Identifying Motivation and Interpersonal Performance Using Peer Evaluations

Michelle M. Zazanis  
U.S. Army Research Institute  
Alexandria, Virginia

Stephen J. Zaccaro  
Department of Psychology  
George Mason University

Robert N. Kilcullen  
U.S. Army Research Institute  
Alexandria, Virginia

Peer evaluations have historically shown high predictive validity, but the reason for this strength has not been clear. This research used an assessment center and subsequent training program to investigate the hypothesis that interpersonal performance and motivation are two key dimensions of performance that may account for the historical strength of peer assessments. Consistent with previous research, results showed that peer rankings from the assessment center predicted final training outcomes better than did staff ratings. Congruent with our hypothesis, results from the training program demonstrated that, when rating a student’s expected future on-the-job performance, peers placed significantly more importance on interpersonal performance and motivation than did staff, and they placed significantly less importance on task performance. Additional longitudinal research is needed to explicitly link peer evaluations of interpersonal skills and performance motivation with future job performance.

Decades of research reveal that peer evaluations are reliable (Gordon & Medland, 1964; Hollander, 1956, 1957) and are able to predict a variety of outcomes, includ-
ing managerial performance (Mayfield, 1970; Orpen, 1983), commissioned officer promotions (Downey, Medland, & Yates, 1976), noncommissioned officer promotions (Amir, Kovarsky, & Sharan, 1970), job advancement (Shore, Shore, & Thornton, 1992), student performance in training (Reynolds, 1966), and soldier performance in combat (Dugan, 1953; Williams & Leavitt, 1947). Moreover, studies comparing peer evaluations to other methods of assessment found that peer evaluations were more highly predictive of future performance than measures such as military course grades (O’Connor & Berkshire, 1958), instructor ratings (Kraut, 1975; Wherry & Fryer, 1949), and supervisor ratings (Williams & Leavitt, 1947).

**INTERPRETATIONS OF PEER RATINGS**

The reason for the predictive validity of peer ratings has not been explicitly defined. One possibility is that peers are very focused on each other’s performance to determine the quality of their own performance (Mumford, 1983). This focus on social comparisons may increase the accuracy of their assessments of others’ task performance skills.

Another hypothesis is that different groups of raters (supervisors, peers, self, and subordinates) have different opportunities to view an individual’s performance and, therefore, use different information in making evaluations (Borman, 1974; Knapp & Campbell, 1993; Murphy & Cleveland, 1991). Murphy and Cleveland suggested that the supervisor’s presence generally elicits an individual’s maximal performance, as opposed to typical performance. In contrast, peers are thought to have a greater number of opportunities to observe an individual’s day-to-day performance and to gain insights into attitudes and personality characteristics that affect such performance.

Previous research suggests that peer evaluations do not relate well to nonpersonality characteristics such as knowledge (Schumacher, Scogin, Howland, & McGee, 1992) or academic grades (Williams & Leavitt, 1947). On the other hand, peer evaluations seem to contain more information than supervisor evaluations about an individual’s personality and interpersonal skills (Borman, White, & Dorsey, 1995; Oppler, Peterson, & McCloy, 1994). Borman et al. found that peer ratings contained additional information about the ratee’s obnoxious characteristics that were not found in ratings made by supervisors. In addition, peers may be better able than supervisors to distinguish between effort and skill dimensions of performance (Klimoski & London, 1974). Overall, these findings suggest that peer evaluations of performance demonstrate high predictive validity because they include more information about the interpersonal and motivational dimensions of performance than do other assessments. If this is the case, peer evaluations might be a useful method to measure these hard-to-measure constructs. They would also provide a more appropriate criterion for interpersonal and motivational tests and measures.
APPLICATION TO THE MILITARY

Interpersonal skills and motivation are critical to performance in most organizations, and the military is no exception. This is especially true for soldiers who work in special operations units, such as the Army Special Forces. Special Forces soldiers work closely with 8 to 11 teammates in a small setting that demands high levels of team coordination and considerable autonomy from higher authorities. The Special Forces mission also requires that soldiers work closely with indigenous populations, providing them with training in tactical methods and technical information as well as conducting integrated training exercises. In addition, Special Forces soldiers frequently engage in joint exercises with members of other services and work closely with members of the State Department or other agencies. These performance requirements define the need for high levels of communication, interpersonal skills, and motivational attributes, such as initiative and persistence in the Special Forces job (Russell, Crafts, Tagliareni, McCloy, & Barkley, 1996).

Joining the Special Forces requires successfully completing a 3-week assessment center program, and a 3- to 6-month training program (training length differs based on the soldier’s assigned specialty area within the Special Forces). Both the assessment program and the training program have phases that require team interaction and coordination to complete field exercises. In the assessment program, the exercises involve transporting heavy objects such as telephone poles, and in the training program, the exercises involve demonstrating tactical knowledge and skills. Soldier performance during both programs is observed and rated by trained program staff as well as by peers.

HYPOTHESES

This research compared the predictive validity of peer and staff ratings of assessment center program performance for future performance in training and examined the relative emphasis given to motivational, interpersonal, and task aspects of performance by peers and staff when rating performance. Peer and staff ratings were collected both at the assessment center and in training. Information regarding whether the individual graduated from the training program was also measured. The peer and staff ratings that were made in the training program were independent of the peer and staff ratings made during the assessment center. That is, these ratings were not test–retest ratings made by a single set of peer and staff raters; rather, they were made by two different sets of raters. Peer ratings during assessment were expected to have higher predictive validity than staff ratings, and peers were expected to give greater importance to interpersonal skills and motivation.

During the assessment center program, peers have more opportunities than staff to see behaviors that indicate what individuals “will do” and to engage in interper-
sonal interactions. Although staff are expected to be strictly observational in the assessment program, once in training, the close supervision training environment allows staff more intensive time for observation and interaction with students. Even in training, however, interpersonal contact and opportunities to see behaviors indicating what students will do are somewhat more likely for peers than staff.

H1: Because peers had a better opportunity to observe soldiers, we predicted peer evaluations, rather than staff ratings, during the assessment center program would predict graduation from training more strongly (Kraut, 1975; Wherry & Fryer, 1949; Williams & Leavitt, 1947).

H2: The correlation between assessment center program ratings (Time 1) and training ratings (Time 2) would be greater for peers than for staff.

H3: We expected more agreement between peers and staff in ratings made during training than in those made during assessment.

H4: Because peers and staff were getting different amounts or types of information regarding motivation and interpersonal skills, we expected more agreement between peers and supervisors on task performance than on motivation or interpersonal skills.

H5: We predicted that peers’ ratings of overall current performance in training and expected future performance on the job would include significantly more information about motivation and interpersonal skills than would staff ratings.

METHOD

Participants

Participants included 329 male Army soldiers who started training for the Special Forces between March and October 1995. The majority were active duty soldiers (76%), with 24% in the National Guard. Thirty-four percent were officers, primarily captains, and 66% were enlisted soldiers (specialists, sergeants, and staff sergeants). The average age of the soldiers was 26 years old, with an average of 4.5 years of time in service.

Setting

Before starting training for the Special Forces, all soldiers attended and passed the 3-week assessment program, Special Forces Assessment and Selection (SFAS).¹

¹The descriptions provided here for the Special Forces programs are those that were in place at the time data was collected for this research.
The first phase of SFAS requires individual events such as psychological test batteries, runs, marches with rucksacks, navigating across terrain, and an obstacle course. The second phase requires 10 team-oriented events in which soldiers are placed in groups of 10 to 14 and must develop and execute plans for moving different types of heavy objects from one location to another, given limited time and resources. These events take place across 5 days, and each day a staff member observes and assesses the team during the events. At the conclusion of the 5 days, peers are asked to rank their team members with respect to their overall contribution to the team effort. During the assessment program, soldiers leave for three primary reasons: (a) voluntary withdrawal, (b) a medical problem or injury, or (c) removal for an integrity violation or for posing a danger to himself or others due to poor performance. At the conclusion of the program, soldiers may also be removed by the final selection board for psychological reasons or for failing to demonstrate a pattern of success across the events in the program.

After selection from the assessment program, soldiers attend the first phase of training. The time lag between assessment and training varies for each soldier, and may be as short as a month or could be longer than a year; typically it might be a few months. The first phase of training teaches land navigation and small unit tactics and spans about 5 weeks. Soldiers are assigned to squads of 10 to 14 members, who function as a team during training. Each team is assigned a Special Forces-qualified soldier, who trains and evaluates student performance. In this study, both trainer evaluations and peer evaluations were obtained at the conclusion of this first phase.²

**Measures**

**Assessment center program ratings.** SFAS staff observed and recorded soldier performance during the assessment program with respect to 13 dimensions of performance that included attributes such as physical fitness, teamwork, motivation, responsibility, judgment, and decisiveness. Staff members used a 3-point rating scale in which 1 was unsatisfactory, 2 was satisfactory, and 3 was outstanding. Because very few soldiers received outstanding ratings, we created an overall staff rating for each soldier by counting the total number of unsatisfactory ratings the soldier received from staff across all the events. Scores were then translated so that higher staff rating values would indicate higher performance. The distribution of this variable did show significant negative skew, $z = -3.29, p < .01$, a condition that would be likely to attenuate its relation with normally distributed variables. Previous factor analysis of these ratings suggested three primary factors—effort and de-

²The second two phases of Special Forces training include specialty training (e.g., medical, weapons, and communications) and a final collective training phase.
pendability, judgment, and physical fitness—that were relatively stable across four samples (Russell et al., 1995).

At the conclusion of the assessment program, peers were asked to rank the members of their team, excluding themselves, with respect to the soldier’s overall contribution to the team effort. This peer ranking was then divided by team size to equalize across teams of different sizes. These scores were also translated so that higher peer rankings would indicate higher performance.

**Training ratings.** We obtained seven ratings from peers and trainers at the conclusion of the first phase of training: (a) interpersonal performance, (b) motivation (described as effort and persistence), (c) physical fitness, (d) leadership, (e) military tactics, (f) overall current performance, and (g) predicted future on-the-job performance. Ratings of physical fitness, leadership, and military tactics were averaged to form an overall task performance rating. The 5-point rating scales provided the rater with behavioral indicators of performance to consider when making their decisions. Higher scores indicated higher performance, where 1 represented low performance, 3 represented average performance, and 5 represented high performance.

**Training outcome.** Final training outcome was obtained from records after students either completed or failed to complete the entire Special Forces training course. There were three reasons for failure: (a) voluntary withdrawal; (b) an honor code violation; or (c) a decision by a review board at the conclusion of any of the phases of training, based on the staff recommendation, that the soldier failed to meet the required standards. If a student fails to meet the required standards, he is typically given at least one opportunity to repeat or recycle the phase. This information was used to create a 3-point scale for a training outcome variable. One point was assigned for soldiers who failed to complete training (12% of the sample), 2 points were given to soldiers who completed training but required one or more recycles (27% of the sample), and 3 points were granted to soldiers who graduated from their first attempt at the course (61%). Because ratings from training staff and peers are the primary input for pass or fail decisions by the review board, and because training staff members provide the recommendations for passing or failing, we will not use training ratings to predict training outcome in these analyses.

---

3Although some research suggests this may overcorrect the scores (Willingham, 1959), the size of the groups in this research varied only slightly, with 91% ranging in size from 10 to 13 members, and the other 9% with either 9 or 14 members.

4Behavioral indicators were not provided for the overall current performance and predicted future on-the-job performance rating scales.
RESULTS

Predicting Success in Training

Correlational results showed that peer rankings from SFAS significantly predicted final training outcomes, \( r = .29, p < .01 \), but staff ratings from SFAS did not, \( r = .00, ns. \) As expected, peer assessments from SFAS predicted final training outcome more strongly than staff ratings from SFAS, \( t(317) = 2.85, p < .01 \), supporting the first hypothesis. In addition, peer assessments from SFAS showed stronger correlations with peer assessments from training than staff assessments from SFAS showed with staff assessments from training (see Table 1), supporting the second hypothesis. In fact, assessment center ratings from staff were not significantly correlated with staff ratings from training.

As predicted in Hypothesis 3, results indicated that the correlations between peer and staff ratings during training were higher than they were between peer and staff ratings made during assessment. The correlation between peer and staff assessments made in SFAS was .06, which was not significant. The correlations between peer and staff ratings from training, however, were all significant \( (p < .01) \) and high in magnitude: For motivation, \( r = .66 \); for interpersonal performance, \( r = .65 \); for task performance, \( r = .81 \); for overall current performance, \( r = .80 \); and for predicted future on-the-job performance, \( r = .77 \). A summary of these relations is shown in Figure 1.

Finally, as predicted in Hypothesis 4, the correlation between peer and staff training ratings was significantly higher for task performance, \( r = .81, p < .01 \), than for motivation, \( r = .66, p < .01 \), and interpersonal performance, \( r = .65, p < .01 \); \( z = 4.22, p < .01 \), and \( z = 4.45, p < .01 \), respectively. The nature of this higher agreement between peers and staff is related to the higher reliability evidenced by the task performance ratings. Interrater reliabilities, although only

<table>
<thead>
<tr>
<th></th>
<th>Peer Rating</th>
<th>Staff Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motivation</td>
<td>.38**</td>
<td>.02</td>
</tr>
<tr>
<td>Interpersonal performance</td>
<td>.40**</td>
<td>.09</td>
</tr>
<tr>
<td>Task performance</td>
<td>.43**</td>
<td>.08</td>
</tr>
<tr>
<td>Overall current performance</td>
<td>.41**</td>
<td>.10</td>
</tr>
<tr>
<td>Future on-the-job performance</td>
<td>.40**</td>
<td>.10</td>
</tr>
</tbody>
</table>

**\( p < .01 \).
available for the peer training ratings, showed a coefficient alpha of .95 for task performance, .81 for motivation, and .76 for interpersonal performance. Task performance was more easily observed, generating higher reliability or agreement within the rater group as well as a higher correlation or agreement between the rater groups.

Data Transformations

Because many of the rating distributions demonstrated significant skew and showed different levels of skew, it is possible that the differences in correlations are not based on true construct differences but rather represent artifacts of the distribution differences. To investigate this, ratings that were significantly skewed were transformed to normalize their distributions. Most distributions were able to achieve normality through a transformation (see Table 2), which was chosen because it provided the lowest $z$ score in testing the skewness. Using the transformed variables, none of the results in Hypotheses 1 through 4 changed in magnitude or significance; each hypothesis was still supported using the transformed variables.

Dimensions of Overall Performance

Intercorrelations among the five peer and five staff ratings from the training program were quite high, suggesting potential multicollinearity problems with regressing overall performance ratings for current and future performance on the component ratings (motivation, interpersonal performance, and task performance;
see Table 3). Green (1978) suggested multicollinearity might be a problem with intercorrelations in the range of .80 or .90.

Ratings of overall current performance and predicted future on-the-job performance by both staff and peers were each regressed simultaneously on ratings of task performance, interpersonal performance, and motivation. These three ratings (task performance, interpersonal performance, and motivation) are subcompon-

<table>
<thead>
<tr>
<th>Rating</th>
<th>Initial Skew Value</th>
<th>Transformation</th>
<th>Final Skew Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peer assessments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Forces Assessment and Selection</td>
<td>ns</td>
<td></td>
<td>ns</td>
</tr>
<tr>
<td>Training–motivation</td>
<td>-4.78**</td>
<td>Log 10</td>
<td>ns</td>
</tr>
<tr>
<td>Training–interpersonal</td>
<td>-3.74**</td>
<td>Log 10</td>
<td>ns</td>
</tr>
<tr>
<td>Training–task</td>
<td>-2.60**</td>
<td>Log 10</td>
<td>-1.71*</td>
</tr>
<tr>
<td>Training–overall current</td>
<td>-3.58**</td>
<td>Log 10</td>
<td>ns</td>
</tr>
<tr>
<td>Training–future on-the-job</td>
<td>-4.43**</td>
<td>Log 10</td>
<td>ns</td>
</tr>
<tr>
<td>Staff assessments</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Special Forces Assessment and Selection</td>
<td>-3.29**</td>
<td>Square root</td>
<td>ns</td>
</tr>
<tr>
<td>Training–motivation</td>
<td>-1.67*</td>
<td>Square root</td>
<td>-1.81*</td>
</tr>
<tr>
<td>Training–interpersonal</td>
<td>ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training–task</td>
<td>ns</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Training–overall current</td>
<td>-1.77*</td>
<td>Square root</td>
<td>-1.97*</td>
</tr>
<tr>
<td>Training–future on-the-job</td>
<td>-2.37**</td>
<td>Square root</td>
<td>ns</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.

<table>
<thead>
<tr>
<th>Peer Ratings</th>
<th>Staff Ratings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Motiv</td>
<td>Interp</td>
</tr>
<tr>
<td>Motiv</td>
<td>1.00</td>
</tr>
<tr>
<td>Interp</td>
<td>.82</td>
</tr>
<tr>
<td>Task</td>
<td>.85</td>
</tr>
<tr>
<td>Overall</td>
<td>.90</td>
</tr>
<tr>
<td>Future</td>
<td>.91</td>
</tr>
</tbody>
</table>

Note. All correlations are significant at p < .01. Motiv = motivation; Interp = interpersonal.
ents of the summary ratings (overall current performance and predicted future on-the-job performance). Simultaneous multiple regression was used to determine the extent to which each of these component ratings represented unique variance in the broader summary rating, with the goal of comparing the policies of peers and staff in making the overall performance decisions. Results showed that each of the three dimensions explained unique variance in peer and staff ratings of overall current performance and future on-the-job performance (see Table 4). That is, although there was a high level of intercorrelation among the ratings, each of the component ratings explained a significant amount of unique variance in the ratings of overall current performance and predicted future on-the-job performance. For overall current performance, peers included a significantly greater amount of information about motivation than did staff. For predicted future on-the-job performance, staff placed significantly more emphasis on task performance than did peers, and peers placed significantly more emphasis on interpersonal performance and motivation than did staff.

As with the analyses for the first four hypotheses, because many of the rating distributions demonstrated significant skew, it is possible that the differences in regression coefficients were not based on true construct differences but instead represent artifacts of the distribution differences. Using the transformed variables, results did change for the regression analyses using current performance as the dependent measure but did not change with future on-the-job performance as the dependent measure (see Table 5).

Using the transformed variables, there were no significant differences in the regression coefficients for peer and staff ratings of motivation, interpersonal performance, and task performance in predicting overall current performance ratings.

<table>
<thead>
<tr>
<th>TABLE 4</th>
<th>Comparison of Regression Coefficients for Peer and Staff Raters</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Peers</td>
</tr>
<tr>
<td></td>
<td>β</td>
</tr>
<tr>
<td>Overall current performance regressed on component ratings</td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>.67**</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>.13**</td>
</tr>
<tr>
<td>Motivation</td>
<td>.23**</td>
</tr>
<tr>
<td>Future on-the-job performance regressed on component ratings</td>
<td></td>
</tr>
<tr>
<td>Task</td>
<td>.49**</td>
</tr>
<tr>
<td>Interpersonal</td>
<td>.24**</td>
</tr>
<tr>
<td>Motivation</td>
<td>.30**</td>
</tr>
</tbody>
</table>

*p < .05. **p < .01.
When predicting future on-the-job performance, however, the regression coefficient for staff ratings of task performance was significantly greater than that for peer ratings of task performance, and the regression coefficients for peer motivation and peer interpersonal performance were significantly greater than those for staff.5

**DISCUSSION**

These analyses were designed to examine the premise that peers include more unique information than staff about interpersonal performance and motivation when rating an individual’s overall performance. When peers and staff were rating an individual’s expected future performance on the job, based on observations during training, this premise was supported. It was not supported, however, when peers and staff were rating an individual’s current performance in the training program. When rating expected future on-the-job performance and current performance in training, both peers and staff included significant unique information about task performance, interpersonal performance, and motivation. Both peers and staff placed comparable amounts of importance on the information from these three dimensions when rating overall performance in training; task performance showed the highest unique contribution, motivation the next

---

5 Although researchers may argue that multicollinearity necessitates that variables with intercorrelations in the range of .80 or .90 be combined (Green, 1978), these results emphasize that this decision depends on the purpose of the research. In this case, maintaining the multiple rating dimensions provided insight into the content of the ratings and the nature of the rating process.
highest contribution, and interpersonal performance the lowest unique contribution of the three.

When rating predicted future on-the-job performance, peers placed significantly more importance on interpersonal performance and motivation than did staff, and they placed significantly less importance on task performance. The magnitudes of the differences are particularly strong for task performance and motivation. This provides support for the theory that peers may emphasize an individual’s motivation—or what the individual actually will do on the job—and not just what he is maximally capable of doing.

Both peers and staff emphasized task performance more in rating overall current performance than predicted future on-the-job performance, which is appropriate given that performance during that phase of training focuses predominantly on tactical knowledge and skills. That may also be why peers did not show a greater emphasis on interpersonal skills and motivation in this particular setting. Future on-the-job performance for these soldiers will require a greater emphasis on interpersonal skills, given that they will be interacting with other cultures and members of other services. It is logical, therefore, that interpersonal performance will be more important in predicting future on-the-job performance than it is for predicting performance in this first tactical phase of training. In addition, although the training situation will typically elicit maximal performance from students, on-the-job performance will be much more affected by the individual’s personality and typical behavior, so the importance of the individual’s motivation would be expected to increase on the job. These results clearly indicated that the relative importance of task, interpersonal, and motivational dimensions of performance depend on the rating situation and the particular question that is posed to the rater.

Peer and staff ratings of expected future on-the-job performance contained different emphases on task, interpersonal, and motivational dimensions of performance; it was not possible to distinctly determine from this whether the peer evaluations are providing better predictive information than the staff. Determining this would require following these soldiers into the field and obtaining measures of their on-the-job performance; this would be a useful future study.

Taken together, however, these results do present evidence that peers include some critical information in their ratings that is not included in ratings made by staff. First, in the assessment center, we found that peer ratings significantly predicted later success in training, whereas staff ratings did not. Second, in the training course, when staff had longer and more opportunities to observe individuals’ performances than they had in the assessment center, their ratings demonstrated a higher correlation with peer ratings of performance. Third, there was more agreement among peers, as well as between peers and staff on assessments of task performance (a dimension that is more visible), than there was for assessments of the motivational and interpersonal dimensions. This is convergent with
Rothstein’s (1990) findings that the interrater reliability of job performance ratings increased with an increased opportunity to observe performance. Interpersonal performance and motivation may be key factors in the structure of peer assessments that account for their historical strength as assessment tools, especially in predicting future performance.

Measurement Concerns

One potential concern that might be raised with the results reported for the assessment center program is that the peer and staff assessments used different measurement techniques, rankings and ratings, respectively, and that this difference in measurement techniques was responsible for the differential prediction obtained from peers and staff. Similarly, the nonsignificant correlation between peer and staff assessments in the assessment center program could also be driven by differences in measurement, as opposed to differences in raters’ judgments. Previous research by Love (1981), however, indicates that the correlation between peer and supervisor assessments does not change significantly when rankings are substituted for ratings; that is, the magnitude of the relation between the peer and staff assessments is only slightly affected by changes in the measurement technique.

In addition, there are several compelling facts about this research that suggest that the peer assessments were at a disadvantage, and the conclusions drawn here are conservative with respect to Type I error. First, the format of the staff assessments in the assessment center program was closer in nature to the format of the training assessment criteria; both were ratings of performance across specific dimensions. From a measurement perspective, the peer assessments as a ranking of overall performance represent a greater departure in format from the training ratings, suggesting that the differences in measurement format would attenuate the peer ranking correlations to a greater degree than the staff rating correlations. Second, the distributions of the nontransformed staff ratings in the assessment center program were more similar to the distributions of the criteria ratings, which, again, would suggest greater attenuation for correlations with peer ranking. Third, the final training outcome criteria is based on the recommendations of staff, and as such, we might expect that staff ratings in the assessment center program would be a stronger predictor of this criteria than would peer ratings. That is, if anything, more bias exists that would favor a relation between staff ratings in the assessment center program and the final training outcome. Fourth, the assessment staff members are trained observers; they have completed a structured training program to learn how to provide accurate assessments, and they are experienced in making ratings. The quality of their assessments should be less prone to rater errors than those of the peers, and therefore, the staff correlations should be less attenuated by error. In sum, we feel that whatever biases were present in this study favored staff assess-
ments over peer assessments, which makes the findings herein with respect to peers even more impressive.

Future Research and Applications

Although the results presented here suggest that peers may possess better information about interpersonal performance and motivation, it cannot be explicitly determined from these data whether peers possess more accurate information—or whether peers simply believe that interpersonal performance and motivation are more important to future performance than do staff. A more definitive way to answer this question would be to compare the antecedents of peer and staff ratings. If peer ratings of interpersonal skills and motivation are more accurate, results should demonstrate higher path coefficients for relevant motivational and interpersonal antecedents. Research by Oppler et al. (1994) found higher validities using peer evaluations as criteria for personality measures and provides some initial evidence that peer evaluations may be more accurate. One challenge of this approach will be defining appropriate antecedent constructs and finding appropriate antecedent measures.

This research has important implications for applied Army settings. Performance in all organizational positions is maximized if soldiers maintain high levels of motivation under low levels of supervision. For jobs requiring high levels of autonomy, such as those in the Special Forces, this is particularly critical. Many military jobs also require high levels of interpersonal interaction to succeed, whether that entails working within one’s own squad or work team or interfacing with other units, companies, services, or international peacekeeping forces.

If peers are particularly adept at assessing interpersonal performance and motivation, this information would be useful for purposes of training and self-development or for selection into screened units such as special operations, recruitment, or drill instruction. The goal would be to use peer evaluations to obtain critical personality or interpersonal information that is sometimes difficult to get from other sources, such as supervisors, military records, or paper-and-pencil assessment measures.

Using peer evaluations in any setting requires a theoretically based approach that specifically identifies the purpose of obtaining the peer evaluations and delineates the critical attributes to be assessed. The first step is to identify the purpose of the assessments. The next step is to consider whether the peers are interactive enough to provide useful information. The peers of some individuals, for example, leaders such as company commanders, may not have extensive interactive contact with fellow company commanders. It is important, therefore, to use peer ratings to measure interpersonal and motivational performance only in those situations in which peers are in a position to observe this performance. The final step is to iden-
tify the specific dimensions of interest, particularly focusing on constructs that perhaps have not been measured well through other means.

REFERENCES


