering high performance* (pp. 188–212), San Francisco: Jossey-Bass.

Lawler EE III, Mohrman SA, Ledford GE Jr. (1995). *Creating high performance organizations: Practices and results of employee involvement and total quality management in Fortune 1000 companies*. San Francisco: Jossey-Bass.

Leavitt HJ. (1951). Some effects of certain communication patterns on group performance. *Journal of Abnormal and Social Psychology, 46*, 38–50.

LePine JA, Hanson MA, Borman WC, Motowidlo SJ. (2000). Contextual performance and teamwork: Implications for staffing. In Ferris GR (Ed.), *Research in personnel and human resource management* (vol. 19, pp. 53–90). New York: JAI Press.

LePine JA, Hollenbeck JR, Ilgen DR, Hedlund J. (1997). Effects of individual differences on the performance of hierarchical decision-making teams: Much more than *g*. *Journal of Applied Psychology, 82*, 830–811.

LePine JA, Van Dyne L. (1998). Predicting voice behavior in work groups. *Journal of Applied Psychology, 83*, 853–868.

LePine JA, Van Dyne L. (2001). Voice and cooperative behavior as contrasting forms of contextual performance: Evidence of differential effects of Big Five personality characteristics and general cognitive ability. *Journal of Applied Psychology, 86*, 326–336.

Liden RC, Mitchell TR. (1983). The effects of group interdependence on supervisor performance evaluations. PERSONNEL PSYCHOLOGY, *36*, 643–653.

Lind EA, Tyler TR. (1988). *The social psychology of procedural justice*. New York: Plenum Press.

Locke EA. (1986). Generalizing from laboratory to field: Ecological validity or abstraction of essential elements? In Locke E (Ed.), *Generalizing from laboratory to field settings* (pp. 3–9). Lexington, MA: Lexington Books.

Marks MA, Mathieu JE, Zaccaro SJ. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review, 26*, 356–376.

McGrath JE. (1964). *Social psychology: A brief introduction*. New York: Holt, Rinehart, and Winston.

Mohrman SA, Cohen SG, Mohrman AM Jr. (1995). *Designing team-based organizations: New forms for knowledge work*. San Francisco: Jossey-Bass.

Mohrman AM, Lawler EE. (1983). Motivation and performance appraisal behavior. In Landy F, Zedeck S, Cleveland J (Eds.), *Performance measurement and theory* (pp. 173–189). Hillsdale, NJ: Erlbaum.

Mount MK, Sytsma MR, Hazucha JF, Hold KE. (1997). Rater–ratee race effects in developmental performance ratings of managers. PERSONNEL PSYCHOLOGY, *50*, 51–69.

Murphy KR, Cleveland JN. (1991). *Performance appraisal*. Needham Heights, MA: Allyn & Bacon.

Naiper N, Latham G. (1986). Outcome expectancies of people who conduct performance appraisals. PERSONNEL PSYCHOLOGY, *39*, 827–837.

Platt JR. (1964). Strong inference. *Science, 146*, 347–353.

Prewett-Livingston AJ, Field HS, Verses JG III, Lewis PM. (1996). Effects of race on interview ratings in a situational panel interview. *Journal of Applied Psychology, 81*, 178–186.

Ringelmann M. (1913). Recherches sur les moteurs animes: Travail de l'homme [Research on animate sources of power: The work of man]. *Annales de l'Institut National Agronomique, 2e* serie-tome XII. 1–40.

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

Schwab DP. (1999). *Research methods for organizational studies*. Mahwah, NJ: Erlbaum.

Shepperd JA. (1993). Productivity loss in performance groups: A motivation analysis. *Psychological Bulletin, 113*, 67–81.

Szymanski K, Harkins SG. (1987). Social loafing and self-evaluation with a social standard. *Journal of Personality and Social Psychology, 53*, 891–987.

Tuckman B. (1965). Developmental sequence in small groups. *Psychological Bulletin, 63*, 384–399.

Van Dyne L, LePine JA. (1998). Helping and voice extra-role behavior: Evidence of construct and predictive validity. *Academy of Management Journal, 41*, 108–119.

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

Wonderlic & Associates. (1983). *Wonderlic Personnel Test manual*. Northfield IL: ER Wonderlic & Associates.