

The research register for this journal is available at
http://www.mcbup.com/research_registers



The current issue and full text archive of this journal is available at
<http://www.emerald-library.com/ft>

Journal of
Management
Development
20,6

508

Applying the social relations model to self and peer evaluations

Gary J. Greguras

Louisiana State University, Baton Rouge, Louisiana, USA

Chet Robie

Niagara University, New York, USA, and

Marise Ph. Born

Vrije Universiteit Amsterdam, The Netherlands

Keywords *Management development, Rating, Performance evaluations*

Abstract *Peer evaluations of performance increasingly are being used to make organizational decisions and to provide individuals with performance related feedback. Using Kenny's social relations model (SRM), data from 14 teams of undergraduate students who completed performance ratings of themselves and other team members were analyzed. Results indicated a significant target variance effect for the majority of performance dimensions and a significant perceiver variance effect for all performance dimensions. Results further indicated that, in general, how individuals see themselves is not congruent with how others see them, how individuals see themselves is congruent with how they see others, how individuals are seen on a particular dimension is related to how they are seen on other performance dimensions, and, how a person is seen by others does not relate to how that individual sees others. Implications, limitations, and suggestions for future research using the SRM are discussed.*

In response to the changing nature of work (e.g. team-based work, telecommuting), organizations are employing different methods and procedures to measure, evaluate, and improve managerial job performance. One recent organizational intervention targeted at measuring and improving managerial performance has been the implementation of multi-source performance rating systems (e.g. 360-degree feedback systems) in which raters from different organizational levels (e.g. subordinates, peers) provide managers with performance-related feedback and evaluations (London and Smither, 1995). Multi-source rating systems assume that raters from different organizational levels provide different information that aids in the identification of employee performance strengths and weaknesses (Murphy and Cleveland, 1995). Although each rater source is assumed to provide useful information, several researchers have suggested that peers may be a uniquely valuable source of information (e.g. Murphy and Cleveland, 1995).

As Murphy and Cleveland (1995) discuss, there are three main reasons why peers may be the single best source of performance information. First, peers

The authors would like to thank David Marcus in helping them to run and interpret the results of the SOREMO program. Readers interested in obtaining a copy of the SOREMO program may access the program at <http://nw3.nai.net/~dakenny/kenny.htm>

generally work closely together. This proximity increases peers' opportunities to observe performance-related behaviors which presumably increases the accuracy of their ratings (Wherry and Bartlett, 1982). Second, peers probably observe a less biased sample of behavior than other rater sources. That is, employees may alter or censor their behaviors while around supervisors or customers in order to project a certain image, but may be more likely to display their typical levels of performance when around their peers. Third, because peers generally are more numerous than supervisors, aggregating their ratings increases the reliability of their ratings and removes some of the biases of individual raters. In addition to these potential advantages, the shift from individually based work to team-based work in US organizations (Reilly and McGourty, 1998) has increased the relevancy and importance of peer feedback and evaluations for managerial development.

Peer ratings, however, have several potential limitations. For example, Landy and Farr (1983) note that peers often work closely together and often develop relationships with one another. Because peers often become friends, they may be unwilling to provide accurate ratings and instead may rate their friends too leniently. Another concern with peer ratings is that peer raters will rate everyone similarly in order to not cause friction within the workgroup (Murphy and Cleveland, 1995). Both of these concerns are primarily a result of interpersonal relationships among peers. That is, interpersonal relationships may cause peers to be too lenient when rating their friends or to not discriminate among friends. Unfortunately, most of the peer rating research has ignored these relationship factors (Hennen and Barnes-Farrell, 1997). Relationship factors and interdependencies among peers likely contain interesting and useful information for more fully understanding peer ratings.

The purpose of the present paper is to introduce organizational researchers to the social relations model (SRM) (Kenny, 1994; Warner *et al.*, 1979) and to use the SRM to analyze peer ratings. The SRM is a methodology that assesses the interdependencies among peer ratings to provide information regarding how relationship factors among peers influence their ratings. This paper begins by providing a brief overview of the SRM. Next, one rater error (i.e. halo error) that is typically encountered in many areas of research using observer ratings is discussed within the framework of the SRM. Finally, the SRM is used to analyze peer ratings and the advantages of using the SRM over other statistical methods are highlighted. Results of these analyses are used to illustrate and discuss how the SRM can be used to answer substantive and methodological questions in peer-based performance evaluation research that are not possible with traditional methods.

Social relations model

When managers evaluate the performance of their peers, their ratings are likely to be dependent on one another's ratings. For example, in a team setting, Jack's evaluation of Mary likely is related to Mary's evaluation of Jack. As Hennen and Barnes-Farrell (1997) note, previous research on peer appraisals has

ignored this nonindependence of data. Ignoring these interdependencies, researchers often use traditional ANOVA designs to analyze peer ratings. However, the violation of the assumption of independence often distorts the results of ANOVA (Kenny and Judd, 1983). In addition to the possibility of the results being distorted, meaningful information about the interdependencies among the peers is lost. In contrast to the traditional ANOVA designs, one advantage of the SRM is that it does not require an assumption of independence and, in fact, the nonindependence among peers is of interest and is analyzed (Marcus and Kashy, 1995).

Developed out of the person-perception literature, the SRM provides both a theoretical basis and a statistical tool to answer questions regarding the impact of rater and ratee characteristics on peer evaluations. The SRM is a special case of generalizability theory (Cronbach *et al.*, 1972) with the basic model consisting of a two-way, random-effects analysis of variance with perceiver as one factor, target as another factor, and relationship as the interaction between the perceiver and the target effects. The SRM is designed to decompose perceptual variance on a construct into several different components:

- perceiver effects – a person’s tendency to rate all persons similarly;
- target effects – a person’s tendency to elicit similar ratings from all persons; and
- relationship effects – factors unique to specific dyadic relationships.

To estimate perceiver, target, and relationship effects using the SRM, multiple perceivers must rate multiple targets. This requirement generally is met by employing a round-robin data collection design in which each member of a group rates, and is rated by, every other member. The round-robin design accommodates the interdependencies inherent in such data, making the SRM distinct from more typical generalizability theory models in which multiple perceivers rate a given target but the given target does not in turn rate those perceivers (Cronbach *et al.*, 1972).

Mainstream generalizability theory and SRM also differ in their typical application. Mainstream generalizability theory is typically used to estimate the sources of variance in ratings and to use this information to project the reliability of those ratings under different measurement conditions (for an example, see Greguras and Robie, 1998). The SRM is also used to estimate the sources of variance in ratings but not necessarily with the intention of studying the reliability of those ratings. Instead, the focus of the SRM is on isolating those sources of variance and correlating them with each other and self-ratings to try to help explain social perceptual phenomena.

In order to illustrate the typical use of the SRM, consider the following example. John and Ann are team members. John perceives Ann as a cooperative co-worker and, therefore, evaluates her favorably on this dimension. Several possible explanations exist for John’s rating of Ann. First, John’s rating of Ann could be the result of John’s tendency to rate or perceive everyone on his team

as cooperative, as would be evidenced by a significant perceiver effect. When there is considerable perceiver variance (i.e. some raters consistently see others as cooperative and other raters see them as uncooperative), there is *assimilation* in ratings (Kenny, 1994). In a generalizability theory framework, assimilation would be evidenced as a main effect due to the rater. Second, John's ratings of Ann also may be attributed to Ann's tendency to evoke similar ratings from all team members, as would be evidenced by a significant target effect. That is, perhaps people on the team (not just John) see Ann as a very cooperative co-worker. When raters consistently see some team members as more cooperative than others (i.e. there is considerable target variance), there is *consensus* in ratings (Kenny, 1994). In a generalizability theory framework, consensus would be evidenced as a main effect due to the person (i.e. ratee). Thus, consensus is sometimes viewed as "true" variance if the ratee is considered the object of measurement. Third, John's appraisal of Ann may be due to factors that are uniquely dyadic; that is, effects due to the unique relationship between two co-workers. The relationship effects represent, for example, the degree to which John's evaluation of Ann cannot be explained by perceiver or target effects. If John does not see most co-workers as cooperative and most people do not consider Ann to be cooperative, yet John repeatedly sees Ann as cooperative, then there would be a large relationship effect. In essence, the relationship effect is an interaction effect as it occurs at the level of the dyad, unconfounded by individual characteristics (i.e. perceiver and target effects).

When multiple performance dimensions are being assessed, the relationships among the constructs can be assessed using either the perceiver or target effects for each performance dimension. Consider the example of the relationship between the performance dimensions of cooperativeness and reliability. If one were using the perceiver effects (i.e. a rater's tendency to evaluate others similarly) from these performance dimensions, a positive perceiver-perceiver correlation would indicate that raters who tend to rate their peers as cooperative also tend to rate them as reliable. If one were using the target effects (i.e. a ratee's tendency to evoke similar ratings from the raters) from these performance dimensions, a positive target-target correlation would indicate that raters who are seen as cooperative are also seen as reliable. In general, interest is usually focused more on the target-target correlation matrix because the target effects reflect variance attributable primarily to the ratee (i.e. the degree to which raters can agree on a ratee's level of a given construct) whereas perceiver effects reflect variance attributable primarily to the rater (i.e. the degree to which raters tend to rate similarly across ratees on a given construct).

Another advantage of the SRM relates to its analysis of self-ratings on the performance dimensions of interest. Typically, self-other agreement in performance evaluation research is assessed by correlating self-ratings with the mean of observer ratings; however, a mean of observer ratings contains both perceiver and target variance and, therefore, is a less precise estimate of self-other agreement than are self-other agreement indices calculated using the

SRM. Using the SRM, it is possible to investigate self-other agreement (e.g. are people who see themselves as cooperative seen as cooperative by others?) by correlating the target's self-report ratings with the target effects. Likewise, one may investigate assumed similarity (e.g. do people who see themselves as cooperative also see others as cooperative?) or contrast projection (e.g. do people who see themselves as cooperative assume others are not cooperative?) by correlating the perceiver effects with self-report ratings. Additionally, in the presence of individual differences on the construct (e.g. significant target effects), significant perceiver effects should be considered a rater response bias. It is then of interest to examine whether the raters' views of themselves are somehow related to how they are rating others. Understanding the processes leading to self- and observer-rating agreement could greatly influence theory development and field applications regarding multi-source ratings.

Finally, issues of reciprocity can be addressed with the SRM. The SRM distinguishes between generalized (i.e. individual level) and dyadic reciprocity (Kenny, 1994). Generalized reciprocity is the individual level correlation between a person's target effect and that person's perceiver effect. For example, do individuals who are seen as cooperative also see others as cooperative? Dyadic reciprocity is the dyadic-level correlation between relationship effects. For example, if David rates Bob as more cooperative than he rates any other individual, will Bob then rate David as more cooperative than he rates any other individual? Reciprocity is different from both self-other agreement, assumed similarity, and contrast project in that the examination of reciprocity requires the computation of correlations between components of variance of observer ratings (e.g. perceiver effects, target effects); whereas the examination of self-other agreement, assumed similarity, and contrast projection requires computation of correlations between self-ratings and components of variance of observer ratings. The study of reciprocity is a fundamental area of inquiry in social perception (Kenny, 1994). For example, in a team setting, a finding of generalized reciprocity for a given construct may suggest a two-way elicitation process wherein the behavior of the ratee is eliciting similar (or dissimilar in the case of a negative correlation) behaviors from the raters. A rater bias that has been studied extensively in the performance evaluation literature that the SRM can lend unique insight into is halo error.

Halo error

Halo error is defined as "a rater's failure to discriminate among conceptually distinct and potentially independent aspects of a ratee's behavior" (Saal *et al.*, 1980, p. 415). Halo error has been identified across disciplines as a common cognitive error that can occur in observer ratings (Berman and Kenny, 1976; Cooper, 1981). Researchers typically make a distinction between "true" and "illusory" halo. "True" halo is not a result of cognitive errors but results instead from true correlations among the performance dimensions (e.g. managers who perform favorably on all aspects of their jobs should receive favorable ratings on all performance dimensions). In contrast, "illusory" halo results from a

cognitive error in which the rater incorrectly generalizes from performance on one dimension to other dimensions rather than carefully considering and making distinctions on each separate aspect of performance.

The SRM provides a methodology to attempt to separate these different forms of halo. Specifically, as noted above, the target-target matrix of correlations among constructs can be used to identify the amount of “true” halo in a set of performance ratings. Recall that significant target variance indicates the amount of stable variance attributable to how consistently a group of raters rates an individual on a construct of interest. Therefore, the target-target correlation matrix should be a more accurate reflection of the true intercorrelation among dimensions than simply examining the correlation matrix between the mean of observer ratings which confounds “true” (i.e. target) variance with perceiver, relationship, and random error variance and thus confounds “true” with “illusory” halo.

Summary

The SRM provides a theoretical basis and statistical tool to decompose perceptual variance into a perceiver effect, target effect, and relationship effects. Few studies have applied the SRM to the study of group dynamics (see Marcus, 1998 for a review of those studies). In the present paper, a sample of students organized in group project teams is used to illustrate how the SRM can be applied to team member performance evaluations. The purpose of the current study is to analyze self and peer ratings using the SRM. The focus of the current study is on the size of the target and perceiver variance components (consensus and assimilation), their relations with self-ratings (self-other agreement and assumed similarity/contrast projection), and their relationships with each other (generalized reciprocity).

Method

Participants

Participants were 59 undergraduate students enrolled in two sections of introduction to industrial/organizational psychology (40 women and 19 men; 30 Caucasian and 29 non-Caucasian) at a large southern university. Participants formed 11 mixed-gender groups of four individuals each and three groups of five individuals each for a total of 14 groups. Although this sample size may appear small, Kenny (1994) suggests that a minimum of six groups is necessary to obtain stable estimates for groups of between four to eight members. Accordingly, the sample size in the current study should produce stable estimates.

Procedure and materials

Data were collected as part of a class project that was required for completion of the class requirements. Students were required to complete a group-based project that entailed reading a case study of an organization that was undergoing a myriad of problems and then creating both written and oral

reports that detailed the diagnosis of the problems and suggested interventions. Data were collected on the last day of class after written and oral reports had been completed. Students were asked to rate themselves and each of their group members on their performance pertaining to the group project.

Self and peer ratings were completed using a modification of the instrument described in Abson (1994). Abson's original instrument contained the performance dimensions of cooperation, ideas, effort, and reliability (each measured with one item), scored on the following scale:

- 0 (no work);
- 1 (not good);
- 2 (average);
- 3 (good); and
- 4 (excellent)

As part of a class exercise, the instrument was modified to include additional relevant performance dimensions that were described in behavioral terms. The modified instrument contained six performance dimensions (cooperation, ideas, effort, reliability, quality, and overall performance), scored on the following scale:

- 1 (unacceptable);
- 2 (very poor);
- 3 (below average);
- 4 (average);
- 5 (above average);
- 6 (very good); and
- 7 (excellent)

Cooperation was defined as a willingness to work together and evidence good communication skills. Ideas was defined as the quality and quantity of ideas. Effort was defined as the amount of time and effort put into the project and one's display of initiative. Reliability was defined as being accessible, punctual, following through on commitments, and regularly attending group meetings. Quality was defined as the quality of individual outcomes produced. Overall performance was defined as the overall contribution of the individual to the group project.

Students were informed that the overall performance dimension averaged across self and observer ratings accounted for 1/7th of the students' total grade in the class. All other performance dimensions were not be used for grading purposes. Confidentiality of the ratings was assured. However, given the fact that participants were informed that the ratings would be used in determining the grade in the course, we believe that the results of the study are probably best generalizable to peer performance appraisal settings in which

the ratings would be used for some form of administrative decision making versus the use of ratings for developmental purposes. Although the vast majority of multi-source feedback research has been conducted on developmental ratings, research suggests that multi-source feedback systems are frequently used for administrative decisions (London and Smither, 1995). Thus, the current study contributes to the literature not only by applying the SRM to peer ratings, but by investigating peer ratings made for administrative purposes.

Results

An overview of the social relations analysis is presented, followed by a description of each step of the analyses. For reasons described in the discussion section, we did not include multiple replications of the constructs; therefore, we were not able to separate relationship from error variance. Thus, our focus will be on the size of the target and perceiver variance components (consensus and assimilation), their relations with self-ratings (self-other agreement and assumed similarity/contrast projection), and their relationships with each other (generalized reciprocity).

Overview of the social relations analyses

The SRM analyses were conducted in four steps. First, the participant's ratings of each other (i.e. dyadic data) were decomposed into target variance, perceiver variance, and relationship/error variance. This first step allowed us to examine the degrees of consensus and assimilation in ratings of cooperation, ideas, effort, reliability, quality, and overall performance. This step also provided the components for the remaining steps. In the second step, the target and perceiver effects were correlated with the participants' self-ratings. This step allowed us to examine the degree of self-other agreement, assumed similarity, and contrast projection. The third step involved calculating the intercorrelations among the target effects for the six performance dimensions to examine the "true" correlation matrix. We consider the target effects intercorrelation matrix to be a matrix of "true" correlations because, as noted previously, the target effects reflect variance attributable primarily to the ratee (i.e. target) and not the rater (i.e. perceiver). The final step examined the correlations between the target effects and the perceiver effects for each of the six performance dimensions that allowed the individual-level of generalized reciprocity of the ratings to be assessed. All of the steps were performed using Kenny's (1998) FORTRAN program SOREMO, which performs social relations analyses on data collected using a round-robin design. The formulas used in this program may be found in Kenny (1994).

Variance partitioning

The total absolute variance and the relative variance partitioning for the six performance dimensions are provided in Table I. The relative variances indicate the percentage of variance of each rating that could be attributed to the

Table I.
Relative variance
partitioning for the
performance
dimensions

Dimension	Target	Perceiver	Total absolute variance
Cooperation	0.27* (0.60)	0.25* (0.59)	1.53
Ideas	0.27* (0.66)	0.34* (0.71)	1.81
Effort	0.19 (0.49)	0.28* (0.59)	1.82
Reliability	0.33* (0.67)	0.22* (0.57)	2.02
Quality	0.23 (0.60)	0.38* (0.72)	1.70
Overall	0.28* (0.67)	0.35* (0.72)	1.32

Note: * $p < 0.05$. The combined relationship and error variances were not tested for significance and thus have been omitted for purposes of clarity. Reliability estimates for the variance estimates are in parentheses

target and the percentage of each rating that could be attributed to the perceiver. Relative variances were reported for ease of interpretation but the significance tests are actually performed on the absolute variances. Absolute variance for each variance component can be computed by multiplying the relative variance by the total absolute variance. As mentioned above, the relationship variance component and concomitant significance testing for it was not reported because multiple replications for each performance dimension were not available and thus the relationship variance could not be separated from the random error variance (Ingraham and Wright, 1986).

The reliabilities of the target and perceiver effects are given in parentheses in Table I. An effect may be statistically significant but not highly reliable. Thus, the reliabilities provide the user with a sense of the degree of confidence with which one can meaningfully interpret target and perceiver effects for a given variable. In the SRM, reliability is calculated and interpreted as described below.

If person i rates person j in group k on trait x , then score x_{ijk} is assumed to be composed of four components: group, actor (perceiver or rater), partner (target or ratee), and relationship components. In equation terms, the model is:

$$x_{ijk} = M_k + \alpha_{ik} + \beta_{jk} + \gamma_{ijk} \quad (1)$$

where x_{ijk} is the rating by i of j in group k , M_k the group effect for group k , α_{ik} the actor (perceiver) effect for person i in group k , β_{jk} the partner (target) effect for person j in group k , and γ_{ijk} the relationship effect (Kenny *et al.*, 1986, pp. 3-4). A measure of reliability for the mean ratee effect adjusted for the rater effect ($\bar{x}_{\cdot jk'}$) is:

$$\frac{\sigma_{\beta}^2}{\sigma_{\beta}^2 + \frac{n-1}{n(n-2)}\sigma_{\gamma}^2 + \frac{1}{n(n-2)}\rho_2\sigma_{\gamma}^2} \quad (2)$$

where σ_{β}^2 is a random variable representing the partner (i.e. target) variance, σ_{γ}^2 is a random variable representing the relationship variance, ρ_2 is the dyadic reciprocity correlation (i.e. the correlation between γ_{ij} and γ_{ji}), and n is group size (Kenny *et al.*, 1986, pp. 13-14). Thus, the above measure of reliability is not unlike other measures of reliability that position a “true” variance component in the numerator and the total variance component (composed of “true” and “error” variance) in the denominator. As a result, this measure of reliability can be interpreted in a similar manner to other measures of its kind. However, given that the SRM captures more variance in general than other models (which will almost certainly increase the denominator but not necessarily the numerator in comparison to traditional models), expectations of levels of reliability should be concomitantly adjusted in comparison to necessary levels of reliability for data analyzed with traditional models.

Results indicated a significant target variance (i.e. consensus) for four of the six performance dimensions (see Table I). The greatest degree of consensus was on the reliability performance dimension, indicating that 33 percent of the variance in ratings was accounted for by the target. That is, there was considerable agreement among peers when evaluating the reliability of their group members. Further, the target variance was not significantly different from zero for the effort and quality performance dimensions, indicating that raters did not agree on the performance levels of their peers on these dimensions. Results further revealed significant perceiver variance effects (i.e. assimilation) for all of the six performance dimensions, suggesting that some raters consistently saw their peers as higher on the performance dimensions whereas other raters saw their peers as lower on the performance dimensions (i.e. a statistically significant proportion of the variance in ratings is due to how the raters perceive others in general).

Self- and other-perception

Table II presents the correlations between the self-ratings and the target effects. The correlations of the self-ratings with the target effects for effort and quality are not reported, because these effects failed to reach statistical significance and it is only appropriate to analyze self-other agreement when consensus exists (Marcus and Leatherwood, 1998). Underlined correlations represent self-other agreement correlations (i.e. self-other correlation on the same performance dimension). In general, the correlations between self-report ratings and the target effects were low with only two correlations reaching statistical significance. The correlation between the self-report rating of cooperation and the target effect for cooperation was statistically significant ($r = 0.39, p < 0.05$), suggesting that those who consider themselves cooperative

Self-report	Cooperation	Target effect		Overall
		Ideas	Reliability	
Cooperation	<u>0.39*</u>	0.28	0.22	0.21
Ideas	<u>0.20</u>	<u>0.22</u>	0.09	0.04
Effort	0.13	<u>0.20</u>	-0.02	0.02
Reliability	0.38*	0.25	<u>0.32</u>	0.20
Quality	0.07	0.12	<u>-0.24</u>	-0.11
Overall	0.23	0.26	-0.10	<u>0.07</u>

Notes:

These correlations have been disattenuated based on the reliability of the variance estimates. Significance tests were based on the uncorrected correlations. Correlations with the target effects for the effort and quality performance dimensions are not presented because ratings for these dimensions did not evidence statistically significant target variance. Underlined values indicate self-other agreement

* $p < 0.05$

Table II.

Self (row) by target (column) correlations

were also seen as cooperative by their peers. The correlation between the self-report rating of reliability and the target effect for cooperation was also statistically significant ($r = 0.38, p < 0.05$), suggesting that those who consider themselves reliable were seen as cooperative by their peers. The paucity of significant correlations suggests that how individuals see themselves is not congruent with how others see them.

Assumed similarity and contrast projection

A positive correlation between a self-rating and a perceiver effect suggests that individuals see others as they see themselves for that dimension (i.e. assumed similarity). A negative correlation between a self-rating and a perceiver effect suggests that individuals see others as opposite of themselves for that dimension (i.e. contrast projection). Table III presents the correlations between self-ratings and perceiver effects. All of the correlations were positive (indicating assumed similarity) and most of the correlations were statistically

Self-report	Cooperation	Ideas	Perceiver effect			Overall
			Effort	Reliability	Quality	
Cooperation	0.95**	0.53**	0.64**	0.56**	0.45**	0.61**
Ideas	0.75**	0.73**	0.70**	0.38	0.54**	0.60**
Effort	0.56**	0.64**	0.68**	0.34	0.62**	0.68**
Reliability	0.66**	0.49**	0.48*	0.57**	0.48**	0.50**
Quality	0.64**	0.69**	0.78**	0.38	0.71**	0.65**
Overall	0.75**	0.72**	0.76**	0.41*	0.65**	0.83**

Notes:

These correlations have been disattenuated based on the reliability of the variance estimates. Significance tests were based on the uncorrected correlations

* $p < 0.05$; ** $p < 0.01$

Table III.

Self (row) by perceiver (column) correlations

significant (the exceptions being for three of the self-ratings correlations with the reliability perceiver effect). These findings suggest that how individuals see themselves is strongly related (in a positive direction) to how they see others (i.e. a strong degree of assumed similarity).

Target-target correlations

The target-target correlations are reported in Table IV. The correlations for the effort and quality target effects were not reported because the effects for these dimensions failed to reach statistical significance. All of the correlations were statistically significant and very high (average $r = 0.89$). These high positive correlations suggest that those who are seen as high (or low) on one dimension will also be seen as high (or low) on other dimensions (i.e. the “true” halo is very high). This is consistent with research that suggests that a general performance factor, not necessarily the result of halo, exists in many ratings of performance (see also Viswesvaran, 1993).

Generalized reciprocity

The target-perceiver (or generalized reciprocity) correlations for the dimensions with statistically significant target and perceiver effects all failed to be statistically significant. Thus, there appears to be no relationship between how one is seen and how one sees others for any of these performance dimensions. Dyadic reciprocity correlations are not reported because of the inability of the present study’s design to separate relationship variance from random error variance.

Discussion

This paper used the SRM to examine self and peer evaluations in a team setting. Peer ratings are frequently used in multi-source feedback systems used for managerial development and decision making (London and Smither, 1995). Unlike the majority of research investigating peer feedback, the current study described and used the SRM to analyze peer ratings. Participants’ responses were partitioned into target and perceiver components and were then correlated

Target effect	Cooperation	Ideas	Reliability	Overall
Cooperation	1.00			
Ideas	0.89*	1.00		
Reliability	0.93*	0.74*	1.00	
Overall	0.98*	0.87*	0.92*	1.00

Notes:

These correlations have been disattenuated based on the reliability of the variance estimates. Significance tests were based on the uncorrected correlations. Target-target correlations for the effort and quality performance dimensions are not presented because ratings for these dimensions did not evidence statistically significant target variance

* $p < 0.05$

Table IV.
Target-target
correlations

with self-ratings and each other. The target effects for the different performance dimensions were also correlated with each other. The results provided insight into the perceptual processes that individuals may have used in making their ratings of themselves and their peers.

The target effects for the effort and quality performance dimensions were not statistically significant. This may suggest that these performance dimensions are either defined in too vague a manner, that behaviors of this type are not highly observable and thus are prone to perceptual biases, that individuals simply disagree on the definition of the construct, or that raters disagree on the level of performance of their peers on these performance dimensions. Based on this information, perhaps an attempt at solidifying the definition of these two dimensions could be made or raters could be trained to more accurately rate such behaviors. Nonetheless, these results indicate that these performance dimensions, with this sample, did not contain significant target (i.e. true) variance and therefore the information provided by, and the decisions based on, these performance dimensions are questionable. This information is useful to users of this instrument and to the instrument's developers. In contrast, the reliability dimension had the highest percentage of target variance, the least percentage of perceiver variance, and the highest amount of total absolute variance. It is likely that these results are partially a function of the high degree of behavioral observability of this dimension. Partitioning variance into target and perceiver effects may help practitioners identify performance dimensions that lack target variance which, in turn, could help to identify rater training and test development needs.

As might be expected, the overall performance dimension evidenced the least amount of total absolute variance in comparison to the other performance dimensions. Participants were aware that their grade would be partially based on the ratings on this dimension. It is likely that this factor resulted in a restriction of range in the ratings for this dimension – a finding consistent with research that indicates less variability in ratings made for administrative versus feedback decisions (e.g. Zedeck and Cascio, 1982). It is surprising, however, that even with this potential pressure to distort the ratings for this dimension, both the target and perceiver effects were statistically significant. One might expect only a significant perceiver effect or no significant effects at all in this instance.

The correlations of the self-ratings with the target (i.e. “true”) effects give some insight into the accuracy of the self and observer ratings. If one defines accuracy in self-assessment as the relationship between self-ratings and the “true” component of observer ratings, then the relationships between self-ratings and target effects should be statistically significant to infer accuracy. Very few self-target effect correlations were statistically significant, suggesting that, for the present sample, how one sees oneself on a given performance dimension does not correspond to how one is seen by others on that same performance dimension. Likewise, only one self-other agreement correlation (i.e. cooperation) was statistically significant. These findings again indicate

that, in general, ratings of oneself are not related to other individuals' ratings of oneself on a particular dimension. This finding is consistent with previous research that has found low correlations between self and peer ratings (e.g. $r = 0.19$) (Conway and Huffcutt, 1997). Note, however, that low agreement between rater sources perhaps should be expected and may even be desirable (e.g. Tornow, 1993).

Results also indicated that the overwhelming majority of the self-perceiver effects correlations were statistically significant. These results suggest that how one sees oneself on a given performance dimension is related to how one sees others (e.g. if you see yourself as very reliable, you are likely to see others as highly reliable). Thus, in the present sample, individuals seem to be applying their schemas of their own performance to their ratings of others' performances. The information obtained from the correlations between self-ratings and target effect, perceiver effect, and other ratings provides a lot more information for ratees to consider and reflect on in terms of how they see themselves, others, and how others perceive them than does the typically reported self-other correlation. Given that multi-source feedback systems are assumed to improve managerial performance primarily as a result of increased ratees' insights about their performance (London and Smither, 1995), this additional information provided by the SRM analyses may add incremental value to multi-source ratings systems. Additionally, this information could be used as a diagnostic assessment for the need to train individuals on accurate self-ratings, as information to identify dimensions (e.g. cooperativeness) in which self-ratings and observer ratings converge, or as a point of discussion between peers.

The high target-target correlations suggest that the true intercorrelations or "true" halo among performance dimensions is very high. In other words, individuals who perform well on one dimension are likely to perform well on other dimensions. It should be noted that these correlations are not necessarily unrealistically high in that they have been corrected for unreliability. The pattern of correlations make sense in that individuals who are seen as unreliable probably are not seen as very cooperative ($r = 0.93$); whereas the link between reliability and ideas may not be as strong because individuals may contribute some very high quality ideas without having to attend every session ($r = 0.74$).

The lack of significant generalized reciprocity correlations suggests a general lack of interdependency in the performance ratings. For example, in the present sample, it is unlikely that individuals who were seen as more cooperative also tended to see others as more cooperative.

Implications for practice

There are several practical applications based on the current study's findings. First, this study illustrated the advantages of the SRM over more commonly used analyses. Organizations that use teams could provide both ratees and raters with more information regarding their performance, their ratings, how

they are viewed by other team members, and the relations between self and other-ratings by analyzing such ratings using the SRM. Because a goal of 360-degree feedback is increased ratee self-awareness (Tornow, 1993), this additional information provided by the SRM analyses could facilitate such increases in self-awareness. However, more work must be done on creating person-level SRM indices before this becomes a tenable reality. For example, the SRM should be expanded to include: a target effect (or “true” score) that could be provided to each ratee, a perceiver effect that could be provided to each rater (or “bias” score), and relationship effects that could be provided to each rater-ratee dyad. The SRM was developed out of a need to test theories regarding person-perception and thus needs some modification to make it maximally useful in an applied setting and to distinguish it from typical multi-rater technology (e.g. 360-degree feedback systems).

Second, the SRM may be used to identify performance dimensions that do not have significant target variance components. Performance dimensions without significant target variance components should probably not be used to make administrative decisions.

Third, identifying performance dimensions that lack significant target variance components could highlight rater training or test development needs.

Future research

There are several avenues for future research in this area. First, a study should be conducted in an organizational team setting to more conclusively investigate generalized and dyadic reciprocity and to strengthen arguments for generalization to field settings. Saavedra and Kwun (1993), using non-SRM methods, found that outstanding contributors tended to be the most discriminating raters in a peer evaluation context. A field SRM study much like the one in the present study would more appropriately answer the questions posed by Saavedra and Kwun (1993) by examining reciprocity and not just discrimination. For example, do those individuals who are seen as outstanding contributors see others, in general, as outstanding or below average?

Second, a taxonomy of peer evaluation constructs and instruments should be developed that emphasizes ratee characteristics (i.e. target variance) and de-emphasizes rater characteristics (i.e. perceiver variance) so that peer evaluations can be made maximally reliable, valid, and fair.

Third, contextual characteristics (e.g. group size, support for the appraisal system, demographic similarity, roles) should be identified that predict perceiver, target, and relationship variance components to provide a better understanding of the factors that influence self and peer ratings.

Fourth, investigation of how the perceiver, target, and relationship variance components may be used to predict various individual (e.g. perceptions of appraisal fairness), group (e.g. team performance), and organizational (e.g. profitability) outcomes also should be pursued and are easily incorporated into the SOREMO program.

Fifth, the incremental utility of providing information regarding self-other rating agreement, assimilation, and contrast projection over simply providing self-other rating correlations in improving rater insights and performance should be investigated.

Sixth, the study of meta-perceptions (i.e. our perceptions of how other people perceive us) in relation to performance evaluation constructs in team settings may be helpful in diagnosing and remediating conflicts within teams (see Kenny, 1994 for a full review).

Seventh, future studies should include either multiple indicators for each performance dimension or multiple replications of the constructs so that the relationship variance component could be estimated separately from the random error variance component.

Eighth, future research could explore other individual (e.g. personality), group (e.g. type of project team, cohesiveness), and organizational (e.g. politics) factors that might influence self- and peer ratings.

Limitations of the present study

The current study has several limitations. First, single-item indicators of the performance constructs were used; multiple-item performance dimensions would probably be more reliable.

Second, we did not conduct multiple replications of the constructs and thus could not separate relationship variance from random error variance. In so doing, we limited ourselves to exploring only certain parts of the model. The problem in employing multiple replications with performance evaluation data is that real changes in performance tend to occur over time and people differ in their rate of change (Hofmann *et al.*, 1993). This may result in a confounding of substantive, systematic variance with random error variance. An alternative that would enable one to estimate relationship variance separately from random error variance is to use the items in a multi-item composite as replications.

Third, our sample consisted of students and therefore the generalizability of our results may be questioned. However, like many project teams in an organization, participants interacted over the course of several months on this project.

Fourth, ratings in the current study were made for administrative purposes and therefore the generalizability of the current study's findings to situations where ratings are made only for developmental purposes is questionable.

Conclusion

The current study illustrated the utility of the SRM in the study of self and peer ratings. Fully understanding the factors that influence self and peer ratings is important due to the increasing emphasis on ratings from these rater sources as part of an individual's development program. In addition, peer evaluations are becoming more important as organizations increasingly organize work around teams. A model such as the SRM is needed to capture and explore such

interdependencies in peer performance evaluations and any other perceptual or behavior realm that affects group and organizational functioning. Future research in the organizational sciences increasingly will call on methodologies that can model such complexity.

References

- Abson, D. (1994), "The effects of peer evaluation on the behaviour of undergraduate students working in tutorless groups", in Foot, H.C., Howe, C.J. and Warden, D.A. (Eds), *Group and Interactive Learning*, Computational Mechanics, Southampton, pp. 153-8.
- Berman, J.S. and Kenny, D.A. (1976), "Correlational bias in observer ratings", *Journal of Personality and Social Psychology*, Vol. 34, pp. 263-73.
- Conway, J.M. and Huffcutt, A.I. (1997), "Psychometric properties of multisource performance ratings: a meta-analysis of subordinate, supervisor, peer, and self-ratings", *Human Performance*, Vol. 10, pp. 331-60.
- Cooper, W.H. (1981), "Ubiquitous halo", *Psychological Bulletin*, Vol. 90, pp. 218-44.
- Cronbach, L.J., Gleser, G.C., Rajaratnam, N. and Nanda, H. (1972), *The Dependability of Behavioral Measurements*, Wiley, New York, NY.
- Greguras, G.J. and Robie, C. (1998), "A new look at within-source interrater reliability of 360-degree feedback ratings", *Journal of Applied Psychology*, Vol. 83, pp. 960-8.
- Hennen, M.E. and Barnes-Farrell, J.L. (1997), "Appraisal in self-managing work groups: a social relations model", presented at the 12th Annual Conference of the Society for Industrial and Organizational Psychology, St Louis, MO, April.
- Hofmann, D.A., Jacobs, R. and Baratta, J.E. (1993), "Dynamic criteria and the measurement of change", *Journal of Applied Psychology*, Vol. 78, pp. 194-204.
- Ingraham, L.J. and Wright, T.L. (1986), "A cautionary note on the interpretation of relationship effects in the social relations model", *Social Psychology Quarterly*, Vol. 49, pp. 93-7.
- Kenny, D.A. (1994), *Interpersonal Perception: A Social Relations Analysis*, Guilford Press, New York, NY.
- Kenny, D.A. (1998), "SOREMO Version V.2", unpublished manuscript, University of Connecticut, Storrs, CT.
- Kenny, D.A. and Judd, C.M. (1983), "Consequences of violating the independence assumption in analysis of variance", *Psychological Bulletin*, Vol. 99, pp. 422-31.
- Kenny, D.A., Lord, R. and Garg, S. (1986), "A social relations analysis of peer ratings", unpublished paper, Storrs, University of Connecticut, Storrs, CT.
- Landy, F.J. and Farr, J. (1983), *The Measurement of Work Performance: Methods, Theory, and Applications*, Academic Press, New York, NY.
- London, M. and Smither, J.W. (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*, Vol. 48, pp. 803-39.
- Marcus, D.K. (1998), "Studying group dynamics with the social relations model", *Group Dynamics: Theory, Research, and Practice*, Vol. 2, pp. 230-40.
- Marcus, D.K. and Kashy, D.A. (1995), "The social relations model: a tool for group psychotherapy research", *Journal of Counseling Psychology*, Vol. 42, pp. 383-9.
- Marcus, D.K. and Leatherwood, J.C. (1998), "The interpersonal circle at zero acquaintance: a social relations analysis", *Journal of Research in Personality*, Vol. 32, pp. 297-313.
- Murphy, K.R. and Cleveland, J.N. (1995), *Understanding Performance Appraisal: Social, Organizational, and Goal-based Perspectives*, Sage, Thousand Oaks, CA.

- Reilly, R.R. and McGourty, J. (1998), "Performance appraisal in team settings", in Smither, J.W. (Ed.), *Performance Appraisal: State of the Art in Practice*, Jossey-Bass, San Francisco, CA, pp. 244-77.
- Saal, F.E., Downey, R.G. and Lahey, M.A. (1980), "Rating the ratings: assessing the psychometric quality of rating data", *Psychological Bulletin*, Vol. 88, pp. 413-28.
- Saavedra, R. and Kwun, S.K. (1993), "Peer evaluation in self-managing work groups", *Journal of Applied Psychology*, Vol. 78, pp. 450-62.
- Tornow, W.W. (1993), "Perceptions or reality: is multiple-perspective measurement a means or an end?", *Human Resource Management*, Vol. 32, pp. 221-30.
- Viswesvaran, C. (1993), "Modeling job performance: is there a general factor?", unpublished doctoral dissertation, University of Iowa, Iowa City, IA.
- Warner, R.M., Kenny, D.A. and Soto, M. (1979), "A new round robin analysis of variance for social interaction data", *Journal of Personality and Social Psychology*, Vol. 37, pp. 1742-57.
- Wherry, R.J. Sr and Bartlett, C.J. (1982), "The control of bias in ratings: a theory of rating", *Personnel Psychology*, Vol. 35, pp. 521-51.
- Zedeck, S. and Cascio, W.F. (1982), "Performance appraisal decisions as a function of rater training and purpose of appraisal", *Journal of Applied Psychology*, Vol. 67, pp. 752-8.