

The relationship between rater affect and three sources of 360-degree feedback ratings.

Author/s: David Antonioni

David Antonioni [a][*]

Heejoon Park [b]

Abstract

We investigated whether rater affect has a similar effect on the leniency of ratings from three of the sources of 360-degree feedback (downward, upward, and peer) and whether there is an interaction between a rater's affect and the time he or she has spent observing the ratee. The findings indicate that the influence of rater affect on the leniency of ratings was significantly greater in upward and peer feedback than in downward feedback and that the influence increased as raters' observation time increased. [C] 2001 Elsevier Science Inc. All rights reserved.

1. Introduction

Many organizations currently use 360-degree feedback to provide information on work performance to individual employees (Antonioni, 1996; Edwards & Ewen, 1996; Romano, 1994; Tomow & London, 1998). In the 360-degree feedback process, individuals receive ratings from three or four different sources: they assess themselves, and they receive assessments from their supervisors, from their peers, and, if they are managers, from their subordinates (hereafter referred to as direct reports). According to a recent estimate, approximately 40% of organizations use some form of 360-degree feedback (Bracken, Timmreck, & Church, 2000).

Organizations primarily use 360-degree feedback for developmental purposes, to provide information to ratees about how raters perceive their leadership and work behaviors. Researchers have suggested that the advantages of using multiple raters include the ability to observe and rate various job facets of each ratee's performance (Borman, 1974), greater reliability, enhanced fairness, and increased ratee acceptance (Latham & Wexley, 1982). Previous empirical research has addressed the benefits of 360-degree feedback (London & Beatty, 1993; Tornow, 1993), the benefits of peer and upward appraisals (Bettenhausen & Fedor, 1997), and the extent of self-other agreement in ratings (Atwater, Ostroff, Yammarino & Fleenor, 1998; Atwater, Roush & Fischthal, 1995; Atwater & Yammarino, 1992, 1993).

However, some researchers (e.g., London & Smither, 1995) have argued that research on 360-degree feedback has not kept pace with the practice and that there are insufficient research models and data available to guide organizations through the use of this type of feedback. Thus, there is a risk that the 360-degree feedback process, when implemented, will go amok (Waldman, Atwater & Antonioni, 1998). One aspect of 360-degree feedback that researchers have largely overlooked is the possibility that specific rater characteristics may influence ratings. This oversight is cause for some concern because 360-degree feedback programs depend on the quality of ratings from multiple sources (Antonioni & Woehr, 2000).

Previous studies have demonstrated that performance appraisals are influenced by

various rater factors such as rater's demographic characteristics, rater cognitive variables, and interpersonal affect (Decotiis & Petit, 1978; DeNisi, Cafferty & Meglino, 1984; Dipboye, 1985; Feldman, 1981; Ilgen & Feldman, 1983; Landy & Farr, 1980; Lee & Alvares, 1977; Park, Sims & Motowidlo, 1986). However, most studies on this topic have focused on performance appraisals from a single source, primarily traditional downward appraisals. Few have compared multiple sources of performance appraisals. Therefore, it remains to be seen whether performance ratings from supervisors, peers, and direct reports are influenced by the same factors and in similar ways.

This study focuses on the influence of one factor, interpersonal affect. Specifically, it investigates the relationship between a rater's interpersonal affect toward ratees and the leniency of ratings from 360-degree feedback. A rater's interpersonal affect toward a ratee is defined as liking (Murphy & Cleveland, 1991). Liking is an emotional reaction (positive, neutral, or negative) to a specific person (Zajonc, 1980). Even though rater affect is irrelevant to evaluating an individual's performance of given tasks, the fact that a rater likes or dislikes a ratee can influence the ratings and consequently determine that rater's evaluation of performance (Cardy & Dobbins, 1986; Decotiis & Petit, 1978; Dipboye, 1985; Park, Sims & Motowidlo, 1986).

Leniency and severity are terms used to describe a rater's tendency to give higher or lower ratings than is warranted by the ratee's behavior (Holzback, 1978; Saal, Downey & Lahey, 1980). Such rating errors can lead ratees to believe that their performance is better or worse than their actual performance (Waldman & Thornton III, 1988). Furthermore, when various organizational outcomes, such as merit pay and promotion, are linked to performance ratings, lenient or severe ratings could result in perceptions of unfairness, depletion of the merit budget, and legal difficulties (Kane, Bernardin, Villanova & Peyrefitte, 1995).

We expected that rater affect would influence ratings regardless of whether the rater was a supervisor, a peer, or a direct report of the ratee. Raters who like ratees may give higher ratings than raters who dislike ratees and this may contribute to lenient or severe ratings. However, there is reason to believe that rater affect would have a stronger influence on peer and upward feedback ratings than on downward feedback ratings because peers and direct reports have less experience in evaluating others and less accountability as raters than managers do. This study also examined whether the amount of time a rater has spent observing the ratee (a rater's observation time) moderates the influences of rater affect on evaluations of work behaviors. Increased observation time gives observers more information for accurate ratings, but it may also amplify the influence of rater affect because raters look for "data" that confirms their feelings about ratees.

2. Theory and hypotheses

Performance appraisal literature has recognized that interpersonal affect may play a significant role in performance appraisals (Decotiis & Petit, 1978; Dipboye, 1985; Park, Sims & Motowidlo, 1986). Some studies focusing on downward appraisals have found that rater affect is an independent variable influencing performance ratings (Cardy & Dobbins, 1986; Judge & Ferris, 1993; Robbins & DeNisi, 1994; Wayne & Ferris, 1990). Dipboye (1985) has suggested that a rater's interpersonal affect may influence performance evaluations independent of objective information and apart from the cognitive processes associated with performance evaluations. Raters may simply give higher ratings to the ratees they like and lower ratings to those they dislike. This may occur if a rater is less concerned about producing accurate ratings than about increasing his or her attractiveness in the eye of the ratee (Kumar & Beyerlein, 1991) or preserving a prior relationship with the ratee (Kingstrom & Mainstone, 1985). Such an impression management strategy is widely recognized as ingratiation in a political perspective (Ferris & Judge, 1991; Jones, 1964). Longenecker, Sims, and Gioia (1987)

argued that raters often deliberately manipulate performance appraisals for political purposes.

Interpersonal affect may also influence performance evaluations by influencing cognitive processes (Robbins & DeNisi, 1994; Varma, DeNisi & Peters, 1996). Thus far, researchers have developed several explanations of the ways a rater's liking of a ratee can influence the cognitive processes underlying performance evaluations. Liking may dictate what raters observe (Isen, Shaker, Clark & Karp, 1978); it may affect the attributions raters make about ratees' behaviors (Feldman, 1981); and it may influence the information raters retrieve from memory at the time of the evaluation (DeNisi, Cafferty & Meglino, 1984; Srull & Wyer, Jr. 1980).

Based on the foregoing discussion, we posited that rater affect toward a ratee may influence 360-degree feedback ratings regardless of the source of ratings (downward, upward, or peer). In a related study of the relationship between rater affect and rating errors, such as leniency, for downward, upward, and peer appraisals Tsui and Barry (1986) found a relationship between rater affect and ratings. The results of their study indicate that positive affect is related to higher ratings (high leniency) and negative affect is related to lower ratings (high severity) in all three types of performance evaluation.

However, their findings should be interpreted with caution because of the possibility of nonindependence in their data. In their study, more than one rating was obtained for each ratee. This practice is likely to introduce dependence among ratings. Glass, Peckham, and Sanders (1972) have pointed out that a violation of the assumption of independence is far more serious than the violation of other statistical assumptions. In a regression analysis, a violation such as this one may result in a bias in the standard error; thus, the significance test would be misleading (Kennedy, 1998). Furthermore, in an analysis of variance, the expected mean squares and their F ratios would also be biased if calculated with nonindependent observations (Kenny & Judd, 1986). In the present study we build on Tsui and Barry's findings by explicitly addressing the nonindependence issue. Therefore, the following hypothesis was tested to determine the relationship between rater affect and 360-degree ratings:

Hypothesis 1: There will be a positive relationship between the leniency of ratings of ratees' work behaviors and rater affect towards ratees in downward, upward, and peer feedback.

We felt it was important to establish whether rater affect had different levels of influence on downward, upward, and peer ratings. There are reasons to believe that some factors associated with particular types of feedback may increase the influence of rater affect. Previous studies have suggested that the impact of rater affect on ratings is stronger when a rater faces greater ambiguity in the appraisal process. For example, the influence of rater affect was stronger in trait-like ratings than in task-related ratings in studies conducted by Park, Sims, and Motowidlo (1986) and Varma et al. (1996). Similarly, Alexander and Wilkins (1982) found that the influence of rater affect was greater in subjective performance measures than in objective measures. Varma et al. (1996) interpreted these findings by explaining that rater affect may serve as an important interpretation cue when a rater faces ambiguity. In addition, Ferris and Judge (1991) suggested that the use of political influence tactics, such as ingratiation, is to likely flourish under ambiguity.

Another important factor in determining the influence of rater affect is rater accountability. Accountability refers to the amount of social pressure raters feel to justify their ratings of others (Tetlock & Kim, 1992). A high level of accountability means that raters anticipate having to explain their ratings in the future because accountability also means that there are outcomes, implicit or explicit rewards or punishments, for the raters (Fink & Klimoski, 1998). Research results indicate that

raters who feel accountable tend to be more lenient in ratings than raters who do not feel such pressure (Antonioni, 1994; Fisher, 1979; Klimoski & Inks, 1990). However, this tendency does not necessarily mean that the influence of rater affect is less strong when raters feel accountable. Rather, it is reasonable to expect that the accountability itself may serve as a check on the extent to which raters allow their affect to influence ratings. Schmitt and Klimoski argued that "accountability may increase the care and rigorousness with which the evaluation is made" (Schmitt & Klimoski, 1991: 182). In a similar vein, McBrearty (1982) argued that anonymity may result in irresponsible evaluations. This suggests that raters who feel accountable may exert more care when they rate others. Previous studies support this argument. Accountable raters made more accurate ratings in a performance evaluation task (Rosenbaum, Lehman & Holcom, 1993) and accountability was negatively associated with rating errors in downward appraisals (Mero & Motowidlo, 1995). Haeggberg and Chen (1999) found similar results in upward appraisals.

The amount of ambiguity and accountability faced by an individual rater may depend on his or her position in the organization and the type of feedback he or she is giving. Because of these factors, we hypothesized that the influence of rater affect would be stronger in peer and upward ratings than in downward ratings. This hypothesis is based first of all on a logical link between experience in rating and levels of ambiguity.

Because most organizations require managers to conduct annual performance appraisals and list performance evaluation as part of managers' job responsibilities, managers tend to have more experience at rating work-related traits and behaviors. In addition, organizations usually provide managers with training on conducting performance appraisals and giving performance feedback (Latham & Wexley, 1982; Smith, 1986). However, this is not the case with peers and direct reports. Because 360-degree feedback is a relatively new development, peers and direct reports usually have very little experience in rating others and are rarely trained to improve their rater performance (Antonioni & Woehr, 2000).

Different levels of experience with giving performance ratings may affect the ways people go about giving ratings. Supervisors may face less ambiguity than those giving peer and upward appraisals because they have more rating experience. Experience giving performance ratings may influence the extent of ambiguity because experienced raters tend to develop a set of cognitive categories for evaluation. Scholars believe that these cognitive categories influence the cognitive processes underlying performance appraisals, processes such as giving attention, categorizing, interpreting, and retrieving relevant information (Cardy et al., 1987; Ostroff & Ilgen, 1992; Srull & Wyer Jr., 1980); thus, the extent to which these categories are developed can determine how well the ratings reflect the ratee's performance (Borman, 1987; Ilgen & Feldman, 1983). Empirical studies support this argument, showing that raters with more experience conducting performance appraisals tended to develop cognitive categories that were more sensitive to relevant work behaviors (Cardy et al., 1987). This suggests that a rater's experience in evaluating performance may lead to more sophisticated evaluative categories (Murphy & Cleveland, 1991), which is likely to reduce the amount of ambiguity that a rater faces in ratings.

Furthermore, raters in downward ratings are more likely to be held accountable than raters giving peer and upward ratings. Organizational norms associated with performance evaluations support the concept that managers are accountable for the outcomes of the constructive feedback they give their direct reports. Specifically, managers expect to see their direct reports make improvements based on their constructive feedback (Frink & Klimoski, 1998). In addition, in the 360-degree feedback process, ratings are usually anonymous; however, the ratings from one's manager are almost always identifiable because the ratee usually has only one manager. Thus, raters giving downward appraisals may feel more accountable for their ratings. In addition,

some practitioners of 360-degree feedback strongly recommend that direct reports discuss their downward ratings with their managers; in these cases managers know that they will need to explain or justify the ratings they give in greater detail (Edwards & Ewen, 1996). By contrast, peer and direct report ratings are usually obtained from multiple anonymous raters and rarely require any face-to-face discussion. Because accountability may limit the extent to which raters allow their affect to influence ratings, we expected that the influence of rater affect on ratings would be greater in peer and upward feedback than in downward feedback.

Because peers and direct reports may face more ambiguity when giving performance evaluations and may feel less accountable for their ratings than managers do, peer and upward ratings may tend to be more susceptible to rater affect than downward ratings. Therefore, the following hypothesis was tested:

Hypothesis 2: The influence of rater affect on the leniency of ratings will be stronger in peer and upward ratings than in downward ratings.

We also investigated the extent to which a rater's observation time moderated the influences of interpersonal affect on appraisal ratings. Scholars have recommended that raters improve their observation skills to raise the quality of rater performance (Borman, 1974; Landy & Farr, 1980). In a field study of students' ratings of instructors, Bernardin & Walter (1977) found that students who were provided a structure for making observations of instructors and trained to keep a formal diary of teaching behaviors throughout the semester had significantly less leniency in ratings than untrained students. In addition, the results of a lab study, in which students were asked to evaluate a ratee's performance on a video, indicated that the amount of time subjects spent observing the ratee's work behavior correlated positively with accuracy in ratings (Favero & Ilgen, 1989). However, because raters only viewed a video of a ratee with whom they never interacted, they may not have experienced any affect towards the ratee; therefore, the results do not help us understand the interaction between observation and rater affect in real-world situations in which affect may be a factor.

It seems logical that increasing the rater's observation time should also increase the amount of information the rater has, which should in turn reduce the ambiguity in ratings. However, as mentioned earlier, rater affect can influence the cognitive process involved in performance appraisals in many ways. Rater affect may influence what raters observe. If raters have a strong positive or negative affect (liking) toward certain ratees, then the raters may notice specific work behaviors that conform to their feelings about the ratee. Raters may actively seek information consistent with their affect toward ratees as a way to preserve their own feelings about the ratees or preserve prior relationships with them (Robbins & DeNisi, 1994; DeNisi, Cafferty & Meglino, 1984; Snyder & Cantor, 1979). This phenomenon was demonstrated most clearly by an experimental field study that examined how structured diary keeping, an intervention previously shown to reduce raters' reliance on fallible memory, influences the effect of rater affect during appraisals (Varma et al., 1996). Much to their surprise, these researchers found that the increased rater observation that resulted from keeping a diary increased the influence of rater affect on ratings. Their further investigation revealed that rater affect was positively associated with the percentage of positive incidents recorded in diaries. These findings suggest that a rater may actively seek information consistent with his or her affect toward the ratee.

Rater affect may also influence how a rater interprets information. Feldman (1981) found that raters appear to attribute qualities to ratees that are consistent with their affect towards ratees. One study found that raters who liked their ratees attributed good performance to internal factors, such as ability, and poor performance to external factors, such as bad luck; the attributions tended to be reversed when the rater disliked

the ratee (Regan, Straus & Fazio, 1974). In an attempt to understand how rater affect influences ratings, Robbins and DeNisi (1994) conducted a lab study and found that rater affect influenced both the process and outcomes of ratings. Their results show that the raters tended to perceive affect-consistent information as more meaningful than affect-inconsistent information in the rating process.

The foregoing discussion suggests that rater affect influences raters' attention and interpretation in the rating process. Through repeated observations of ratees, raters may store up increasing amounts of affect-consistent information and interpretation. According to this model, the more a rater observes a ratee, the more likely it is that the ratings will become affect-consistent. Thus, rater observation time may moderate the influence of rater affect on ratings so that increased rater observation time results in greater influence of rater affect. Therefore, the following hypothesis was tested:

Hypothesis 3: The influence of rater affect on the leniency of ratings will increase as the rater observes the ratee more frequently.

3. Methods

3.1. Participants

A total of 433 out of 456 employees of a midsize insurance company volunteered to be part of the study. Participants represented all functional areas and positions within the company, as well as all levels of management. A number of years before the study the company decided to use self-managed teams; therefore, each of the 51 managers in the organization had only a few individuals as direct reports. Complete data existed for 163 downward ratings, 103 upward ratings, and 1027 peer ratings. All of the managers who participated had at least two years of experience conducting annual performance appraisals.

3.2. Measures

3.2.1. Independent variables

An attractive relationship scale developed by Tsui and Barry (1986) measured interpersonal affect. This scale is composed of three items that measure the extent to which a rater likes a ratee. Raters responded to statements such as "I like this person" using a seven-point response scale ranging from strongly agree (1) to strongly disagree (7). Cronbach's coefficient alpha for the interpersonal affect scale was 0.90.

Three items created specifically for this study measured the amount of rater observation time. Raters responded to statements such as "I frequently have the opportunity to observe the work behaviors of the person [name] I am evaluating," using a seven-point response scale ranging from Strongly Disagree (-3) to Strongly Agree (+3). Cronbach's coefficient alpha was 0.79 for the observation scale. Summing the three items produced scores on this measure ranging from -5 to 9.

3.2.2. Dependent variable

The performance measured by the present study is best conceptualized as contextual work behaviors (Borman & Motowidlo, 1993). As opposed to task performance, contextual work behaviors do not directly contribute to the technical core. But they do support the organizational, social, and psychological environment in which the technical core must function (Borman & Motowidlo, 1993). Such contextual behaviors are vital for the organization's efficient performance (George & Brief, 1992). Motowidlo and Van Scotter (1994) also found that task performance and contextual behaviors contribute

independently to overall performance. Because the types of contextual work behaviors required may vary across organizations, Organ (1988) recommended the use of site-specific measures to capture organizational conditions.

In the present study, 20-items measured contextual behaviors that were desired by the company. The desired work behaviors were linked to the company's vision and values. The measure included items such as "Demonstrates strong work ethic," "Treats others as equals," "Provides honest and constructive feedback," and "Confronts conflict directly." A five-point scale ranging from Strongly Disagree (1) to Strongly Agree (5) was used. To examine the dimensionality of the measure, a principal component factor analysis with varimax rotation was performed. Scree plot analysis suggested that a single factor solution accounted for the data adequately. Cronbach's coefficient alpha for the 20 items was 0.92. Thus, the work behavior measure was constructed by averaging the 20 items.

Researchers have operationalized leniency largely in two ways. If true scores

are known, leniency can be measured in an absolute sense by comparing ratings and the true scores (e.g., Farh & Werbel, 1986). However, true scores are largely unavailable in field settings. In such cases, researchers have operationalized leniency in a relative sense by comparing a rater's rating with the ratings of other raters (e.g., Kane et al., 1995). Because the true scores for contextual behaviors were not available in the present study, we operationalized leniency as the difference between an individual rating and the overall mean rating for that rater.

3.2.3. Control variables

Control variables, such as age, gender, and job tenure, were obtained from the company's archival sources.

3.3. Procedure

The research was done with a midsize insurance company that was commencing a 360-degree performance feedback process. The company intended to use 360-degree ratings as feedback for developmental purposes.

At the recommendation of the first author, all participants received two hours of training regarding the 360-degree feedback process. The training outlined the purpose of the 360-degree feedback and described how the 360-degree process provides data for assessing changes in valued work behaviors. Along with a review of the 360-degree feedback instrument, the training provided information on how peer raters were to be selected, how to prevent common rating errors (rater affect was not specifically mentioned), when to expect a summary feedback report, and how to set improvement goals and action plans. In addition, participants were informed that the company was conducting a Study on the 360-degree feedback process with a university. Participants were told that measurement instruments would have code numbers on them to link instruments together; however, all their individual responses would remain strictly confidential. They were also informed that participation in the study was voluntary. Participants were asked to select at least three peers who could evaluate them and each was required to get a supervisor's approval for the list of peer raters. This was done to make sure that participants selected peers who knew their work behaviors and performance. To reduce the influence of common method variance, measurement instruments were administered at different times (Podsakoff & Organ, 1986). The 360-degree instrument was administered first, and participants mailed their completed questionnaires to a data-entry company. Four weeks later the affect and observation questionnaire was administered. Each participant received one questionnaire for each individual he or she rated. Completed questionnaires were mailed to the first author at

his university address.

3.4. Analysis

Ratings in 360-degree feedback are unlikely to be independent. That is, downward ratings are nested within rater and upward ratings are nested within ratee. Peer ratings are nested both within rater and within ratee. This is likely to introduce dependencies in the data and thus violate the independence assumption that is the basis of most statistical analyses. Error terms might be correlated within the evaluations of one ratee if unobserved ratee characteristics affected ratings. Error terms might also be correlated within the data taken from a single rater if that rater has a general tendency to underrate or overrate others.

Breusch and Pagan's (1980) Lagrange multiplier tests were used to test the independence assumption within rater and within ratee. The Lagrange multiplier test determines whether the variance of rater- or ratee-specific error is zero. If the observations are independent within rater or within ratee, the variance of rater- or ratee-specific error should be zero. Because two potential sources of nonindependence existed, we first created a dummy variable for each rater and each ratee. Then, the set of dummy variables for ratees (or raters) was included in the Lagrange multiplier tests for independence within rater (or ratee). This controlled for the potential dependence within ratee (or within rater) while testing independence within rater (or within ratee). In all models, the Lagrange multiplier tests consistently rejected the hypothesis that the variance of rater- and ratee-specific errors was zero. This implies that the ratings were not independent within rater and within ratee.

The presence of nonindependence made it difficult to use conventional statistical methods for data analysis. Although regression coefficients from ordinary least squares (OLS) estimations are still unbiased, the standard errors associated with the coefficients are biased if observations are nonindependent (Kennedy, 1998). The fact that nonindependence stemmed from two sources (rater and ratee) made it more difficult to correct the problem. Several methods exist to correct nonindependence, but all of them deal with one source at a time. To correct for the nonindependence within rater and within ratee simultaneously, we adopted two different approaches in this study. To correct the nonindependence within rater, we estimated Huber and White's sandwich estimator of variance, which provides robust standard errors even when error terms show within group correlation (Huber, 1967; Rogers, 1993; White, 1980, 1982).

To correct the nonindependence within ratee, we used a fixed effects model. Although fixed effects models have been used largely in the context of panel data to control for unobserved individual characteristics that do not change over time (e.g., Gerhart, 1988; Gerhart & Milkovich, 1990), they can be used to control for nonindependence from unobserved heterogeneity whenever observations have an appropriate group structure (Chamberlain, 1980). Specifically, we took difference scores from their respective overall ratee means for each independent variable. Using such difference scores for both independent and dependent variables effectively corrects for nonindependence by removing all between-ratee variance, and only within-ratee variance was used to estimate the parameters of interest (Hsiao, 1986; Liker, Augustyniak & Duncan, 1985; Mundlak, 1978).

Please note that fixed effects models not only correct for nonindependence, but also eliminate any bias because of unobserved ratee characteristics that do not vary across observations. Researchers have argued that the association of rater affect with ratings does not necessarily represent bias. A rater may simply like a good performer (Lefkowitz, 2000; Varma, DeNisi & Peters, 1996). The fixed effects approach effectively controls for a ratee's performance that does not vary across raters.

4. Results

Table 1 reports descriptive statistics and correlations of the variables. Thirty three percentage of raters and 30% of ratees were male. The average ages of raters and ratees were 38.8 and 38.3, respectively. Raters had 2.81 years of job experience on average. The average job tenure of ratees was 2.86. In general, the raters had positive affect toward the ratees ($M = 5.69$) and they had an opportunity to observe the ratees ($M = 5.58$). The raters also tended to rate the ratees favorably ($M = 4.00$). Interpersonal affect ($r = 0.45$, p [less than] .001) and observation ($r = 0.10$, p [less than] .001) were positively associated with ratings.

Table 2 reports the results of fixed effects models with robust standard errors. The first column reports the results of the base model, which includes interpersonal affect, dummy variables representing peer and upward feedback ratings, and demographic variables for the raters. In the second model, the interaction between interpersonal affect and the source of evaluations was added to the base model. In the third model, the interaction of interpersonal affect and observation was added to the base model.

The first column shows that interpersonal affect ($[\beta] = 0.17$, p [less than] .001) was significantly related to leniency. Raters who expressed positive affect toward their ratees tended to be more lenient than raters who had negative affect. Interpersonal affect along with the control variables accounted for 17% of the total variance in the leniency of ratings ($[R.\text{sup}.2] = 0.17$, p [less than] .001).

The second column reports the interaction between interpersonal affect and the sources of performance feedback. In all three directions of 360-degree ratings, interpersonal affect was statistically significant ($[\beta] = 0.06$, p [less than] .05 for downward feedback, the omitted category; $[\beta] = 0.06 + 0.13 = 0.19$, p [less than] .001 for peer feedback; and $[\beta] = 0.06 + 0.14 = 0.20$, p [less than] .001 for upward feedback). Regardless of the source of performance feedback, rater affect was positively associated with rating leniency.

The results also show that the influence of interpersonal affect was stronger in peer ($[\beta] = 0.13$, p [less than] .001) and upward feedback ($[\beta] = 0.14$, p [less than] .01) than in downward feedback. However, the difference between peer and upward feedback ratings was not statistically significant ($d = 0.01$, n.s.). Adding interaction terms increased $[R.\text{sup}.2]$ by 0.01 (p [less than] .001).

The third column reports the effects of interaction between interpersonal affect and observation. A significant interaction effect between interpersonal affect and observation ($[\beta] = 0.03$, p [less than] .001) suggests that the influence of interpersonal affect on leniency increased as the rater observed the ratee more. Adding the interaction effects of interpersonal affect and observation increased $[R.\text{sup}.2]$ by 0.02 over the base model (p [less than] .001).

5. Discussion

The results of this study indicate that a rater's interpersonal affect does influence all three sources of 360-degree feedback (downward, upward, and peer), which is consistent with the findings of Tsui and Barry (1986). In addition, with this study we improved upon the methodology of their study by statistically controlling for unobserved ratee characteristics and by correcting for nonindependence among ratings. We also found that different sources in the 360-feedback process were influenced by interpersonal affect unequally. A major contribution of the present study is the finding that the influence of interpersonal affect was stronger in upward and peer ratings than it was in downward feedback. This finding raises some questions about the generalizability

of previous performance appraisal research. We believe that each source of 360-degree feedback may be influenced differently by the same factors; they may even be influenced by different factors.

As expected, the interaction between the amount of time raters observed ratees and the extent to which they liked the ratees was significant. That is, the more raters observed ratees, the influence of rater affect on ratings increased. Apparently raters are so strongly influenced by positive or negative affect that increased observation means noticing specific behaviors that conform to their affect or retrieving affect-consistent information at the time of the evaluation.

5.1. Implications for practitioners and future research

Practitioners in organizations that are using or planning to use 360-degree appraisal ratings to evaluate and determine merit raises need to be mindful of the results of this study. A rater's strong positive or negative feelings toward a ratee should not be overlooked or discounted. The authors recognize the challenge of acknowledging the impact of rater affect in organizations that strongly desire 360-degree ratings for determining merit raises. It may be equally difficult to convince raters themselves that liking ratees tends to bias their ratings. However, the results of this study provide empirical data suggesting that upward and peer 360-degree ratings may be biased by rater affect; therefore, at this point, these ratings should be used for the sole purpose of providing ratees with developmental feedback.

In organizations that use 360-degree feedback for developmental feedback, a couple of possibilities exist for managing the influence of raters' feelings toward ratees. It is important for those making decisions based on 360-degree feedback to acknowledge that rater affect plays an important role in the 360-degree feedback process. As Longenecker et al. (1992) indicated, we need to acknowledge that affect is a defining factor in performance evaluations in organizations. This acknowledgment does not mean, however, that these evaluations should be discounted. In fact, Varma et al. (1996) have indicated that liking or disliking someone's work behaviors and traits is as important as work results in determining overall performance, especially if human relations skills are important to the roles and responsibilities of the job. However, to minimize the influence of rater affect at the time of evaluation, raters should be trained to control their emotional reactions toward ratees (Cardy & Dobbins, 1986).

Robbins and DeNisi (1994) suggested guidelines for controlling the influence of affect on ratings. The following suggestions are based on modifications of those guidelines. First, administrators should acknowledge the impact of rater affect and address it by helping raters identify any strong positive or negative emotional reactions toward ratees. If a rater has strong feelings, then that individual needs to learn to make a relatively objective evaluation of the ratee. This means that raters need to know how to assess the diagnostic value of information they obtain from their observations to reduce the influence of their affect as they interpret information. This type of training should not be a one-time event, but an ongoing process focused on improving the ways in which raters observe and rate others. As Vance et al. (1983) state, most of the variance in leniency and halo is attributable to the behaviors of raters, rather than the work of the ratees. Therefore, it is important for academicians to continue conducting a systematic inquiry on rater affect (Longenecker et al., 1992) and for organizations to acknowledge Izard's (1982) statement that there is no evaluation without affect.

As with any study, there are limitations to the interpretations and generalizations made from this research. Because of the design of the study, no causal effects can be inferred from the results. A second limitation is potential common method variance. Each rater provided both ratings of rater affect and contextual behaviors. However, as Podsakoff and Organ (1986) recommended, we attempted to reduce the influences of common method variance by gathering the data at different times. Finally, because only one

organization was used, the ability to generalize the results from this study may be limited until the study is replicated in other organizations, although the results of our study do support the findings of Tsui and Barry (1986).

In the future, an experimental field study should investigate whether it is possible for raters to control for their affect toward ratees. A group trained to control for affect while making peer, upward, or downward evaluations should be compared to groups that receive no training. Another experimental field study should be conducted to determine whether training raters to assess the value of the information they obtained from their observations moderates the influence of the raters' interpersonal affect toward ratees.

David Antonioni, Ph.D. is associate professor of management and Chair of Executive Education in the School of Business at University of Wisconsin-Madison. He conducts applied research in the area 360-degree feedback. His publications appear in *Personnel Psychology*, *Academy of Management Executive*, *Human Resource Development Quarterly*, *Organizational Dynamics*, and *Industrial Management*.

Heejoon Park is a Ph.D. candidate at the Industrial Relations Research Institute, University of Wisconsin-Madison. His research interests include performance appraisal, dispute resolution, and the effect of institutions on organizational behaviors and performance.

(a.) University of Wisconsin--Madison, School of Business-Executive Education, 601 University Avenue, Madison, WI 53715-1035, USA

(b.) University of Wisconsin--Madison, Industrial Relations Research Institute, 1180 Observatory Drive, Madison, WI 53706, USA

(*) Corresponding author. Tel.: + 1-800-292-8964. E-mail address: dantonioni@bus.wisc.edu (D. Antonioni)

References

Alexander, E. R., & Wilkins, R. D. (1982). Performance rating validity: the relationship of objective and subjective measures of performance. *Group and Organizational Studies*, 7, 485-496.

Antonioni, D. (1994). The effects of feedback accountability on upward appraisal ratings. *Personnel Psychology*, 47, 349-356.

Antonioni, D. (1996). Designing an effective 360-degree appraisal feedback process. *Organizational Dynamics*, 25, 24-38.

Antonioni, D., & Woehr, D. J. (2000). Improving the quality of multi-source rater performance. In D. W. Bracken, C. W. Timmreck, & A. H. Church (Eds.), *Handbook of Multisource Feedback* (pp. 114-129). San Francisco: Jossey-Bass.

Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, S. W. (1998). Self-other agreement: does it really matter? *Personnel Psychology*, 51, 577-598.

Atwater, L. E., Roush, P., & Fischthal, A. (1995). The influence of upward feedback on self and follower ratings of leadership. *Personnel Psychology*, 48, 35-59.

Atwater, L. E., & Yammarino, F. J. (1993). Personal attributes as predictors of superiors' and subordinates' perception of military academy leadership. *Human Relations*, 46, 645-

Atwater, L. E., & Yammarino, F. J. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? *Personnel Psychology*, 45, 141-164.

Bernardin, H. 3., & Walter, C. S. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. *Journal of Applied Psychology*, 62, 64-69.

Bettenhausen, K. L., & Fedor, D. B. (1997). Peer and upward appraisals: a comparison of their benefits and problems. *Group and Organizational Management*, 22, 236-263.

Borman, W. C. (1974). The rating of individuals in organizations: an alternative approach. *Organizational Behavior and Human Performance*, 12, 105-124.

Borman, W. C. (1987). Personal constructs, performance schema, and "folk theories" of subordinate effectiveness: explorations in an army officer sample. *Organizational Behavior and Human Decision Processes*, 40, 307-322.

Borman, W. C., & Motowidlo, S. J. (1993). Expanding the criterion domain to include elements of contextual performance. In N. Schmitt & W. C. Borman, *Personnel selection in organizations* (pp. 71-98). San Francisco: Jossey-Bass.

Bracken, D. W., Timmreck, C. W., & Church, A. H. (2000). History and development of multiple-source feedback as a methodology. In D. Bracken, C. Timmreck, & A. Church (Eds.), *Handbook of MultiSource Feedback* (pp. 3-14). San Francisco: Jossey-Bass.

Breusch, T., & Pagan, A. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. *Review of Economic Studies*, 47, 239-253.

Cardy, R. L., Bernardin, H. J., Abbott, J. G., Senderak, M. P., & Taylor, K. (1987). The effects of individual performance schemata and dimension familiarization on rating accuracy. *Journal of Occupational Psychology*, 60, 197-205.

Cardy, R. L., & Dobbins, G. H. (1986). Affect and appraisal accuracy: liking as an integral dimension in evaluating performance. *Journal of Applied Psychology*, 71, 672-678.

Chamberlain, G. (1980). Analysis covariance with qualitative data. *Review of Economic Studies*, 47, 225-238.

Decotiis, T., & Petit, A. (1978). The performance appraisal process: A model and some testable propositions. *Academy of Management Review*, 3, 635-646.

DeNisi, A. S., Cafferty, T. P., & Meglino, B. M. (1984). A cognitive view of the performance appraisal process: a model and research propositions. *Organizational Behavior and Human Performance*, 6, 109-155.

Dipboye, R. L. (1985). Some neglected variables in research on discrimination in appraisals. *Academy of Management Review*, 10, 116-127.

Edwards, M., & Ewen, A. J. (1996). 360-degree feedback--the powerful new model for employee assessment & performance improvement. New York: AMACON.

Farh, J., & Werbel, J. D. (1986). Effects of purpose of the appraisal and expectation of validation on self-appraisal leniency. *Journal of*

Applied Psychology, 71, 527-529.

Favero, J. L., & Ilgen, D. R. (1989). The effects of ratee prototypicality on rater observation and accuracy. *Journal of Applied Social Psychology*, 19, 932-946.

Feldman, J. M. (1981). Beyond attribution theory: cognitive processes in performance appraisal. *Journal of Applied Psychology*, 66, 127-148.

Ferris, G. R., & Judge, T. A. (1991). Personnel/human resources management: a political influence perspective. *Journal of Management*, 17, 447-488.

Fisher, C. C. (1979). Transmission of positive and negative feedback to subordinates: a laboratory investigation. *Journal of Applied Psychology*, 64, 533-540.

Frink, D. D., & Klimoski, R. J. (1998). Toward a theory of accountability in organizations and organizations and human resources management. *Research in Personnel and Human Resources Management*, 16, 1-51.

George, J. M., & Brief, A. P. (1992). Feeling good-doing good: a conceptual analysis of the mood at work-- organizational spontaneity relationship. *Psychological Bulletin*, 112, 310-329.

Gerhart, B. (1988). Sources of variance in incumbent perceptions of job complexity. *Journal of Applied Psychology*, 73, 154-162.

Gerhart, B., & Milkovich, G. T. (1990). Organizational differences in managerial compensation and financial performance. *Academy of Management Journal*, 33, 663-691.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of failure to meet assumptions underlying fixed effects analyses of variance and covariance. *Review of Educational Research*, 42, 237-288.

Haeggberg, R. L., & Chen, P. Y. (1999). Can accountability assure the accuracy of upward appraisals? Paper presented at the Fourteenth Annual Conference for the Society for Industrial and Organizational Psychology, Atlanta.

Holzback, R. L. (1978). Rater bias in performance ratings: superior, self, and peer ratings. *Journal of Applied Psychology*, 63, 579-588.

Hsiao, C. (1986). *Analysis of panel data*. Cambridge: Cambridge University Press.

Huber, P. J. (1967). The behavior of maximum likelihood estimates under non-standard conditions. In *Proceedings of the Fifth Berkeley Symposium in Mathematical Statistics and Probability* (pp. 221-233). Berkeley, CA: University of California Press.

- Ilgén, D. R., & Feldman, J. R. (1983). Performance appraisals: a process focus. In B. M. Staw & L. L. Cummings (Eds.), *Research in organizational behavior*. (Vol. 5, pp. 141-197). Greenwich, CT: JAI Press.
- Isen, A. M., Shalcker, T. E., Clark, M. S., & Karp, L. (1978). Positive affect, accessibility of material in memory, and behavior: a cognitive loop. *Journal of Personality and Social Psychology*, 36, 1-12.
- Izard, C. E. (1982). Comments on emotion and cognition: can there be a working relationship? In M. S. Clark & S. T. Fiske (Eds.), *Affect and Cognition* (pp. 229-240). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.
- Jones, E. E. (1964). *Ingratiation*. New York: Appleton-Century-Crofts.
- Judge, T. A., & Ferris, G. R. (1993). Social context of performance evaluation decisions. *Academy of Management Journal*, 36, 80-105.
- Kane, J. S., Bernardin, H. J., Villanova, P., & Peyrefitte, J. (1995). Stability of rater leniency: three studies. *Academy of Management Journal*, 38, 1036-1051.
- Kennedy, P. (1998). *A guide to econometrics*. Cambridge, MA: The MIT Press.
- Kenny, D. A., & Judd, C. M. (1986). Consequences of violating the independence assumption in analysis of variance. *Psychological Bulletin*, 99, 422-431.
- Klimoski, R., & Inks, L. (1990). Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes*, 45, 194-208.
- Kingstrom, P. O., & Mainstone, L. E. (1985). An investigation of the rater-ratee acquaintance and rater bias. *Academy of Management Journal*, 28, 641-653.
- Kumar, K., & Beyerlein, M. (1991). Construction and validation of an instrument for measuring ingratiation behaviors in organizational settings. *Journal of Applied Psychology*, 76, 619-622.
- Landy, F. J., & Farr, J. L. (1980). Performance ratings. *Psychological Bulletin*, 87, 72-107.
- Latham, G. P., & Wexley, K. N. (1982). *Increasing productivity through performance appraisal*. Reading, MA: Addison-Wesley.
- Lee, D. M., & Alvares, K. M. (1977). Effects of sex on descriptions and evaluations of supervisory behavior in a simulated industrial setting. *Journal of Applied Psychology*, 62, 405-410.
- Lefkowitz, J. (2000). The role of interpersonal affective regard in supervisory performance ratings: a literature review and proposed causal model. *Journal of Occupational and Organizational*

Psychology, 73, 67-85.

Liker, J. K., Augustyniak, S., & Duncan, G. J. (1985). Panel data and models of change: a comparison of first difference and conventional two-wave models. *Social Science Research*, 14, 80-101.

London, M., & Smither, J. W. (1995). Can multi-source feedback change perceptions of goal accomplishments, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, 48, 803-839.

London, M., & Beatty, R. W. (1993). 360-degree feedback as a competitive advantage. *Human Resource Management*, 32, 353-374.

Longenecker, C. O., Sims, H. P., Jr., & Gioia, D. A. (1987). Behind the mask: the politics of employee appraisal. *Academy of Management Executive*, 7, 183-193.

Longenecker, C. O., Jaccoud, H. P., Sims, H. P., Jr., & Gioia, D. A. (1992). Quantitative and qualitative investigations of affect in executive judgment. *Applied Psychology: An International Review*, 41, 21-41.

McBrearty, P. (1982). We should abolish anonymous evaluations of teachers by their students. *Chronicle of Higher Education*, 7, 25-48.

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.

Motowidlo, S. J., & Van Scotter, J. R. (1994). Evidence that task performance should be distinguished from contextual performance. *Journal of Applied Psychology*, 79, 475-480.

Mundlak, Y. (1978). On the pooling of time series and cross sectional data. *Econometrica*, 46, 69-86.

Murphy, K. R., & Cleveland, J. N. (1991). *Performance appraisal: an organizational perspective*. Englewood Cliffs, NJ: Prentice Hall.

Organ, D. (1988). *Organizational citizenship behavior: the good soldier syndrome*. Lexington, MA: Lexington Books.

Park, O. S., Sims, H. P., & Motowidlo, S. J. (1986). Affect in organizations. In H. P. Sims & D. A. Gioia (Eds.), *The thinking organization*. (pp. 77-90). San Francisco, CA: Jossey-Bass.

Podsakoff, P., & Organ, D. W. (1986). Self reports in organizational research: problems and prospects. *Journal of Management*, 12, 531-544.

Regan, D. T., Straus, E., & Fazio, R. (1974). Liking and the attribution process. *Journal of Experimental Social Psychology*, 10, 385-397.

Robbins, T. L., & DeNisi, A. S. (1994). A closer look at interpersonal

affect as a distinct influence on cognitive processing in performance evaluations. *Journal of Applied Psychology*, 79, 341-353.

Rogers, W. H. (1993). Regression standard errors in clustered samples. *Stata Technical Bulletin*, 13, 19-23.

Romano, C. (1994). Conquering the fear of feedback. *HR Focus*, 71, 9-10.

Rosenbaum, A. E., Lehman, W. E. K., & Holcom, M. L. (1993). Effects of personal accountability and rating consequences on performance evaluation. Paper presented at the Eighth Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Snal, F. E., Downey, R. G., & Lahey, M. A. (1980). Rating the ratings: assessing the psychometric quality of rating data. *Psychological Bulletin*, 88, 413-428.

Schmitt, N., & Klimoski, R. (1991). *Research methods in human resource management*. Cincinnati: Southwestern Publishing.

Smith, D. E. (1986). Training programs for performance appraisal: a review. *Academy of Management Review*, 11, 22-40.

Snyder, M., & Cantor, N. (1979). Testing hypotheses about other people: the use of historical knowledge. *Journal of Experimental Social Psychology*, 15, 330-342.

Srull, T. K., & Wyer, R. S., Jr. (1980). Category accessibility and social perception: some implications for the study of person memory and interpersonal judgments. *Journal of Personality and Social Psychology*, 38, 841-856.

Tetlock, P. E., & Kim, J. I. (1987). Accountability and judgment processes in a personality prediction task. *Journal of Personality and Social Psychology*, 52, 700-709.

Tornow, W. W. (1993). Editor's note: introduction to special issue on 360-degree feedback. *Human Resource Management*, 32, 21 1-220.

Tornow, W. W., London, M., & CCL Associates. (1998). *Maximizing the value of 360-degree feedback. A process for successful individual and organizational development*. San Francisco: Jossey-Bass.

Tsui, A. S., & Barry, B. (1986). Interpersonal affect and rating errors. *Academy of Management Journal*, 29, 586-599.

Vance, R. J., Winne, P. S., & Wright, S. E. (1983). A longitudinal examination of rater and ratee effects in performance ratings. *Personnel Psychology*, 36, 609-621.

Varma, A., DeNisi, A. S., & Peters, L. H. (1996). Interpersonal affect and performance appraisal: a field study. *Personnel Psychology*, 49, 341-360.

Waldman, D., Atwater, L. A., & Antonioni, D. (1998). Has 360-degree feedback gone amok? *Academy of Management Executive*, 12, 86-94.

Waldman, D., & Thornton, G. C., III. (1988). A field study of rating conditions and leniency in performance appraisal. *Psychological Reports*, 63, 835-840.

Wayne, S. J., & Ferris, G. R., 1990. Influence tactics, affect, and exchange quality in supervisor-subordinate interactions: a laboratory experiment and a field study. *Journal of Applied Psychology*, 75, 487-499.

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48, 817-830.

White, H. (1982). Maximum likelihood estimation of mis-specified models. *Econometrica*, 50, 1-25.

Zajonc, R. B. (1980). Feeling and thinking: preferences need no Inferences. *American Psychologist*, 35, 15 1-175.

Table 1 Means, standard deviations, and correlations

	Mean	S.D.	Correlation			
			(1)	(2)	(3)	
(1) Ratee age	38.33	9.40				
(2) Rater age	38.79	8.76	.12 ***			
(3) Ratee gender a	.30	.46	.00	.00		
(4) Rater gender a	.33	.47	-.01	.04	.32 ***	
(5) Ratee job tenure	2.86	2.06	.29 ***	.04	.08 **	
(6) Rater job tenure	2.81	2.00	.07 **	.29 ***	.04	
(7) Affect	5.69	1.12	.03	.05	.05	
(8) Observation	5.58	1.34	.06 *	.10 ***	.02	
(9) Ratings	4.00	.51	-.03	.05	.05	
	(4)	(5)	(6)	(7)	(8)	
(1) Ratee age						
(2) Rater age						
(3) Ratee gender a						
(4) Rater gender a						
(5) Ratee job tenure	.07 **					
(6) Rater job tenure	.08 **	.16 ***				
(7) Affect	.03	-.02	.02			
(8) Observation	.10 ***	.16 ***	.06 *	.39 ***		
(9) Ratings	.02	-.06 *	-.09 **	.45 ***	.10 ***	

(a)Male = 1, female = 0.
 (***)p [less than] .001.
 (**)p [less than] .01.
 (*)p [less than] .05.

Table 2 Results of fixed effects models with robust standard errors

	(1)		(2)		(3)
	[beta]	S.E.	[beta]	S.E.	[beta]
Affect	.17 ***	.02	.06 *	.03	.04
Affect* peer			.13 ***	.03	
Affect* upward			.14 **	.05	
Observation					-.08 ***
Affect* observation					.03 ***
Peer feedback a	-.08	.04	-.78 ***	.20	-.08 *
Upward feedback b	-.06	.06	-.84 **	.29	-.07
Rater age	.00	.00	.00	.00	.00
Rater job tenure	-.04 ***	.01	-.04 ***	.01	-.04 ***
Rater gender c	-.07	.04	-.06	.04	-.06
[R.sup.2]		.17 ***		.08 ***	
[R.sup.2] Change d				.01 ***	
	S.E.				
Affect	.04				
Affect* peer					
Affect* upward					
Observation	.04				
Affect* observation	.01				
Peer feedback a	.04				

Upward feedback b .06
Rater age .00
Rater job tenure .01
Rater gender c .04
[R.sup.2] .09 ***
[R.sup.2] Change d .02 ***
(a)1 if peer feedback ratings, 0 otherwise.
(b)1 if upward feedback ratings, 0 otherwise
(c)Male = 1, female = 0.
(d)Change from the first model.
(***)p [less than] .001.
(**)p [less than] .01.
(*)p [less than] .05.

-

COPYRIGHT 2001 JAI Press, Inc.
in association with The Gale Group and LookSmart. COPYRIGHT
2001 Gale Group

-