

Group & Organization Management

<http://gom.sagepub.com>

Peer Assessment, Individual Performance, and Contribution to Group Processes: The Impact of Rater Anonymity

Peter A. Bamberger, Ido Erev, Michal Kimmel and Tali Oref-Chen
Group Organization Management 2005; 30; 344
DOI: 10.1177/1059601104267619

The online version of this article can be found at:
<http://gom.sagepub.com/cgi/content/abstract/30/4/344>

Published by:



<http://www.sagepublications.com>

On behalf of:



[Eastern Academy of Management](http://www.eam.ac)

Additional services and information for *Group & Organization Management* can be found at:

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

Subscriptions: <http://gom.sagepub.com/subscriptions>

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

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

Citations <http://gom.sagepub.com/cgi/content/refs/30/4/344>

Peer Assessment, Individual Performance, and Contribution to Group Processes

THE IMPACT OF RATER ANONYMITY

PETER A. BAMBERGER

IDO EREV

MICHAL KIMMEL

TALI OREF-CHEN

Technion—Israel Institute of Technology

Drawing from game theoretical research, we explore the degree to which rater anonymity affects consequences of peer assessment with respect to individual effectiveness and contribution to group processes. Using longitudinal data collected in a kibbutz-owned manufacturing facility in Israel, we found peer assessment to be associated with a general improvement of supervisory ratings of subordinates over time. More important, this improvement was found to be significantly greater among individuals working in departments assigned to a nonanonymity assessment condition relative to those assigned to an anonymous condition. The theoretical and practical implications of these findings are discussed.

Keywords: peer assessment; multisource feedback; game theory; performance evaluation; rater anonymity

Peer assessment methods have been increasingly adopted as one key component of multisource feedback systems (Hedge, Borman, & Birkeland, 2001) and have become particularly attractive to organizations adopting self-managing work groups and team-based work processes (Bamberger &

The authors wish to thank P. Gluzman, Y. Lieberman, I. Meshulam, T. Bar-Nir, and A. Orenstein for their assistance in the data collection phases of this study, as well as Angelo DeNisi, Dafna Eylon, Edward J. Lawler, Manuel London, and Steven Wolff for comments on an earlier draft of this article. Special thanks to Professor Ayala Cohen and Dr. Etti Doveh for statistical assistance. This study was supported by the Technion Institute of Management and the Goldberg Research Fund.

Group & Organization Management, Vol. 30 No. 4, August 2005 344-377

DOI: 10.1177/1059601104267619

© 2005 Sage Publications

344

Meshoulam, 2000). Moreover, decades of research have consistently shown that most systems of peer assessment offer a high degree of reliability and validity (DeNisi & Stevens, 1981; Reilly & Chao, 1982). Employees also appear to place a higher degree of faith in peer-based assessment data (Maurer & Tarulli, 1996), which may in turn elicit greater behavioral changes on the part of employees than traditional supervisory ratings (Fedor & Bettenhausen, 1989).

However, examination of the social and performance-related consequences of peer assessment appears to lead to mixed results. On one hand, Hazucha, Hezlett, and Schneider (1993) found peer assessment and other multisource feedback components to have a positive, long-term impact on managerial effectiveness. Similarly, Druskat and Wolff (1999, p. 69) concluded that peer appraisals "can have an immediate positive impact on group member perceptions of open communication, group task focus, group viability and member relationship."

On the other hand, many other researchers examining the social consequences of peer assessment have concluded that such appraisals generally harm relationships and impair group functioning. For example, Kane & Lawler (1978) suggest that users may not always cooperate in peer assessments because such systems implicitly request privileged information about areas that may seriously disturb the peer group. To the extent that they do cooperate, their appraisals may be subject to rater biases (Saavedra & Kwun, 1993) that can affect members' perceptions about the group (DeNisi, Randolph, & Blencoe, 1983) as well as their morale (Cederblom & Lounsbury, 1980). Others (Kane & Lawler, 1978) note that because peer appraisals can turn into popularity contests, they can have severe negative implications with respect to team member relationships. Finally, DeNisi et al. (1983) found that negative peer assessment feedback had detrimental effects on group cohesiveness and satisfaction, as well as on subsequent peer ratings on job tasks.

We believe that these mixed results reflect two robust but contradictory behavioral effects of peer assessment: one positive and one negative. The objective of this article is to examine the role of rater anonymity as one factor that may influence the balance between these two contradictory effects. We begin this examination by reviewing the possible explanations for both positive and negative effects of peer assessment on individual job performance and contribution to group processes and then focus on how rater anonymity may serve as a critical, determining factor. Specifically, we suggest that there may be a number of reasons why peer assessment may, in general, have positive consequences with respect to individual job performance and contribution to group processes. However, we also point out that under conditions of

rater anonymity, these positive effects may be diminished and even offset. Our argument is based on the assumption that under such conditions, self-enhancement biases are likely to prime a self-reinforcing tendency to assume that others' ratings of oneself have been downwardly biased. Such an assumption on the part of raters is by no means unreasonable, given Antonioni's (1994) finding that upward appraisal ratings of managers were significantly lower when their subordinate raters were assigned to use an anonymous (as opposed to an accountable) appraisal procedure. Because beliefs about peers' motivations may shape individuals' own rating behaviors, we argue that such perceived peer biasing may result in increasing levels of actual downward biasing, which may in turn have a negative impact on a variety of performance-related work behaviors. Such negative outcomes are less likely under conditions of rater nonanonymity, because the lack of anonymity provides raters with a disincentive to downwardly bias their ratings of others (i.e., raters may be more likely to believe that they will be held individually accountable for any self-serving ratings of others).

In this sense, the current study offers a unique approach to the study of peer assessment. Whereas previous studies, as noted above, have focused primarily on the reliability, validity, and user acceptability of peer assessment, our concern is with its behavioral and performance-related consequences. Furthermore, whereas the limited research on the issue of rater anonymity and accountability has tended to focus on assessment accuracy and user acceptability (Antonioni, 1994; London, Smither, & Adsit, 1997), we examine how rater anonymity may influence the behavioral consequences of peer assessment. Finally, our study is unique in terms of its research methodology in that, unlike much of the previous research on peer assessment, we test our hypotheses on the basis of data collected in the context of a multiyear, longitudinal field experiment.

THE POSITIVE EFFECTS OF PEER ASSESSMENT

The positive consequences of peer assessment may stem from the useful information that peers are able to provide to one another (Fedor & Bettenhausen, 1989), as well as the tendency—at least in some studies—of the assessment data to generally be positive in nature (Druskat & Wolff, 1999). Such positive feedback could potentially boost team cohesion and consequently motivate enhanced individual and unit performance (Druskat & Wolff, 1999).

Furthermore, the literature on social dilemmas (Kerr & Brunn, 1981) suggests that when opportunities for social loafing exist, a positive effect may emerge as a consequence of the enhanced degree of monitoring and

sanctioning inherent in peer assessments. Assuming at least a minimal degree of task and/or outcome interdependence among work unit members (Wageman, 1995), such monitoring and sanctioning may reduce the incentive to "free ride" (Albanese & Van Fleet, 1985). Consequently, under such conditions, peer assessment should be associated with less social loafing and enhanced individual contributions to group processes. Previous studies using student subjects have operationalized these two broad outcomes in terms of enhanced individual initiative and motivation to achieve group outcomes (Druskat & Wolff, 1999) as well as individual task performance and task-related behavior (DeNisi et al., 1983). Field studies using employee samples have operationalized these outcomes in terms of peer helping and coaching (mentoring), and information sharing and teamwork (Wageman, 1995). All of these parameters have been identified as critical to effective work group functioning (McIntyre & Salas, 1995; Wageman, 1995). As such, we would expect to find a generally beneficial effect of peer assessment with respect to the level of individual initiative, and cooperative behaviors such as peer mentoring, teamwork, and motivation to assist in the achievement of group objectives. Furthermore, given that enhanced employee cooperation tends to have a direct, positive effect on effectiveness (Tjosvold, 1984), and consistent with Hazucha et al.'s (1993) findings noted above, we would also expect peer assessment to have a generally beneficial impact on individual effectiveness on the job. Indeed, in a recent quasi-experiment using student subjects, Erez, Lepine, and Elms (2002) found peer assessment to have a positive impact on performance and demonstrated that the effects of peer evaluation on performance are mediated by increased member cooperation and workload sharing.

THE NEGATIVE EFFECTS OF PEER ASSESSMENT

However, such beneficial effects may be greatly diminished when raters are concerned that their peers' ratings may be at least somewhat motivated by competitive self-interest and thus unjustifiably biased downward. Regardless of the peer raters' actual intent or rating behavior, such concerns may initially arise when employees perceive some downward bias in the peer assessments received. Given the tendency of individuals to overestimate their own contributions relative to others (Ross & Sicoly, 1979), such a perception of a downward rater bias may not be so unusual. The lower-than-expected results in early peer assessment rounds may subsequently lead to the suspicion of a competitive orientation on the part of one's peers, thus providing a cognitive-based stimulus for the adoption of a competitive orientation on the part of the recipient. Perceived bias on the part of one's peers may thus generate actual

bias in subsequent ratings provided to others, particularly if peer raters believe that the benefits to be gained by strategically and downwardly biasing their assessments of others outweigh the risks of getting caught and being subject to administrative retribution and/or peer retaliation. These benefits may include an enhanced standing relative to one's peers, not to mention an enhanced ability to defend against becoming a "sucker" (i.e., receiving a lower relative standing as a result of potential downward biasing by others).

Such benefits may become particularly salient when peer assessment data are used for evaluation purposes; that is, as a basis upon which to make decisions regarding employee compensation or job placement (Waldman, Atwater, & Antonioni, 1998), a practice which is increasing in popularity (London & Smither, 1995, p. 807). They may also become salient when, regardless of management's espoused intentions (i.e., to use the assessment data for developmental purposes only), employees perceive that the results of a peer assessment may influence such decisions in the future (Atwater, Waldman, Atwater, & Cartier, 2000). As Tornow and London (1998) suggest, because many organizations fail to adequately communicate the developmental nature of their peer assessment, rater ambiguity about the eventual use of peer assessment data is a relatively common problem. Moreover, as Fedor, Bettenhausen, and Davis (1999, p. 96) note,

Even when an organization clearly states how it intends to use performance appraisal information, the inevitable ambiguities associated with administering the system and employees' efforts to make sense of the entire process can lead employees to perceive differently the extent to which the appraisal system is used for developmental feedback purposes.

Under such conditions, the perceived incentive structure may be such that peers may feel that they have little choice but to proactively bias their ratings of others to provide a minimal degree of protection from becoming the sucker in a ratings game. Moreover, such biasing can result in the creation of a vicious circle of ever-increasing biasing in peer assessment. Specifically, the self-fulfilling prophecy initiated by a desire to avoid becoming a sucker may result in increasing degrees of perceived and actual rater bias as rater suspicions of peer bias are in fact confirmed. As suspicions are confirmed over time, the incentive to downwardly bias one's own assessments of others becomes reinforced, further priming this tendency toward downward biasing.

The perception that assessment results may be downwardly biased may have a negative impact on individual effectiveness and contribution to group processes for a number of reasons. First, if workers perceive assessment

results to be biased, they may be less likely to deem the feedback to be at the level of quality required to justify behavioral change and/or an investment in self-development. Second, if workers believe that their peers are likely to strategically and downwardly bias their assessments (regardless of the individual's true level of performance), they may reduce their own level of effort and cooperation proportionately to avoid placing themselves in a sucker role (Albanese & Van Fleet, 1985). In this sense, the mere suspicion that peers are likely to downwardly bias their assessments of each other may generate a kind of self-fulfilling prophecy of reduced individual effort and unit-wide cooperation. Moreover, interpersonal resentment among peers may develop as individuals begin to suspect that (a) cooperation by others is little more than "window dressing" in an effort to "win points" (Kerr & Brunn, 1981) and that (b) their peers cannot necessarily be trusted to provide bias-free, quality assessments (London et al., 1997, p. 167).

PEER ASSESSMENT AND RATER ANONYMITY

In the majority of workplaces, peer assessment is implemented on an anonymous basis (London & Smither, 1995, pp. 805-807). Indeed, prevailing managerial theory suggests that anonymous assessment reduces the risk of rating inflation (London et al., 1997) and, as Tornow and London (1998, p. 7) indicate, "promotes feedback candor." However, as suggested by London et al. (1997, p. 166), anonymity may also generate less-than-accurate ratings and consequently serve as the Achilles' heel of peer assessment. In this context, we propose that a relatively simple manipulation—the elimination of this convention of rater anonymity—may enhance the long-term contribution of peer assessment to organizations.

This assertion may be supported on two grounds. First, although not examined directly by previous studies, this assertion is consistent with the results of a cross-study comparison. Specifically, the clearest evidence of positive behavioral and attitudinal consequences has been observed in studies (e.g., Erez et al., 2002) in which ratings were provided on a nonanonymous basis. In these studies, rater names were clearly linked to their ratings, allowing the ratee and/or assessment administrator to potentially trace specific ratings back to a rater either at will or when foul play (e.g., purposeful or malicious biasing) was suspected. In contrast, in studies finding peer assessment to have generally adverse behavioral and attitudinal consequences (e.g., Tagger & Brown, 2001), the ratings were generally provided anonymously.

A second reason to expect that anonymity may affect the tradeoff between the positive and negative effects of peer assessment comes from a simple

game theoretic analysis. Game theoretical analysis (see Luce & Raiffa, 1957) implies that from the standpoint of a rational individual, the assessment of a peer is a behavior just like the behavior being assessed, and therefore peer assessor evaluative decisions will be guided by strategic considerations. In particular, game theoretical analysis suggests that it is likely that peers will “defect” or adopt a strategy potentially beneficial to themselves but harmful to their colleagues when there is little probability that such behavior will affect their future interaction with their colleagues, that is, when peers perceive no mechanism by which to retaliate against a defector, or for a defector to retaliate against them (Murnighan & Roth, 1983). Axelrod and Dion (1988) refer to such a situation as one lacking an indefinite “shadow of the future”: the indefinite shadow being what, in effect, “keeps everyone honest.” As has been demonstrated in a number of organizational studies (e.g., Wagner, 1995), such a lack of a shadow of the future may be manifested when individuals are unable to observe or identify potential defectors. The results of these studies indicate that defection behavior (such as social loafing or, as in the case of the current study, the downward biasing of assessments of others in a peer assessment framework) is strongly influenced by identifiability, or the degree to which others can observe and assess an individual’s behaviors. For example, in a study of classroom groups, Wagner (1995) found that the identifiability of individual tasks greatly reduced the level of social loafing. Thus, lack of identifiability may provide a central motivation for certain individuals in a group to defect against, rather than cooperate with, their peers (Albanese & Van Fleet, 1985).

However, beyond this main effect, Luce and Raiffa (1957) suggest that such conditions may motivate others to adopt defection rather than cooperation as a simple matter of self-defense. Luce and Raiffa demonstrated that through a process of “backward induction,” the threat of defection by others may convince individuals that it may be best to take pre-emptive action against one’s peers and defect against them by adopting a similar manner of behavior before their peers do the same. That is, even in the absence of proven defection behavior by one’s peers, individuals may themselves opt to defect simply to minimize the damage likely to be caused to them if their colleagues in fact defect as expected (Axelrod & Dion, 1988).

Assuming that most workers are at least somewhat suspicious that peer ratings may eventually influence managerial administrative decisions, and drawing from the game theory literature cited above, we believe that anonymous peer assessment may create an incentive structure in which peer raters increasingly come to believe that it is in their best interest to downwardly bias their assessments of others. First, anonymity lowers any disincentive for downwardly biasing assessments of others because it is difficult for peers or

appraisal administrators to identify precisely who is giving the biased rating, a key precondition for confrontation or retaliation. At worst, under conditions of true anonymity, a clearly downwardly biased assessment (i.e., one that, when compared with all other assessments of the same individual, is an obvious outlier) may be excluded when estimating an average rating. Consequently, under conditions of anonymity, peers have little to lose by downwardly biasing their ratings of others. In fact, the game theory literature provides substantial support for such a notion using experimental N-person prisoner's dilemma games. For example, Fox and Guyer (1978) found lower rates of defection (higher rates of cooperation) when participants could observe and identify the decisions made by their peers.

Second, anonymity may in fact increase the incentive to provide downwardly biased assessments as a means by which to protect oneself from the eventual downward biasing expected by others. Under conditions of anonymity, raters' fears that their peers may strategically bias their ratings of others downward may be heightened because there is no effective means by which to identify and then confront those providing unjustified ratings. Lacking any reasonable mechanism by which to respond, and consistent with the game theoretical notion of "backwards induction," raters may adopt the defensive strategy of proactively and downwardly biasing their assessments of others. In doing so, they may minimize the risk that the downward biasing of others will have a negative impact on their overall relative standing. Consequently, there are direct, short-term benefits to the anonymous rater for downwardly biasing his or her ratings of others. However, over time, such assessment behaviors may create a self-reinforcing pattern of downward biasing among raters in general, with all of the negative implications noted above.

Requiring that peer raters sign off on their ratings of others (i.e., non-anonymity) may be an effective means by which to avoid such problems. By doing so, it may be possible to alter the perceived incentive structure underlying peer assessment in that these raters must take into account that, regardless of any claim of confidentiality, their ratings of others can always be traced back to them. In this way, the simple practice of nonanonymity increases the risks faced by peer raters when downwardly biasing their assessments of others or otherwise providing poor quality and unjustifiable feedback. At the very least, assessment administrators may question the assessment practices of such individuals and may even enforce sanctions if foul play is suspected. Although little attention has been paid to the issue of rater anonymity in performance assessment, at least one study (Fusilier, 1980) found raters to be less averse to giving low performance ratings when they knew ratees would be unable to identify them as the raters.

Recognition of such sanctions is also likely to reduce the presumption of downward biasing on the part of others and consequently boost the perception of rating accuracy. Antonioni (1994) found that managers who were able to identify their raters viewed the upward feedback process more positively than did managers who received feedback from anonymous raters. Regardless of the degree to which nonanonymity truly does enhance rating accuracy, the simple presumption of enhanced accuracy may be sufficient to prevent any backwards induction of bias by others and the perceived need to proactively bias one's ratings of others in self-defense. Consequently, many of the potential negative effects of peer assessment noted above (particularly the emergence of a self-reinforcing tendency to provide biased ratings of others) may be effectively eliminated.

Moreover, requiring raters to identify themselves may enhance many of the potential positive behavioral consequences of peer assessment. First, by eliminating any tendency to downward bias, peer assessments may, on the whole, become more favorable, thus boosting peers' sense of camaraderie, encouraging ever-greater levels of cooperation, and enhancing members' willingness to contribute to group-level objectives. Antonioni (1994, p. 354) found that subordinates rated their managers more favorably when raters were identified than when they were anonymous. Second, by providing an incentive to invest in assessment decisions and potentially boosting rater conscientiousness, there may be less of a risk that raters will allow their affect to influence their ratings (Antonioni & Park, 2001). A number of studies (e.g., Mero & Motowidlo, 1995) suggest that the potential need to justify a decision can encourage individuals to think longer and more carefully about their rating decisions. Consequently, actual quality and accuracy of assessment data may improve (London et al., 1997, p. 166), thus having direct performance consequences. Based on the discussion above, and assuming that employees are at least somewhat suspicious that peer ratings may eventually influence certain managerial administrative decisions, we hypothesize the following:

Hypothesis 1: On average, peer ratings will be higher among those employed in work units in which peer assessment is conducted on the basis of rater nonanonymity relative to those employed in units in which peer assessment is conducted on the basis of rater anonymity.

Hypothesis 2: Supervisory ratings of contribution to group processes (i.e., individuals' group task motivation, individual initiative, teamwork, and mentoring) and effectiveness (i.e., productivity-related work behaviors, overall work performance) will be higher among those in units in which peer assessment is

conducted on the basis of rater nonanonymity relative to those employed in units in which peer assessment is conducted on the basis of rater anonymity.

EXPERIENCE, LEARNING, AND THE IMPACT OF PEER ASSESSMENT

Although increased attention has been paid to the variability of employee performance over time (Deadrick & Madigan, 1990), and the learning-related factors potentially accounting for this variance (Murphy, 1989), the dynamic performance perspective has largely been neglected by researchers concerned with multirater assessment. For multirater assessment research, this perspective is important in that it suggests that when new assessment practices are adopted, these practices—as elements of the job environment—may themselves trigger shifts in the distribution of employee performance (Deadrick, Bennett, & Russell, 1997). According to Murphy's (1989) model, it is likely to take employees time to comprehend and make sense of any such change in the job environment, and as such, the effects of such a change in assessment practices are not likely to be immediately observed. That is, both the positive and negative effects of the peer assessment-based incentive structures may only emerge with experience (Erev & Roth, 1998). Specifically, the generally positive impact of peer assessment on both individual and overall group outcomes may only emerge with experience, as group members learn that increased effort, information exchange, mentoring, and overall cooperation are indeed associated with more positive peer assessments (Leslie, Gurskiewicz, & Dalton, 1998). Similarly, the negative implications of peer assessment on such outcomes may not be immediately manifested but rather may only emerge and intensify with experience as individuals' fears of others' self-enhancing biasing are confirmed and perceptions of peers' past motivation are crystallized. Indeed, consistent with the findings of DeNisi et al. (1983), Tagger and Brown (2001) found that severity biases in peer ratings were more common when individuals received negative ratings from their peers in earlier rounds of assessment. Given that the incentive structure underlying peer assessment may not be immediately recognized and may only emerge with experience over time, we also hypothesize the following:

Hypothesis 3: The difference in supervisory-assessed outcomes under conditions of rater anonymity and nonanonymity will increase in magnitude over time and will be manifested in terms of a significant Time \times Condition interaction.

METHOD

PARTICIPANTS AND STUDY CONTEXT

Participants were 141 workers employed by a kibbutz-owned cable manufacturing firm in Israel. Sixty-three percent of the workers were members of the kibbutz (mean company tenure = 9.8 years), whereas the remaining 37% were employees hired from the outside (mean company tenure = 4.5 years). Seventy-four percent of the workers were male. Workers were employed in 16 of the company's 17 departments, each with its own supervisor. Ten departments were defined as service-oriented (average of 5 workers per work unit, including such departments as maintenance and production control). Six departments were engaged in production (average of 15 workers per work unit, departments included extrusion and cable wrapping). Hired workers were proportionately divided between each of the 16 departments. In all 16 departments, employees worked as members of a team, deciding jointly on work flow and task assignments and engaging in joint problem solving.

Prior to the current study, no formal system of performance assessment had ever been implemented in any of the kibbutz-owned companies, including the cable company. Recognizing the potential benefits of adopting some system of performance appraisal, the management and employees of the company agreed to participate in a longitudinal study of peer assessment.

As noted above, a core assumption underlying our hypotheses is that employees tend to have some doubt as to how management will in fact use the peer assessment data collected. Consequently, we adopted a study design in which employees received only a very general explanation as to the purpose of peer assessment and the way the data would be used. Specifically, at a meeting with employees held prior to the implementation of the assessment intervention, we first had the plant's personnel manager explain that the system was being adopted to "help develop plant staff and thereby enhance productivity." Next, we explained that each employee would be presented with an aggregated assessment report and would be given the opportunity to discuss these results and strategies for performance improvement with both the investigators and "a member of the management team, such as the personnel manager." The latter point was emphasized to highlight the potential developmental use of the results. Finally, we asked the plant's personnel manager to explain that although the data would be deemed confidential, aggregated assessment reports would be included in each employee's personnel file to allow for further development-oriented follow-up. At no other time during

the study did we communicate any further information regarding the purpose of peer assessment or how it would be conducted.

As a consequence, whereas employees understood that the data would be *formally* used for developmental purposes, our communication was made purposefully ambiguous such that employees could potentially infer that the data could *informally* influence certain administrative decisions regarding, for example, job placement. Given that the kibbutz never implemented any form of contingent pay, it is unlikely that employees inferred a possible impact on compensation. Although—per management's request—no formal manipulation check (i.e., of employee beliefs about the intervention's purpose) was conducted, post-hoc interviews conducted approximately 2 weeks after each round of assessment suggest that our manipulation was successful. These semistructured interviews were conducted with 6 workers in each condition (for a list of the questions posed, see Appendix 1). Although most of those interviewed accepted that the main purpose of the system was to help them improve their work habits and team relations, at least half expressed some degree of skepticism regarding management's true, longer-term intentions. These employees noted that they had little doubt that these assessments would, at the very least, influence staffing decisions in the future. This is not surprising given that, in the absence of any other performance assessment system, the peer assessments became the only ratings of record.

THE PEER ASSESSMENT AND FEEDBACK MANIPULATION

Working with the company's personnel manager, we divided the 16 departments into 8 matched pairs on the basis of production or work processes and group size. We paid particular attention to the latter, because it is conceivable that the incentive to distort ratings may be stronger in smaller groups. Eight departments (5 service and 3 production; $n = 64$) were randomly assigned to the nonanonymous condition, and their 8 matched pairs of roughly equivalent size ($n = 77$) were assigned to the anonymous condition. To ensure that no bias was introduced as a function of our condition assignments, preassessment criterion measures (at both the individual and group levels) were taken for subsequent testing. These criterion measures are described below.

Peer appraisals were conducted on a departmental basis three times during the course of 18 months. Peer assessment forms distributed to workers in departments assigned to the nonanonymous condition differed from those distributed to workers in departments assigned to the anonymous condition. Whereas the forms for those in the nonanonymous condition instructed participants to write their name at the top of the first page, the forms used for

those in the anonymous condition instructed participants to complete the form anonymously. Furthermore, the forms used for those in the nonanonymous condition included only the names of the participant's coworkers and not the name of the participant him- or herself. In contrast, the forms used for those in the anonymous condition included the names—in seemingly random order—of all departmental employees, including that of the rater. In this way, despite the fact that forms were coded in such a way as to allow the investigators (and no one else) to identify the rater, it appeared to the participants in the anonymous condition that it would be impossible to identify the individual completing the anonymous assessment form. Participants in both conditions were told that all ratings would be kept strictly confidential, and in fact, in neither condition did ratees ever learn the identity of their raters. Nevertheless, the 12 post-hoc interviews noted above indicated that the manipulation was successful in that interviewees in the nonanonymous condition were far more likely than those in the anonymous condition to express concerns about the confidentiality of their ratings.

All peer assessment forms included items relating to four main performance dimensions (teamwork, professionalism, overall work attitudes, and efficiency in self-management). These four dimensions were identified by top management as having the most critical impact on departmental performance and are also among those factors most reliably assessed by peer raters (Brutus, Fleenor, & London, 1998). Using a behavioral observation scale framework (Latham & Wexley, 1977), during a period of 3 months and on the basis of both on-site interviews and observations, we collected critical incident data having to do with these dimensions from both kibbutz members and hired employees. Following the recommendations of Van Velsor (1998), these incidents were then used to generate statements relating to each of the four dimensions noted above. This approach is also consistent with Hinkin's (1995) recommended approach to enhancing content validity in that our scales were initially based on a classification scheme recommended in the literature and then further developed on the basis of inductive grouping of the critical incident-based items. As recommended by Hinkin (p. 971), this grouping was done by independent raters (in this case, students), with interrater consistency exceeding .80. Inductively generated items were, in all cases, consistent with one of the four underlying constructs. Using a 6-point Likert-type scale format with 6 = *very accurate* and 1 = *not at all accurate* serving as the anchors, employees were asked how accurately these statements captured the behavior of each of their peers. Sample statements are "shares information with teammates" (teamwork dimension) and "operates equipment in the proper manner" (professionalism). Eigenvalue and scree plot analyses of the Time 1 peer assessment data indicated a four-factor

TABLE 1
Factor Analysis of Peer Assessment Items at Times 1, 2, and 3

Item	Time 1				Time 2				Time 3			
	1	2	3	4	1	2	3	4	1	2	3	4
Familiar with processes	.85	.22	.10	.18	.84	.13	.12	.35	.64	.24	.06	.56
Equipment knowledge	.84	.16	.19	.17	.84	.24	.25	.06	.84	.22	.26	.28
Fund of knowledge	.43	.37	.18	.61	.81	.03	.40	.28	.65	.35	.06	.52
Customer orientedness	.52	.32	.38	.38	.65	.05	.53	.12	.61	.22	.34	.49
Overall professionalism	.81	.26	.21	.23	.90	.13	.26	.18	.81	.24	.46	.42
Organized	.11	.81	.11	.36	.09	.95	0	.02	.15	.86	.24	.26
Transfers knowledge to others	.23	.73	.36	.16	.06	.85	.09	.05	.19	.83	.26	.17
Empowerment of others	.15	.74	.40	.03	.09	.92	.17	.03	.05	.72	.51	.19
Decision making	.25	.73	.23	.33	.23	.86	.09	.13	.26	.82	.37	.11
Initiative	.29	.72	.16	.26	.21	.85	.13	.04	.45	.65	.43	.06
Overall self-management	.30	.80	.27	.15	.07	.96	.05	.10	.30	.84	.33	.21
Positive atmosphere	.24	.17	.78	.24	.21	.10	.87	.17	.07	.48	.77	.20
Accepts criticism	.06	.23	.83	.13	.27	.07	.82	.18	.21	.23	.84	.04
Takes responsibility	.22	.20	.71	.35	.33	.14	.83	.09	.19	.56	.68	.24
Involves others	.13	.48	.58	.26	.30	.11	.16	.78	.18	.40	.71	.15
Overall teamwork	.26	.38	.76	.17	.44	.03	.78	.30	.12	.49	.78	.14
Follows rules	.13	.18	.42	.71	.03	.11	.57	.67	.18	.33	.11	.84
Resource conservation	.10	.44	.21	.62	.45	.16	.14	.78	.21	.20	.02	.89
Work values/focus	.20	.19	.12	.85	.53	.14	.38	.59	.35	.09	.22	.78
Conscientiousness	.32	.22	.27	.75	.50	.01	.47	.61	.38	.04	.25	.83
Eigenvalue	1.64	10.67	1.44	1.28	4.91	8.92	1.64	1.25	0.95	11.73	1.06	3.22

NOTE: Bolded items in each column relate to those items included in the scales associated with each of the following identified factors: Factor 1 = professionalism; Factor 2 = efficiency in self-management; Factor 3 = teamwork; Factor 4 = work attitudes.

solution consistent with the four dimensions noted above. As can be seen in Table 1, this four-factor structure remained quite stable during the remaining two time periods, another indicator of construct validity, according to Schwab (1980). Further evidence of construct validity is indicated by the relatively high reliability coefficients (all above .80) generated for each of the four indices constructed on the basis of the results of the factor analysis.

Prior to each assessment round, participants in the nonanonymous condition were reminded that although their assessments would be kept confidential, they had to write their name at the top of the assessment form "in case any problems were identified or further information was needed from them." No further remarks were made regarding how or when the capability to link raters to their assessment data might be used or who might make use of such a

capability, because, given the high degree of contact between anonymous and nonanonymous peer raters, any such discussion might have highlighted the issue of anonymity, thus exposing the manipulation and placing the entire study at risk.

In contrast, participants assigned to the anonymous condition were reminded not to record their name anywhere on any of the assessment forms. At no time, however, did nonanonymous and anonymous participants ask why (respectively) they were required to write their name or to refrain from doing so. Post-hoc interviews suggested that those in the nonanonymous condition had no idea that peers in the anonymous groups completed the forms on an anonymous basis and vice versa. Nevertheless, these interview data also suggested that, in contrast to those in the anonymous condition, already after the first round, those in the nonanonymous condition had a heightened concern that someone, such as the plant personnel manager, might still hold them accountable for their ratings. This is notable, given the lack of specification as to how or when nonanonymous raters might be held accountable for their ratings.

Within 1 month of each peer assessment, we aggregated the individual assessment data and prepared an individualized peer assessment report for each participant. In the case of those assessed anonymously, aggregate scores incorporated both peer *and* self-ratings because, as far as those in this condition were concerned, there was no way that we could conceivably parcel out the self-ratings from the ratings of others. The inclusion of the self-assessment ratings (which were generally higher than the mean *peer* ratings in a group) upwardly biased the mean scores provided back to each of the participants in the anonymous condition. Because inflated ratings could potentially reduce participants' perceptions of rater bias or a competitive orientation on the part of one's peers, the inclusion of the self-assessments had the potential to weaken the magnitude of the manipulation effect. However, we assumed that were this to occur, it would only create a more conservative context for the testing of our hypotheses.

Within 1 month of each assessment, and in the presence of a member of the management team (usually, the HR manager¹), we conducted one-on-one feedback sessions with each participant, presenting both an aggregated, individual result as well as data on the group norm, and providing a simple verbal summary of the findings. To avoid biasing the results, student assistants who were not fully aware of our research agenda presented these feedback data using pre-scripted verbal summaries generated on the basis of the numerical results and focusing on areas for improvement. Given the espoused developmental nature of the process, participants were encouraged to ask questions and seek clarification from the human resource manager.

RESEARCH DESIGN AND PROCEDURES

In addition to administering the peer assessment instrument to non-supervisory plant personnel three times during the course of 18 months, we also had department supervisors periodically assess the performance of each of their subordinates using the measures described below. These supervisory ratings were at no time presented to the plant employees. At Time 1, about 1 month after project approval, supervisors were asked to appraise each of their subordinates according to the parameters specified below (See Appendix 2 for a timeline of experimental procedures). Supervisors were informed that these data were being collected for research purposes only and would at no time be reported back to their subordinates. Two weeks later, we conducted our first peer assessment, feeding the results back to the participants about 3 to 4 weeks later. Supervisors were at no time given access to the completed but unaggregated peer assessment forms. At Time 2, about 5 months after participants received their feedback, supervisors were again asked to appraise their subordinates' performance during the previous 5-month period. Two weeks after these data were collected, we conducted a second round of peer assessment, with feedback given about 3 to 4 weeks later. Time 3 occurred about 5 months after subordinates received their peer assessment feedback from the previous round. Again, supervisory ratings of performance were collected. A final round of peer assessment was conducted about 2 weeks later, followed by the presentation of feedback and post-hoc interviews with 5 supervisors. Supervisors were not informed about the precise nature of our study and, based on the post-hoc interview data collected, had no idea of the way in which peer assessment data were collected in their respective departments.

DEPENDENT VARIABLES

We asked supervisors to rate their subordinates along six criteria—four relating to the individual's contribution to group processes and two relating to individual effectiveness—using a rating instrument developed and validated by the Israel Productivity Institute (Oldstein et al., 1981) and selected by management. We evaluated the instrument for site-specific content validity (i.e., criterion relevance, comprehensiveness, and clarity at the specific enterprise studied) on the basis of the ethnographic and critical incident data described above. Specifically, we assessed the degree to which each of the six scales assessed its respective domain of interest (Hinkin, 1995, p. 968) by generating a list of key performance criteria suggested by the critical incident data and then ensuring that each issue was covered by at least one scale item.

The four criteria relating to individual contribution to group processes—namely, group task motivation, mentoring, teamwork, and individual initiative—focused on employees' team-enhancing behaviors. Group Task Motivation (4 items) included such items as "The individual does more than is required without asking anything in return." Mentoring (4 items) included such items as "The individual is able to teach others." Teamwork (3 items) included such items as "The individual shares information with other team members." Finally, Individual Initiative (4 items) included such items as "The individual solves work problems on his/her own initiative." The two criteria relating to individual effectiveness focused on team members' productivity-related work behaviors and overall work performance. The productivity-related Work Behaviors Scale (5 items) included such items as "The individual arrives to work on time." The Work Performance Scale (4 items) included such items as "The individual performs work of high quality and with minimal waste or scrap."

To ensure the structural integrity of our dependent variables (i.e., construct validity), we subjected the approximately 120 observations collected for each dependent variable at Time 1 to a confirmatory factor analysis. We evaluated model fit on the basis of three indexes of absolute fit (χ^2 , RMR and RMSEA), two indexes of comparative fit (IFI and CFI), and one index of parsimonious fit (PNFI). Although the model χ^2 (235.86) is significant, this cannot be taken as evidence of poor model fit because the sample size is approximately at that point at which "the fit of the model is likely to be rejected almost every time" (James, Mulaik, & Brett, 1982, p. 151). Furthermore, it should be noted that the model χ^2 is also less than two times the model degrees of freedom ($df = 143$). As Tabachnik and Fidell (2001, p. 721) note, such a ratio gives a rough indication that the model may fit the data. Moreover, along each of the remaining parameters, using the cutoff criteria recommended by Kelloway (1998), the model generated adequate fit results (RMR = .08, RMSEA = .09, IFI = .90, CFI = .90, and PNFI = .65). Further evidence of construct validity stems from the reliability estimates (alphas) reported along the diagonal in Table 2. Evidence of convergent validity is apparent from the high degree of correlation between a number of the managerial-based assessments and their parallel (e.g., teamwork) or similar (motivation and initiative on one hand and peer-assessed work attitudes, performance, and professionalism on the other) peer assessments at Time 1 (all significant at the $p < .05$ level or greater; See Table 1). Finally, whereas, as Schwab (1980, p. 18) notes, convergence provides only limited evidence of criterion-related validity, the relatively high ($r > .50$) correlation between, on one hand, the two supervisory-assessed performance-related outcomes (i.e.,

TABLE 2
Correlations Between Variables by Condition With Cronbach's Alpha Reliability

	Motivation	Mentoring	Teamwork	Initiative	Behavior	Performance	Professionalism	Efficiency	Teamwork	Attitudes	M	SD
Nonanonymous at Time 1 (n = 64)												
Motivation	0.85											
Mentoring	0.34**	0.89										
Teamwork	0.66***	0.45***	0.69									
Initiative	0.70***	0.19	0.56***	0.93								
Behavior	0.39***	0.50***	0.41***	0.18	0.73							
Performance	0.65***	0.27*	0.68***	0.44***	0.36**	0.88						
Peer-assessed												
professionalism	0.44***	0.33**	0.65***	0.54***	0.38***	0.54***	0.89					
Peer-assessed efficiency	0.32**	0.33**	0.54***	0.29*	0.43***	0.54***	0.86***	0.83				
Peer-assessed teamwork	0.22	0.43***	0.47***	0.11	0.45***	0.47***	0.66***	0.81***	0.86			
Peer-assessed work attitudes	0.38**	0.57***	0.65***	0.33**	0.67***	0.54***	0.78***	0.75***	0.73***	0.87		
Anonymous at Time 1 (n = 77)												
Motivation	0.91											
Mentoring	0.26	0.80										
Teamwork	0.66***	0.56***	0.66									
Initiative	0.62***	0.35**	0.64***	0.94								
Behavior	0.26*	0.12	-0.12	0.00	0.82							
Performance	0.76***	0.50***	0.57***	0.62***	0.22	0.85						
Peer-assessed												
professionalism	0.52***	0.33**	0.42***	0.46***	0.06	0.63***	0.83					
Peer-assessed efficiency	0.46***	0.08	0.31**	0.44***	0.00	0.58***	0.81**	0.84				
Peer-assessed teamwork	0.31**	0.49***	0.37**	0.39***	0.00	0.48***	0.73***	0.67***	0.88			
Peer-assessed work attitudes	0.47***	0.36**	0.29*	0.40***	0.25*	0.52***	0.74***	0.66***	0.70***	0.87		

NOTE: Cronbach's alphas are shown along the diagonal in bold.
* p < .10. ** p < .05. *** p < .01.

productivity-related behaviors and overall work performance) and, on the other, company-provided data on individual absenteeism, tardiness, and scrap rate available for approximately half of the workers in the sample, is, at the very least, suggestive of these ratings' criterion-related validity.

ANALYSIS

To test Hypotheses 1 through 3, we ran a series of analyses using a mixed model. As Cohen, Cohen, West, and Aiken (2003, p. 567) point out, "when data are clustered such that individuals within clusters are more like one another than are randomly selected individuals, bias is introduced into inference in regression." Such clustering violates a basic assumption underlying the general linear model, namely the assumption of nonindependence among a set of observations. By adding code predictors that identify such clusters and account for cluster differences, the multilevel or mixed model facilitates the examination of the cross-level interaction between variables that occur at different levels of aggregation and provides estimates that take such biases into account (Cohen et al., 2003). In the current study, the use of a mixed model was necessary because single supervisors provided ratings on all subordinates within their units, causing all within-unit observations to be correlated. Indeed, the random effects of supervisor were statistically significant ($p < .10$) with respect to five of the six supervisory-assessed outcomes when included in a preliminary test of the mixed model. Using the mixed model, and therefore taking these random effects into account, we tested Hypotheses 1 and 2 on the basis of the t values generated for the fixed effects of condition in each of the three time periods. We tested Hypothesis 3 in a similar manner, again using the mixed model, only this time focusing on (a) the F value associated with the fixed effect of the Time \times Condition interaction and (b) a comparison of the variable-specific effect sizes across the three time periods.

RESULTS

Using a dummy variable for employee type (hired worker versus kibbutz member), we ran a series of time-specific effects analyses to examine whether mean supervisory and peer assessments of kibbutz members were significantly different from those of hired employees in any of the three time periods. In each of the 30 mixed models (3 time periods \times 10 variables)² tested, we included the fixed effects of anonymity as well as the random effects of supervisor. No significant differences in mean ratings for members versus nonmembers were found with respect to 5 of the 6 supervisory-

assessed criteria or any of the 4 peer-assessed criteria. Only in the case of supervisory-assessed performance were kibbutz members' ratings significantly different ($t = 2.48, p < .05$) from those of nonmembers at Time 1, but this difference was no longer statistically significant in either Time 2 or Time 3. Moreover, within the context of a multiple comparison, a p value of greater than .01 in one of 30 comparisons cannot reliably be taken as evidence of a true effect. Finally, a similar multiperiod analysis confirmed that employee type had no statistically significant effect with regard to any of the supervisory- or peer-assessed criteria (F ranged from 0.0 to 2.50, $p > .10$, with respect to all 10 supervisory- and peer-assessed outcomes). Consequently, whereas we included a supervisor variable in the mixed models analyzed below, these models excluded the random effects of employee type.

MAIN EFFECTS OF ANONYMITY IN PEER ASSESSMENT

To test Hypothesis 1 (which specified that mean peer assessment ratings would be higher under conditions of nonanonymity relative to conditions of anonymity), we compared the mean composite³ peer assessments of all those employed in departments assigned to the anonymous condition with those of individuals employed in departments assigned to the nonanonymous condition using the t statistics generated on the basis of the mixed model. Regardless of whether self-assessments were included or excluded, the composite peer assessment scores at Time 1 for workers in the nonanonymous departments were not significantly different from those in anonymous departments (see bottom of Table 3). Furthermore, as indicated by the a priori contrast results (see right side of Table 3) regardless of condition, peer assessment scores increased significantly across both subsequent appraisal periods. However, consistent with Hypothesis 1, the results indicate that following the implementation of peer assessment, the mean composite peer ratings received by those assigned to the nonanonymous condition were significantly higher than those assigned to the anonymous condition in both Time 2 and Time 3. Moreover, as indicated by the significant Time \times Condition interaction ($F = 11.82, p < .01$ when self-assessments were included), the effect of anonymity on peer ratings increased over time.

To test Hypothesis 2, we compared the mean supervisory assessments of all those employed in departments assigned to the anonymous condition with those of individuals employed in departments assigned to the nonanonymous condition using the t statistics generated on the basis of the mixed model. As can be seen in Table 3, along all supervisory-rated criteria, mean scores at Time 1 for workers in departments assigned to the nonanonymous condition were not significantly different from those in anonymous departments.

TABLE 3
Descriptive Statistics, *t* Tests, and Repeated Measures ANOVA for Supervisory-Assessed Team Processes and Outcomes and for Peer Assessments^a

Variable	Time 1						Time 2						Time 3						A priori contrast				Repeated measure ANOVA Time × Condition		
	Nonanon.		Anon.		T	SD	Nonanon.		Anon.		T	SD	Nonanon.		Anon.		T	SD	T2-T1		T3-T1			T	F
	M	SD	M	SD			M	SD	M	SD			M	SD	M	SD			M	SD	M	SD			
Supervisory assessment																									
Group process—Motivation	3.74	0.89	3.57	1.03	0.63	4.03	0.79	3.58	1.01	2.03**	4.41	0.78	3.96	0.90	2.65***	1.08	5.79***	6.60***	1.42						
Group process—Mentoring	3.84	0.81	3.87	0.81	0.34	4.03	0.81	3.81	0.84	1.19	4.30	0.76	3.91	0.85	2.44***	0.82	3.41***	3.80***	1.18						
Group process—Teamwork	3.84	0.83	3.66	0.73	0.80	4.07	0.80	3.63	0.84	1.82**	4.44	0.59	3.98	0.77	2.07**	0.55	4.81***	5.57***	1.08						
Group process—Initiative	3.57	0.89	3.55	0.98	0.12	4.06	0.74	3.56	0.99	2.02**	4.35	0.84	3.89	0.93	1.71**	2.25***	5.43***	4.49***	1.02						
Individual performance—																									
Productivity-related work behavior	4.40	0.57	4.39	0.57	0.04	4.57	0.52	4.23	0.67	1.69**	4.82	0.29	4.45	0.57	2.92***	0.14	4.80***	4.94***	5.06***						
Individual performance—																									
Overall work performance	4.05	0.86	3.91	0.83	0.77	4.25	0.66	4.00	0.81	1.52	4.56	0.55	4.21	0.67	2.89***	1.74	5.01***	3.80***	1.15						
Composite peer assessment (including self-assessments in the case of the anonymous condition) ^b :	4.91	0.69	4.88	0.60	0.19	5.18	0.51	4.94	0.54	2.35**	5.29	0.48	4.83	0.64	4.25***	4.31***	3.81***	0.11	11.82***						
Composite peer assessment (excluding self-assessments in the case of the anonymous condition) ^b :	4.91	0.69	4.82	0.66	0.70	5.18	0.51	4.88	0.61	2.79***	5.29	0.48	4.83	0.71	4.04***	4.34***	4.47***	0.61	8.45***						

a. All *t* and *F* statistics are evaluated on the basis of Satterthwaite degrees of freedom. Given the directional nature of our hypotheses, all *t* tests evaluated on the basis of a one-tailed test.
b. Given the sizable correlations between the four peer assessment measures and in order to simplify this table, we present the results for two composite variables constructed on the basis of the average peer ratings across all four peer-assessment measures, one including and one excluding self-assessment scores.
p* < .05. *p* < .01.

Furthermore, the a priori contrast results in this table indicate that, regardless of condition, scores significantly increased over time for all six supervisory-related criteria. However, the results also indicate that with the implementation of peer assessment in Time 1, the mean supervisory-based criterion scores received by those assigned to the nonanonymous condition and those assigned to the anonymous condition began to diverge. Specifically, at Time 2, consistent with Hypothesis 2, mean supervisory assessments of those employed in departments in which peer assessment was conducted on a non-anonymous basis were significantly higher ($p < .05$) than those employed in departments in which the assessment was conducted anonymously with respect to three out of the four individual-contribution-to-group-process criteria (initiative, motivation, and teamwork). Moreover, these significant differences were maintained with respect to all three of these variables through Time 3 (motivation and teamwork), with significant differences appearing with respect to the fourth group-process factor—mentoring—in Time 3.

Similarly, although at Time 1 there was no difference in the mean supervisory-assessed productivity and performance scores of individuals employed in departments assigned to the two conditions, subsequent to the implementation of peer assessment, significant differences did develop over time with respect to both of these individual effectiveness criteria. Whereas the mean level of supervisory-assessed productivity-related behaviors fell for individuals in the anonymous condition between Time 1 and Time 2, for those in the nonanonymous condition, the mean level actually rose during the same time period. This difference was statistically significant in Time 2 ($t = 1.69, p < .05$). Although the mean score on productivity-related behaviors rose regardless of condition between Times 2 and 3, the difference between the two mean scores increased and became statistically significant by Time 3 ($t = 2.92, p < .01$). Similarly, a significant difference in mean overall work performance ratings also emerged only at Time 3 ($t = 2.89, p < .01$), suggesting a possible lag in the consequences of anonymous peer assessment on supervisory-assessed individual effectiveness.

Taken in combination, these results provide strong support for Hypothesis 2. Specifically, it appears that the implications of peer evaluation on cooperative group processes are influenced by whether ratings are provided on an anonymous or nonanonymous basis. The significance and consistency of these findings are particularly noteworthy given the upward biasing of anonymous ratings by self-ratings noted earlier. As can be seen in the composite peer ratings appearing at the bottom of Table 3, the inclusion of self-ratings tended to inflate the peer assessments provided back to the participants in the anonymous condition, thereby weakening the manipulation. The fact that

such consistent and significant differences emerged and were maintained over time nonetheless appears to attest to the robustness of these effects.

TIME \times CONDITION EFFECTS

However, as may also be seen in Table 3, more limited support was found for Hypothesis 3, which specified that the difference in supervisory-assessed outcomes under conditions of rater anonymity versus nonanonymity would increase in magnitude over time. Specifically, a significant Time \times Condition effect was found with respect to only a single supervisory-assessed outcome, namely, productivity-related behaviors ($F = 5.06$, $p < .01$). Still, the failure of the F statistic to reach statistical significance for most of the Time \times Condition models may also stem from the interaction of two methodological factors. Specifically, as noted earlier, the dependent variables were assessed on the basis of relatively coarse rating scales. Such scale coarseness may cause severe information loss and may "greatly reduce the probability of detecting true interaction effects" (Russell & Bobko, 1992, p. 336). This problem may be further exacerbated by the assessment of our criteria over only three points in time, thus setting up an extremely conservative context for the testing of a Time \times Condition interaction and hence increasing the likelihood of a Type II error.

Consequently, in addition to evaluating Hypothesis 3 on the basis of each model's F statistic, for each of the supervisory-assessed outcomes, we also compared condition effect sizes across the three time periods. The effect size was assessed as the standardized difference between means. Given the heterogeneity of variances, a pooled estimate was used to obtain an estimate of the standard deviation (Grissom & Kim, 2001). The variance estimate included the two variance components estimated in the model, namely, the random effects (of supervisor) as well as the pure error. The resulting effect sizes (see Table 4) indicate that, as hypothesized, the positive effect of rater nonanonymity did increase in magnitude over time. Effect size between Time 1 and Time 2 increased with respect to all six supervisory-assessed outcomes. Evidence of increasing effect size between Times 2 and 3 was also found with respect to four of the six outcomes. With respect to all six criterion measures, by Time 3, the effect of anonymity increased to approximately 0.5 standard deviations or more. Indeed, in the case of supervisor-assessed productivity behaviors (the one criterion measure for which the Condition \times Time interaction was significant), the effect size rose from 0.01 at Time 1 to 0.80 at Time 3. In sum, the pattern of effect sizes over time displayed in Table 3 suggests more broadscale support for Hypothesis 3.

TABLE 4
Change in Effect Size Over Time

<i>Variable Name</i>	<i>Time 1</i>	<i>Time 2</i>	<i>Time 3</i>
Supervisor-assessed individual contribution to group process— Motivation	0.16	0.53	0.53
Supervisor-assessed individual contribution to group process— Mentoring	0.11	0.33	0.49
Supervisor-assessed individual contribution to group process— Teamwork	0.22	0.55	0.64
Supervisor-assessed individual contribution to group process— Initiative	0.03	0.57	0.47
Supervisor-assessed effectiveness—Productivity-related work behavior	0.01	0.57	0.80
Supervisor-assessed effectiveness—Overall work performance	0.22	0.44	0.58
Peer assessment—Composite (<i>including</i> self-assessment)	0.04	0.44	0.81
Peer assessment—Composite (<i>excluding</i> self-assessment)	0.13	0.52	0.77

DISCUSSION

The primary objective of our study was to assess the extent to which rater anonymity might explain contradictory behavioral outcomes associated with peer assessment in the research literature. The results presented above suggest that the elimination of rater anonymity in the appraisal procedure may indeed explain such contradictory results. Specifically, our results suggest that the use of appraisal procedures in which raters are required to identify themselves may enhance the longer-term, behavioral implications of peer assessment.

Contrary to some of the research results presented at the beginning of this paper, our data suggest that peer assessment need not be associated with any subsequent decline in either group processes or performance. Indeed, regardless of condition, we found a general improvement in supervisory ratings of subordinates subsequent to the implementation of peer assessment, with strong and significant ($p < .01$ in every case) time effects between Time 2 and Time 3, and between Time 1 and Time 3. These results are consistent with the findings of Druskat and Wolff (1999), suggesting that peer assessment may have generally positive effects on individual contributions to group processes and performance. On the other hand, this unconditional improvement is not surprising given the fact that prior to the experimental manipulation studied here, no system of formal performance appraisal had ever been implemented in this organization. Therefore, we may assume that some of the improvement may have been a function of the adoption of a performance

appraisal system. Nevertheless, due to the lack of a control group, we were unable to determine the extent to which this improvement is directly attributable to peer assessment, the introduction of a system of performance in general, or to other factors such as response shift bias (Martineau, 1998).

We also found no evidence that anonymous peer assessment harmed relationships and impaired group task focus and functioning. Rather, we found consistent evidence that this improvement in managerial ratings of subordinates' attitudes and behaviors over time was greater among individuals working in departments assigned to the nonanonymous assessment condition than among those assigned to the anonymous condition.

ALTERNATIVE EXPLANATIONS

Although we lack the data to fully understand the social dynamics underlying these findings, a number of possible explanations come to mind. For example, a tendency toward more positive peer ratings among members of nonanonymous departments may have contributed to an enhanced sense of camaraderie among these workers, and this in turn may have generated the more positive social consequences reported above. Similarly, the opposite tendency among those in the anonymous condition may have made these employees angry and not motivated to change. However, the proportion of generally positive ratings was nearly similar in both conditions. Excluding self-reports, positive ratings (i.e., 4 to 6 on the rating scale) accounted for an average of 92.6% and 95.1% of the total number of ratings given in the anonymous and nonanonymous conditions, respectively. Moreover, it does not appear that these minor differences had any significant impact on camaraderie. Specifically, although we did not measure camaraderie directly, using O'Reilly, Caldwell, and Barnett's (1989) instrument, we did assess employee perceptions of group cohesiveness ($\alpha = .84$) at approximately the same three points in time that supervisors provided appraisal data. Mean cohesiveness scores—nearly identical for the two groups at Time 1 (3.86 and 3.92, $t = 0.23$)—were slightly higher for those in the nonanonymous group at Times 2 (4.15 vs. 4.09). However, they were lower at Time 3 (4.21 vs. 4.28). Moreover, at no time were these differences statistically significant ($t = 0.39$ and 0.49 at Times 2 and 3, respectively).

Alternatively, it may be that a nonanonymous rating process generated more positive social consequences because such an approach forces raters to gather more evidence and to become more conscientious when conducting peer assessment. If, as a result, peers provided more reliable and valid assessment feedback to one another, this may be what in fact underlies the enhanced process and performance outcomes manifested. Still, the results

presented in Table 2 fail to suggest that the peer assessment measures were consistently more reliable or valid when used under conditions of non-anonymity than when used under conditions of anonymity. Cronbach's alpha coefficients were generally no higher for the peer assessment measures in the nonanonymous condition than in the anonymous condition, and there is no clear pattern of higher correlations between supervisor and peer ratings in the nonanonymous condition relative to the anonymous condition.

Finally, our results may have been generated by a tendency of members of anonymous departments to assume that their individual scores would be downwardly biased by peers engaging in some sort of tactical and self-serving distortion. Consequently, in time, members of anonymous departments may have become increasingly convinced that they had less to gain by going out of their way to work hard and assist their coworkers because, regardless, the latter would have an incentive to tactically down-rate their peers. As a result, it is conceivable that any positive effects of peer assessment (i.e., reduced social loafing) were simply weakened, resulting in lower than expected improvements in teamwork and cooperation. In contrast, given the nature of the rating process in nonanonymous departments, it is far less likely that peers felt any major risk of becoming a sucker. Indeed, whereas the mean peer ratings of those assigned to the anonymous departments remained relatively stable, the mean peer ratings of those assigned to the nonanonymous departments generally increased in time, with the interaction effect being significant ($p < .01$) for both the composite variable (see Table 3) and the individual peer assessment dimensions. It is possible that, in time, workers in the nonanonymous condition learned that enhanced cooperation was recognized and valued by their peers. Moreover, in time, these workers may have learned that rather than using the peer assessment system as a mechanism for tactical self-enhancement, their peers used the system as initially intended, namely, to provide positive feedback where positive feedback was due. Consequently, in contrast to anonymous peer assessment, nonanonymous peer assessment may have been more effective in reinforcing precisely those behaviors that the intervention was designed to influence in the first place.

In sum, although not fully discounting other possible explanations, the peer assessment data presented in Table 3 support the claim that when workers are at least somewhat suspicious that peer ratings may eventually influence managerial administrative decisions, non anonymity may reduce the risk of self-enhancement bias and game playing on the part of peer raters, thus providing a stronger context for the potentially positive consequences of peer assessment to take effect. On the other hand, these data also support the

claim made by many of those discussing the disadvantages of non-anonymous assessment systems, namely that non anonymity may be associated with inflated ratings (Antonioni, 1994; London & Smither, 1995). Yet, the degree to which these ratings are in fact inflated (as opposed to only *appearing* to be inflated when compared to anonymous ratings) is an empirical question. Indeed, as Mero and Motowidlo (1995) suggest, the non-anonymous raters' assessments may reflect "true" performance, and thus the "real" improvements resulting from the adoption of peer assessment as well. In contrast, the anonymous ratings, because of the built-in incentive to score peers in a more tactical manner, may fail to reflect "true" performance. The assessments provided by the anonymous raters may fail to capture any initial improvements in group processes or individual performance due to the built-in "need" for anonymous raters to downwardly bias their peers' scores to enhance their own relative ratings and to avoid emerging as a sucker. This may, in turn, result in a kind of vicious circle in which deflated peer assessments at Time 2 weaken any incentive for workers to avoid social loafing and enhance their cooperative efforts, thus resulting in further deflated peer assessments (albeit this time perhaps more justified) in subsequent rounds.

In a post-hoc analysis, we attempted to assess the relative accuracy of nonanonymous versus anonymous ratings by comparing a subject's individual ratings of a given target with the aggregate, team-based score received by that same target (most likely, a less biased assessment). These analyses in fact confirmed that nonanonymous ratings were no less strongly associated with aggregated peer assessments than were anonymous ratings and vice versa. Specifically, using a multilevel modeling approach (necessary due to the nesting of individual assessments in team assessments) and data from two points in time (Times 2 and 3), we regressed subjects' individual ratings of a target on experimental condition (anonymous versus nonanonymous), aggregated group ratings of that same target and an interaction term reflecting the possible moderating effect of rater anonymity on the relationship between individual and aggregate ratings. Although aggregated group peer assessments were significantly associated ($p < .01$) with individual assessments at both times, at neither point in time did we find the parameter estimate of the interaction term to be statistically significant, indicating that rater anonymity has little impact on the degree of association between individual ratings and aggregated team-level ratings. Consequently, consistent with the argument presented by Mero and Motowidlo (1995), our findings suggest that, regardless of their apparent inflatedness, the assessments provided by the nonanonymous raters may actually be no less accurate than those provided by the anonymous raters.

LIMITATIONS AND FUTURE DIRECTIONS

Although our data suggest that the development of a competitive orientation on the part of anonymous peer raters may underlie the effects reported above, our data provide us with only a limited means by which to rule out alternative explanations. Given the tendency of most researchers in the area of multisource feedback to come out in favor of anonymous feedback (e.g., Tornow & Tornow, 2001), future researchers in this area should attempt to develop experimental designs with the potential to more rigorously test these and other alternative explanations. For example, researchers may want to expand on the current study by assessing the degree to which perceived or actual bias in ratings received by a target in previous rounds may be more or less predictive of biasing in that target's own subsequent ratings of others, depending on the level of anonymity in the rating system.

Researchers may also wish to further explore the degree to which the perceived nature of the peer assessment system (developmental versus evaluative) influences the impact of rater anonymity. Given our assumption that employees tend to be at least somewhat suspicious as to management's true intent with respect to peer assessment, we adopted a design in which the somewhat ambiguous nature of the system was held constant. In the future, researchers might want to adopt a 2×2 design, manipulating both rater anonymity and system type (i.e., developmental vs. evaluative). Such a design would allow researchers, for example, to determine the degree to which the potential leniency effects of non anonymity are influenced by the developmental or evaluative nature of the rating system.

Alternatively, researchers may want to examine the moderating role of collectivistic versus individualistic organizational cultures on the impact of anonymity on peer assessment. It may be that the negative impact of anonymity is more severe in more collectivistic organizational cultures than in individualistic cultures because, in the former, such assessment systems are likely to destroy pre-existing meta-norms of cooperation (Axelrod & Dion, 1988).

However, before any of these avenues of future research are pursued, there is a need for the results of the current study—a single-case field experiment—to be replicated. Such replication is imperative given the contextual and methodological shortcomings of the current study. Specifically, the findings reported above were based on data collected in a small company owned by an Israeli kibbutz in which no formal performance assessment system had even been implemented. Given the unique, collectivist nature of Israeli kibbutzim, as well as our inability to determine the extent to which the lack of any pre-existing evaluation system increased or decreased the salience of

peer assessment, these findings may offer limited generalizability to larger, more traditionally owned and managed firms in the United States and elsewhere. Furthermore, due to the field-based nature of this study, we cannot rule out the possibility that some supervisors may have, formally or informally, learned about the aggregated peer assessment results from one or more of their subordinates or colleagues, or from simply gaining access to a subordinate's personnel file. Consequently, although our post-hoc interview data with five supervisors suggest that this did not occur, there is some risk of confounding with respect to several of our primary criterion measures. There is also the possibility that because peer feedback was not limited to the formal rating process, informal discussions among peers about performance expectations, demands, and gaps may have also influenced behavior and performance outcomes. Such informal feedback processes have the potential to generate both positively and negatively biased ratings (perhaps depending on the nature of the informal feedback) to an unknown degree. Because we did not directly assess such informal processes or the attitudes and perceptions potentially generated by them, we were unable to specify the extent to which they may have biased our results in either direction. Finally, although our qualitative data suggest otherwise, given the lack of a quantitative assessment, there remains the possibility that our manipulation was unsuccessful and that those in the nonanonymous condition felt no more accountable than those in the anonymous condition. Neither can we rule out the possibility that beliefs about the developmental or administrative use of the data systematically varied by condition.

In light of these limitations, as well as the lack of a control group, we recommend that our findings be taken with a degree of caution. Although they suggest that by moving away from the standard of rater anonymity in peer assessment systems, it may be possible to generate more positive effects on group processes and outcomes, clearly more research is necessary.

CONCLUSION

Our understanding of the factors that influence the impact of peer assessment on employee behavior is limited. Positive and negative effects are likely to be driven by a wide variety of factors. Nevertheless, the current analysis suggests that the effect of one variable, namely rater anonymity, should not be ignored. Specifically, our results suggest that whereas anonymous peer appraisal procedures may be well institutionalized in organizations and in fact preferred by raters, their utility should not be taken as a given. Indeed, although anonymous rating procedures may enhance the acceptability of the

system to raters in the short run, in the long run, such procedures may place a limit on the overall utility of peer assessment to these same individuals as ratees, as well as to the organizations employing them. Consequently, independent of the relative importance of the other factors that may influence the behavioral consequences of peer assessment, the results of our longitudinal field experiment suggest that the elimination of rater anonymity may generally be expected to have a positive effect.

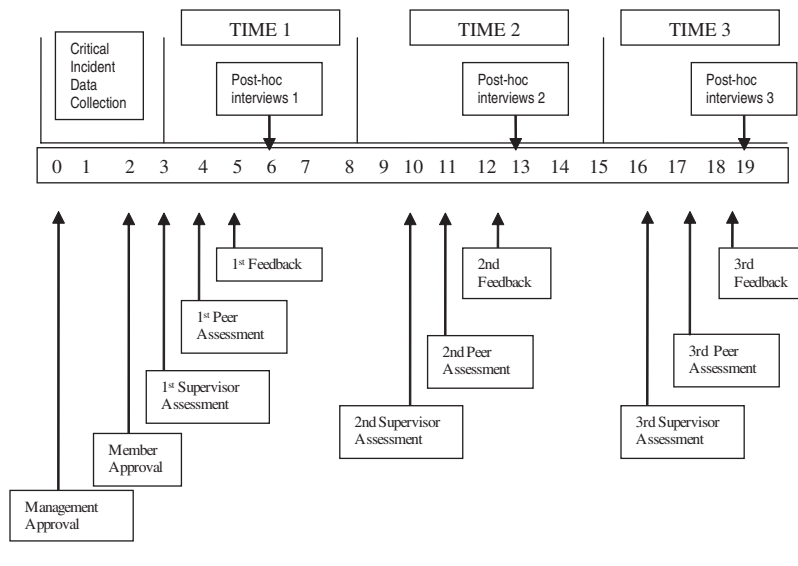
Finally, whereas practitioners may tend to view peer assessment simply as an alternative to more traditional supervisory-based appraisal systems, because of its peer-based nature, the broader social consequences of its adoption cannot be neglected. Although a number of studies have examined some of these broader social consequences, we still know relatively little about the impact of peer assessment on group processes and outcomes. Our findings suggest that simply evaluating the validity, reliability, and user acceptability of peer assessment is not sufficient. Organizations adopting such practices should also closely monitor the system's impact on group processes and group member relations, as well as its implications on both individual and group performance.

APPENDIX 1

Questions Posed in post-Hoc Interviews With Employees:

1. Why do you believe management decided to adopt a system of peer assessment?
 2. How do you think the peer assessment data will be used in the short term? In the long term?
 3. From your understanding of the peer assessment system as it has been implemented to date, how similar are the assessment instruments completed by employees in different departments?
 4. Despite all the precautions taken to ensure confidentiality, to what degree are you concerned that there may be some "backlash" to one of the ratings that you gave a peer; that is, that someone in the factory may criticize you for one of the ratings that you gave to someone else?
-

APPENDIX 2 Timeline of Study



NOTES

1. At the request of the CEO, other managers, including supervisors, were invited to observe the final feedback sessions for training purposes. However, as all supervisors had by then already submitted their final round of assessments, there was no direct risk of criterion contamination.
2. Six supervisory- and four peer-assessed variables.
3. Given the sizable correlations between the four peer-assessment measures, we present the results for two composite variables constructed on the basis of the average peer ratings across all four peer-assessment measures, one including and one excluding self-assessment scores.

REFERENCES

- Albanese, R., & Van Fleet, D. (1985). Rational behavior in groups: The free-riding tendency. *Academy of Management Review*, *10*, 244-255.
- Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology*, *47*, 349-356.
- Antonioni, D., & Park, H. (2001). The relationship between rater affect and three sources of 360-degree feedback ratings. *Journal of Management*, *27*, 479-495.

- Atwater, L., Waldman, D., Atwater, D., & Cartier, P. (2000). An upward feedback field experiment: Supervisors' cynicism, reactions, and commitment to subordinates. *Personnel Psychology, 53*, 275-297.
- Axelrod, R., & Dion, D. (1988). The further evolution of cooperation. *Science, 242*, 1385-1390.
- Bamberger, P., & Meshoulam, I. (2002). *Human resource strategy: Formulation, implementation and impact*. Thousand Oaks, CA: Sage.
- Brutus, S., Fleenor, J., & London, M. (1998). Elements of effective 360-degree feedback. In W. W. Tornow, M. London, & C. C. L. Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 11-27). San Francisco: Jossey-Bass.
- Cederblom, D., & Lounsbury, J. W. (1980). An investigation of user acceptance of peer evaluations. *Personnel Psychology, 33*, 567-579.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Lawrence Erlbaum Associates.
- Deadrick, D. L., Bennett, N., & Russell, C. J. (1997). Using hierarchical linear modeling to examine dynamic performance criteria over time. *Journal of Management, 23*, 745-757.
- Deadrick, D. L., & Madigan, R. M. (1990). Dynamic criteria revisited: A longitudinal study of performance stability and predictive validity. *Personnel Psychology, 43*, 717-744.
- DeNisi, A. S., Randolph, W. A., & Blencoe, A. (1983). Potential problems with peer ratings. *Academy of Management Journal, 26*, 457-464.
- DeNisi, A. S., & Stevens, G. (1981). Profiles of performance, performance evaluation and personnel decisions. *Academy of Management Review, 24*, 592-602.
- Druskat, V. U., & Wolff, S. B. (1999). Effects and timing of developmental peer appraisals in self-managing work groups. *Journal of Applied Psychology, 84*, 58-74.
- Erev, I., & Roth, A. (1998). Predicting how people play games: Reinforcement learning processes on games with unique mixed strategy equilibrium. *American Economic Review, 88*, 848-881.
- Erez, A., Lepine, J. A., & Elms, H. (2002). Effects of rotated leadership and peer evaluation on the functioning and effectiveness of self-managed teams: A quasi-experiment. *Personnel Psychology, 55*, 929-948.
- Fedor, D., & Bettenhausen, K. (1989). The impact of purpose, participant preconceptions and rating level on the acceptance of peer evaluations. *Group and Organization Studies, 14*, 182-197.
- Fedor, D.B., Bettenhausen, K., & Davis, W. (1999). Peer evaluations: Employees' dual roles as raters and recipients. *Group and Organization Management, 24*, 92-121.
- Fox, J., & Guyer, M. (1978). "Public" choice and cooperation in n-person prisoner's dilemma. *Journal of Conflict Resolution, 22*, 468-481.
- Fusilier, M. R. (1980). The effects of anonymity and outcome contingencies on rater beliefs and behavior in a performance appraisal situation. *Proceedings of the 40th Annual Meeting of the Academy of Management, 273-277*.
- Grissom, R. J., & Kim, J. J. (2001). Review of assumptions and problems in the appropriate conceptualization of effect size. *Psychological Methods, 6*, 135-146.
- Hazucha, J. F., Hezlett, S. A., & Schneider, R. J. (1993). The impact of 360-degree feedback on management skills development. *Human Resource Management, 32*, 325-351.
- Hedge, J. W., Borman, W. C., & Birkeland, S. A. (2001). History and development of multisource feedback as a methodology. In D. Bracken, C. W. Timmreck, & A. H. Church (Eds.), *The handbook of multisource feedback* (pp. 15-32). San Francisco: Jossey-Bass.
- Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management, 21*, 967-988.

- James, L. R., Mulaik, S. A., & Brett, J. M. (1982). *Causal analysis: Assumptions, models and data*. Beverly Hills, CA: Sage.
- Kane, J., & Lawler, E. (1978). Methods of peer assessment. *Psychological Bulletin*, *85*, 555-586.
- Kelloway, E. K. (1998). *Using LISREL for structural equation modeling*. Thousand Oaks, CA: Sage.
- Kerr, N., & Brunn, S. (1981). Ringelann revisited: Alternate explanations for the social loafing effect. *Personality and Social Psychology Bulletin*, *7*, 224-231.
- Latham, G. P., & Wexley, K. N. (1977). Behavioral observation scales for performance appraisal purposes. *Personnel Psychology*, *30*, 255-268.
- Leslie, J. B., Gurskiewicz, N. D., and Dalton, M. A. (1998). Understanding cultural differences on the 360-degree feedback process. In W. Tornow, M. London, & C. C. L. Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 196-216). San Francisco: Jossey-Bass.
- London, M., & Smither, J. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations and performance related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, *48*, 803-839.
- London, M., Smither, J., & Adsit, D. (1997). Accountability: The Achilles heel of multi-source feedback. *Group and Organization Management*, *22*, 162-184.
- Luce, R., & Raiffa, H. (1957). *Games and decisions*. New York: Wiley.
- Martineau, J. W. (1998). Using 360-degree surveys to assess change. In W. W. Tornow, M. London, & C. C. L. Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 217-248). San Francisco: Jossey-Bass.
- Maurer, T. J., & Tarulli, B. A. (1996). Acceptance of peer/upward performance appraisal systems: Role of work context factors and beliefs about managers' development capability. *Human Resource Management*, *35*, 217-241.
- 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.
- Mero, N. P., & Motowidlo, S. J. (1995). Effects of rater accountability on the accuracy and the favorability of performance ratings. *Journal of Applied Psychology*, *80*, 517-524.
- Murnighan, J. K., & Roth, A. R. (1983). Expecting continued play in prisoner's dilemma games: A test of several models. *Journal of Conflict Resolution*, *2*, 279-300.
- Murphy, K. R. (1989). Is the relationship between cognitive ability and job performance stable over time? *Human Performance*, *2*, 183-200.
- Oldstein, Z., Carmi, O., Meshulam, I., Saar, M., Rotenstreich, Y., & Rimer, A. (1981). *Manual for the human resource manager* (In Hebrew). Tel Aviv, Israel: Productivity Institute.
- O'Reilly, C., Caldwell, D., & Barnett, W. (1989). Work group demography, social integration and turnover. *Administrative Science Quarterly*, *34*, 21-37.
- Reilly, R. R., & Chao, G. T. (1982). Validity and fairness of some alternative employee selection procedures. *Personnel Psychology*, *78*, 113-123.
- Ross, M., & Sicoly, F. (1979). Egocentric biases in availability and attribution. *Journal of Personality and Social Psychology*, *37*, 322-336.
- Russell, C. J., & Bobko, P. (1992). Moderated regression analysis and Likert scales: Too coarse for comfort. *Journal of Applied Psychology*, *77*, 336-342.
- Saavedra, R., & Kwun, S. (1993). Peer evaluation in self-managing work groups. *Journal of Applied Psychology*, *78*, 450-462.
- Schwab, D. P. (1980). Construct validity in organizational behavior. *Research in Organizational Behavior*, *2*, 3-43.
- Tabachnik, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Boston: Allyn and Bacon.

- Tagger, S., & Brown T. C. (2001, August). *Interpersonal affect and rating errors following prior peer-feedback in teams*. Paper presented at the 61st annual meeting of the Academy of Management, Washington, D.C.
- Tjosvold, D. (1984). Cooperation theory and organizations. *Human Relations*, *37*, 743-767.
- Tornow, W. W., & London, M. (1998). Challenges and implications for maximizing 360-degree feedback. In W. W. Tornow, M. London, & C. C. L. Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 249-258). San Francisco: Jossey-Bass.
- Tornow, W. W., & Tornow, C. P. (2001). Linking multisource feedback content with organizational needs. In D. Bracken, C. W. Timmreck, & A. H. Church (Eds.), *The handbook of multi source feedback* (pp. 48-62). San Francisco: Jossey-Bass.
- Van Velsor, E. (1998). Designing 360-degree feedback to enhance involvement, self-determination and commitment. In W. W. Tornow, M. London, & C. C. L. Associates (Eds.), *Maximizing the value of 360-degree feedback* (pp. 149-195). San Francisco: Jossey-Bass.
- Wageman, R. (1995). Interdependence and group effectiveness. *Administrative Science Quarterly*, *40*(1), 145-180.
- Wagner, J. A. III. (1995). Studies of individualism-collectivism: Effects on cooperation. *Academy of Management Journal*, *38*, 152-172.
- Waldman, D. A., Atwater, L. E., & Antonioni, D. (1998). Has 360 degree feedback gone amok? *Academy of Management Executive*, *12*, 86-94.

Peter A. Bamberger (Ph.D. 1990, Cornell University) is an associate professor at the Faculty of Industrial Engineering and Management, Technion, and a visiting fellow at Cornell University. Current interests include peer relations in the workplace and employee well-being. Recent publications include Mutual Aid and Union Renewal with S. Bacharach and W. Sonnenstuhl (Cornell University Press, 2001) and Human Resource Strategy with I. Meshulam (Sage, 2000).

Ido Erev (Ph.D. 1990, University of North Carolina) is an associate professor at the Davidson Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology. Current research focuses on modeling and deriving the implications of human adaptation to economic incentives. Examples of implications include the design of rule enforcement and bonus systems, training methods, marketing strategies, and pricing.

Michal Kimmel completed her Master of Science degree at San Francisco State University and is currently a doctoral student in human resource management at the Davidson Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology. Her research focuses on the application of attachment theory to peer relations in the workplace.

Tali Oref-Chen completed her Master of Science degree in human resource management at the Davidson Faculty of Industrial Engineering and Management, Technion—Israel Institute of Technology in 2001. She is currently working as an organizational consultant specializing in the area of peer assessment.