

Self-Other Agreement on a 360-Degree Leadership Evaluation

Stefanie K. Halverson  
Rice University

Scott Tonidandel  
Rice University

Cassie Barlow  
United States Air Force

Robert L. Dipboye  
Rice University

Paper presented at the 17<sup>th</sup> Annual Conference of the Society for Industrial and Organizational Psychology, Toronto, Canada. April, 2002.

### ABSTRACT

The current study investigates the relationship between Air Force promotion rate and: self-supervisor (n=825), self-peer (n=285), and self-subordinate (n=256) agreement on a leadership scale. Results indicate that self-subordinate agreement a better predictor of promotion rate than self-superior or self-peer agreement. Analyses were completed using Polynomial Regressions Equations (Edwards, 1993; 1994).

### PRESS PARAGRAPH

With the increase in the use of 360-degree feedback measures as a performance appraisal and development tool (Dalton, 1996), it has become more important to conduct research on statistical and methodological practices associated with this method. The current study investigates the relationship of self-ratings (n=2,117) with supervisor-ratings (n=825), peer-ratings (n=285), and subordinate-ratings (n=256). Based on a leadership survey of Air Force Officers, we found that agreement between self and subordinate was more important in predicting promotion rates than agreement between self and supervisor or self and peers. Analyses were completed using Polynomial Regressions Equations (Edwards, 1993; 1994).

Industrial-Organizational Psychologists, Human Resource Managers, and performance appraisal researchers have tended to focus their efforts on technologies for improving the appraisal process, such as BARS, MSS, BOS, and appraisal/feedback systems such as MBO and split role appraisals. The latest appraisal/feedback systems technology on the rise is the multisource, multirater (or 360-degree) feedback method, in which data is collected from individuals (self-ratings) along with supervisors, peers, and subordinates (Church, 2000). There are estimates that 12% - 29% of organizations are using this method, and that the use of this method is increasing (Church, 2000; Mercer's Fax Facts; Waldman, Atwater, & Antonioni, 1998). The 360-degree method has primarily been employed as a feedback and development tool, under the assumption that increasing one's level of self-other agreement will facilitate the implementation of the necessary changes (Goodstone & Diamante, 1998). However, there is a growing trend in the use of 360-degree feedback as an administrative or performance appraisal tool (Dalton, 1996). Self-other agreement or the similarity/difference in self-ratings compared to other ratings has been considered critical in both uses and has spurred a wealth of research (Atwater et al., 1998; Atwater, Roush, & Fischthal, 1995; Atwater & Yammarino, 1992; Brutus, Fleenor, & Taylor, 1996; Church, 2000; Dalton, 1996; Fleenor, et al., 1996; Goodstone & Diamante, 1998; Johnson & Ferstl, 1999; Van Velsor, et al., 1993).

Self-ratings, in themselves, are not generally seen as accurate predictors of performance (Harris & Schaubroeck, 1988; Hough, Keyes, & Dunnette, 1983) and often suffer from a leniency bias (Podsakoff & Organ, 1986). However, as Atwater and Yammarino (1992) suggested, self-ratings may provide some insight into raters' traits or

dispositions, such as self-esteem, or self-consciousness. Moreover, coupled with information from others' (supervisors, peers, subordinates) ratings, self-ratings should provide insight into one's level of self-awareness. In a review of the self-evaluation literature Mabe and West (1982) found that more accurate (or aware) raters tended to be higher in intelligence, achievement status, and internal locus of control. Wicklund's (1975; 1978; 1979) theory of self-awareness suggests that self-aware persons attend more to others' perceptions of them and, therefore, can utilize this information to appropriately change their behavior.

Recent work by Atwater et al. (1998), Atwater, Roush, and Fischthal (1995), Atwater and Yammarino (1992), and Van Velsor, et al. (1993) has suggested that self-other agreement reflects self-awareness and is related to several outcome measures including leadership and managerial effectiveness. In addition, these studies explore the importance of the direction of disagreement, suggesting that those who are in-agreement with their other receive the highest ratings, followed by under-raters (those who rate themselves below their other), and over-raters (those who rate themselves above their other) receive the lowest ratings (Atwater, Roush, & Fischthal, 1995; Atwater & Yammarino, 1992; Van Velsor, et al., 1993). Subsequent work by Yammarino and Atwater (1997) proposed a 4-group categorization model with overestimators, underestimators, in-agreement good, and in-agreement bad. They propose good performance for in-agreement good, mixed performance for underestimators, and poor performance for in-agreement bad and overestimators. However, Fleenor, et al. (199) reports results that are somewhat at odds with the previous work of Atwater and colleagues. Using a six-factor model including performance (high or low) along with the

self-other agreement indices (high performance/ in-agreement, low performance/in-agreement, high performance/ under-rater, etc.) Fleenor, et al. (1996) found that other ratings alone (whether they were above or below the mean) predicted effectiveness. He also found that self-other agreement no longer predicted effectiveness. Similarly, Brutus, Fleenor, and Taylor (1996) found that neither self-ratings alone, nor self-other agreement predicted performance, while peer ratings was a relevant predictor.

Despite the wealth of research in this area, several concerns have been raised which may account for previous inconsistencies in the findings. First, there has been some disagreement over the method by which self-other agreement is determined. Work in this area used difference scores, correlations (London & Wohlers, 1991; Wohlers & London), and categorical agreement. The majority of the previous research has been done with Profile Similarity Indexes (PSIs), using difference scores. These methods generally use one of three methods to obtain non-negative scores: D-squared, the square root of D-squared, or the absolute value of D. Edwards (1994) has shown the shortcomings of these methods, and the importance of conceptualizing the self-other relationship in a three-dimensional plane (Edwards, 1993; 1994; Edwards & Parry, 1993). Edwards (1993) explains that the D-squared method assigns greater weight to differences of a larger magnitude. Also, the D-squared method, as well as the square root of D-squared, and the absolute value of D method are non-directional (see Edwards, 1993; 1994 for a complete description of these and other difference score methods).

Edwards (1995) explains the advantages to, and theoretical reasons behind, the advantages of polynomial regression equations over alternative approaches to difference scores in this research. First, because the constructs measured generally represent

different constructs (in this case self ratings versus other ratings), they should be measured as different constructs. Second, multivariate analyses allow us to treat each dependent measure jointly, allowing us to test the effect of each variable individually and at the same time test both variables together. Third, multivariate analyses allow us to maintain the directionality of the relationship between two variables (whether  $x > y$ , or  $y > x$ ) as opposed to simply looking at the absolute difference between variables (see Edwards, 1995). Therefore, it is suggested that research using self-other agreement should be analyzed using a multivariate approach.

A second source of concern in this area of research is the type of outcome measures used across these studies. This problem is two-fold. Most of the previous research has used ratings of performance by one of the sources as the outcome measure. Consequently, previous findings are subject to common method bias. In research involving the relationship between self-other agreement and performance/ effectiveness outcomes, it is important that objective outcome measures are used, as opposed to a different set of “other” ratings.

A third limitation concerns the types of 360 survey items used. There is little congruence across the literature in what types of survey items are used, and poor-item development may threaten the validities and reliabilities of 360-degree surveys (Church 2000; Church & Waclawski, 1998a). Moreover, the importance of the “other” may differ with the attributes measured in the 360-degree survey. For example, leadership behavior (see for e.g. Atwater & Yammarino, 1992) may be more relevant in predicting outcomes when subordinate perceptions are the source of the 360-degree survey than when superior or peer perceptions are the source. The current study addresses ratings on a leadership

questionnaire. In this case, self-subordinate ratings may be the best indicator of performance. Many leadership theorists have suggested that followers' perspectives are necessary in order to evaluate leadership ability (Bass, 1990; House et al., 1991). These "follower-centered" theories define leadership in terms of follower attributions. Therefore, when considering leadership performance it may be particularly important that leaders are aware of these perceptions.

The current study was an attempt to address several of the limitations in the previous research on 360-degree appraisals. We compared agreement between self-ratings on a leadership measure and the ratings of others on the same measure for three different sources: supervisors, subordinates, and peers. Agreement on this leadership index was used to predict a concurrent, objective outcome measure in the form of promotion rate using polynomial regression equations. Consistent with Atwater & Yammarino (1992) and Atwater, Roush, and Fischthal (1995), we hypothesized that in-agreement raters will be the highest performers, followed by under-raters, and over-raters for all organizational others. Moreover, because the current study uses a measure of leadership performance, we hypothesized that the relationship between self-subordinate agreement would be particularly important for predicting future performance.

#### Method

Participants were chosen by asking the US Airforce personnel center to choose 1,500 survey recipients from each of these categories: Lieutenant, Captain, Major and Lieutenant Colonel. A survey was then sent to that person (the survey recipient). Each recipient was responsible for sending the survey to his or her superior. That superior was then responsible for sending the survey to one of the recipient's peers and one of the

recipient's subordinates. This method was chosen to reduce that potential bias that may have arisen from the recipient choosing his or her own peer and subordinate. The response rate for the survey was 37%, which is a normal response rate for Air Force surveys. Demographic characteristics of the sample were similar to the Air Force at large in rank, race, and source of commission. There were slightly more male respondents in the survey than there are males in the Air Force at large. The overall sample was 89% Caucasian and 88% Male. A total of 2,117 self(s), 825 supervisors, 285 peers, and 256 subordinates responded.

### Measures

The 360-degree instrument. The 360-degree instrument consisted of a leadership questionnaire that was developed, and tested, by the Air Force Academy's Office of Institutional Research and Assessment (HQUSAF/A/XPR). The measure was based on competencies outlined by a special advisory group concerned with defining the critical attributes of leadership in the Air Force. The final scale consisted of 54 items. Principal Components Analysis, using a Varimax rotation and Kaiser normalization, for self-ratings, resulted in 4 factors: command/critical thinking, communication, teamwork/ peer leadership, management/mentoring. However, internal consistency ratings of the overall scale ranged from .97 to .99, so the overall scale was used.

Promotion rate. The outcome measure (promotion rate) was calculated by computing the number of promotions each officer had received, and the schedule on which each promotion was received (behind schedule, on-time, or ahead of schedule). Each promotion was given a score of one if it was behind schedule, two if it was on time, and three if it was ahead of schedule. Time in the military was not controlled for, as it

would be expected that greater amount of time in the military should affect leadership abilities, and the promotion schedule should separate high performers from low performers, regardless of time within the military.

## Results

Descriptive statistics, intercorrelations and internal consistency reliability estimates for the study variables are presented in Table 1.

Consistent with the recommendations of Edwards (1993; 1994; Edwards & Parry, 1993) the present study uses polynomial regression and response surface methodology to assess the impact of similarity in ratings from different sources on promotion rate. A series of hierarchical regression equations were computed that regressed promotion rate on self-ratings and others' ratings in step 1. The square of self-ratings, the square of other's ratings, and the interaction of self by other's ratings were added in step 2. A significant increase in the amount of variance explained ( $R^2$ ) in step 2 indicates a nonlinear relationship and requires the use of response surface methodology to clarify the nature of the relationship.

The results of a hierarchical regression analysis examining the relationship between self and subordinate ratings on promotion scores are presented in Table 2. Due to the significance of the two models and the significant increase in variance explained in step 2,  $\Delta R^2 = .03$ ,  $F(3,247) = 2.80$ ,  $p = .04$ , the surface that corresponds to the equation was examined (see Figure 1). The slope along the line of perfect agreement (where self rating = subordinate's rating) was positive and significant,  $a_1 = 3.54$ ,  $p < .05$ , and there was a significant concave curvature to the surface,  $a_2 = -2.48$ ,  $p < .05$ . When self-ratings were in agreement with subordinates' ratings, promotion scores increased with higher

ratings, but at a diminishing rate. We also examined the surface at the point of maximum disagreement (self ratings = 7 and subordinate ratings = 1, or vice versa). The slope and curvature of the surface at this point were  $\underline{x}_1 = -0.92$ ,  $p > .05$ , and  $\underline{x}_2 = -2.23$ ,  $p < .05$ , respectively. These results indicate that promotion rate is slightly lower for individuals who severely overestimate their performance and slightly higher for individuals who severely underestimate their performance compared to their subordinates. In general, these results indicate that promotion rate tended to be greater when both self- and subordinate ratings are high than when they are low.

Table 3 contains the results of a hierarchical regression analysis examining the relationship between self and peer ratings on promotion rate. An inspection of Table 3 reveals that neither step 1 nor step 2 of the regression equation was statistically significant,  $F(2,280) = 2.43$ ,  $p = .09$ ,  $F(5,280) = 1.29$ ,  $p = .27$ , respectively. The simultaneous effect of self- and peer ratings fails to have a significant impact on promotion rate.

The results of a hierarchical regression analysis examining the relationship between self and supervisor ratings on promotion rate are presented in Table 4. Because step 2 of the regression equation was non significant,  $F(5,808) = 2.00$ ,  $p = .076$ , and there was not a significant increase in  $R^2$ ,  $\Delta R^2 = .003$ ,  $F(3,808) = .545$ , the nonlinear terms are not necessary to explain the relationship between self- and supervisor ratings and promotion rate. The overall model tested in step 1 was significant,  $F(2,211) = 3.94$ ,  $p = .02$ , and an examination of the results indicates that self-ratings significantly predicted promotion rate,  $\beta = 0.079$ ,  $p = .024$ , whereas supervisor ratings did not,  $\beta = 0.060$ ,  $p = .076$ .

In general, these results suggest that only self- and subordinate ratings are significantly and positively related to promotion rate. Moreover, the response surface that describes the nature of the association between self- and subordinates' ratings indicates that both sets of ratings need to be considered simultaneously in order to understand their relationship to promotion rate.

### Discussion

There are three unique aspects of the present study. The present study is one of only a few studies to date that have used polynomial regression equations in analyzing self-other agreement in 360-degree appraisals (Atwater, et al., 1998; Johnson & Ferstl, 1999). The use of a relatively objective outcome measure, i.e., promotion rate, is also unique to this study given that previous research has tended to rely on ratings to measure the outcome of self-other agreement. Using an employee's rate of promotion over time is helpful in that it reflects effectiveness across several positions and with several superiors, and should reflect other hard measures of effectiveness. Finally, in this study we compared self-peer, self-subordinate, and self-supervisor agreement in ratings of leadership on their ability to predict promotion rate.

Consistent with our hypothesis, self-subordinate agreement was more important than self-peer or self-supervisor agreement in predicting promotion rate from these leadership ratings. Looking at the relationship between self-subordinate agreement and promotion rate, we found that those who were high in agreement had the highest promotion rate, and as ratings (on the 360-degree instrument) dropped, so did promotion rate. These findings are consistent Yammarino and Atwater (1997) predictions for in-

agreement raters. Extreme underestimators had lower promotion rates than those in-agreement, followed by extreme overestimators. This is also consistent with previous findings (Atwater, Roush, & Fischthal, 1995; Atwater & Yammarino, 1992; Van Velsor, et al., 1993). In contrast, we did not find that self-superior or self-peer agreement was related to promotion rate. These findings suggest support for our hypothesis that agreement between self and supervisor and agreement between self and peers on leadership perceptions may not be as important in predicting leadership effectiveness as agreement between self and subordinates.

These findings lend support to a follower-centered perspective of leadership (Bass, 1990; House et al., 1991), which suggests that the followers' perspectives are necessary in order to evaluate leadership ability. Follower-centered theories define leadership in terms of an attribution that the follower (or in this case subordinate) must make towards his/her leader. Therefore, it seems that a leader's self-awareness of his/her subordinates' perceptions of him/her should be especially important to leader effectiveness and more important than either peer or supervisor ratings. This may be particularly relevant in a military setting, where the ability to command the respect and obedience of one's subordinates is important to leadership success.

In conclusion, the present study demonstrates that self-awareness, as operationalized by the discrepancy between self- and other- ratings, contributes above and beyond the component ratings to the prediction of an important outcome measure. Moreover, the findings suggest that agreement of self and subordinates may be especially important when leadership behavior is the basis of the self and subordinate ratings. The

findings are not conclusive but do suggest some important directions for future research on self-awareness as a basis of appraisal in 360-degree systems.

## References

- Atwater, L. E., Ostroff, C., Yammarino, F. J., & Fleenor, J. 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. (1992). Does self-other agreement on leadership perceptions moderate the validity of leadership and performance predictions? Personnel Psychology, *45*, 141 - 164.
- Bass, B. (1990). Bass and Stogdill's handbook handbook of leadership. New York: Free Press.
- Brutus, S., Fleenor, J., & Taylor, S., (1996, April). Methodological issues in 360-degree feedback research. Paper presented at the Annual Conference of the Society for Industrial and Organizational Psychology, Inc. San Diego.
- Church, A. H. (2000). Do higher performing managers actually receive better ratings? A validation of multirater assessment methodology. Consulting Psychology Journal: Practice and Research, *52*, 99 - 116.
- Church, A. H. & Waclawski, J. (1998a). Designing and using organizational surveys. Aldershot, England: Gower
- Dalton, M. (1996). Multirater feedback and conditions for change. Consulting Psychology Journal: Practice and Research, *48*, 12 - 16.
- Edwards, J. R., (1993). Problems with the use of profile similarity indices in the study of congruence in organizational research. Personnel Psychology, *46*, 641 - 665.

- Edwards, J. R. (1994). The study of congruence in organizational behavior research: Critique and proposed alternative. Organizational Behavior and Human Decision Processes, *58*, 683 - 689.
- Edwards, J. R. (1995). Alternatives to difference scores as dependent variables in the study of congruence in organizational research. Organizational Behavior and Human Decision Processes, *64*, 307 - 324.
- Edwards, J. R., & Parry, M. E. (1993). On the use of polynomial regression equations as an alternative to difference scores in organizational research. Academy of Management Journal, *36*, 1577 - 1613.
- Fleenor, J., McCauley, C., Brutus, S. (1996). Self-other rating agreement and leader effectiveness. Leadership Quarterly, *7*, 487 - 506.
- Goodstone, M. S. & Diamante, T. (1998). Organizational use of therapeutic change strengthening multisource feedback systems through interdisciplinary coaching. Consulting Psychology Journal: Practice and Research, *50*, 152 - 163.
- Harris, M., & Schaubroeck, J., (1988). A meta-analysis of super-supervisor, self-peer, and peer-supervisor ratings. Personnel Psychology, *41*, 43 - 62.
- Hough, L, Keyes, M., & Dunnette, M. (1983). An evaluation of three “alternative” selection procedures. Personnel Psychology, *36*, 261 - 275.
- House, R. J., Spangler, W. D., & Woycke, J. (1991). Personality and charisma in the U.S. presidency: A psychological theory of leadership effectiveness. Administrative Science Quarterly, *36*, 364-396.

- Johnson, J. W. & Ferstl, K. L. (1999). The effects of interrater and self-other agreement on performance improvement following upward feedback. Personnel Psychology, 52, 271 - 303.
- London, M., Wohlers, A. (1991). Agreement between subordinate and self-ratings in upward feedback. Personnel Psychology, 44, 375 - 390.
- Mabe, P. & West, S. (1982). Validity of self-evaluation of ability: A review and meta-analysis. Journal of Applied Psychology, 67, 280 - 286.
- Mercer's fax facts survey results: Multisource assessment. (Available from William M. Mercer, Inc., 1166 Avenue of the Americas, New York, NY 10036)
- Podsakoff, P. & Organ, D. (1986). Self-reports in organizational research: Problems & prospects. Journal of Management, 12, 531 - 544.
- Van Velsor, E., Taylor, S., Leslie, J. (1993). An examination of the relationship among self-perception accuracy, self-awareness, gender, and leader effectiveness. Human Resource Management, 32, 249 - 264.
- Waldman, D. A., Atwater, L. E. & Antonioni, D. (1998). Has 360 degree feedback gone amok? Academy of Management Executive, 12, 86 - 94.
- Wicklund, R. (1975). Objective self-awareness. In Berkowitz L (Ed.), *Advances in experimental social psychology* (Vol. 8, pp. 233-275). New York: Academic Press.
- Wicklund, R. (1978). Three years later. In Berkowitz L (Ed.), *Cognitive theories in social psychology* (pp. 509-521). New York: Academic Press.
- Wicklund, R. (1979). The influence of self on human behavior. American Scientist, 67, 187-193.

Wohlers, A., & London, M. (1989). Ratings of managerial characteristics: Evaluation difficulty, coworker agreement, and self- awareness. Personnel Psychology, 42, 235 - 261.

Yammarino, F. J., Atwater, L. E. (1997). Do managers see themselves as others see them? Implications of self-other ratings agreement for human resources management.

Table 1

Intercorrelations and Descriptive Statistics of Variables

	Mean	SD	Alpha+	Self	Subordinate	Supervisor	Peer	Performance
Self rating	6.10	0.55	0.97	----	----	----	----	----
Subordinate	6.12	0.98	0.99	0.02	----	----	----	----
Supervisor	6.19	0.77	0.98	-0.03	-0.03	----	----	----
Peer	5.91	1.25	0.99	-0.08	0.09	0.12*	----	----
Performance	7.74	2.58	----	0.07**	0.11	0.06	0.03	----

\*  $p < .05$ \*\*  $p < .01$ 

+ Cronbach's Alpha

Table 2

Regression of Promotion Scores on Self- and Subordinate Ratings

	<u>Model 1</u>	<u>Model 2</u>
	Betas	Betas
Self rating	0.08	1.31
Subordinate rating	0.13*	2.23*
Self X self		-0.01
Subordinate X subordinate		-0.13
Self X subordinate		-2.34*
R	0.16	0.24
R <sup>2</sup>	0.03	0.06
F	3.26*	3.02*
ΔR <sup>2</sup>		0.03
F <sub>Δ</sub>		2.80*

\* p &lt; .05

Table 3

Regression of Promotion Scores on Self- and Peer Ratings

	<u>Model 1</u>	<u>Model 2</u>
	Betas	Betas
Self rating	0.13*	1.46
Peer rating	0.05	1.00
Self X self		-0.99
Peer X peer		-0.03
Self X peer		-0.96
R	0.13	0.15
R <sup>2</sup>	0.02	0.02
F	2.42	1.29
ΔR <sup>2</sup>		0.00
F <sub>Δ</sub>		0.54

\* p &lt; .05

Table 4

Regression of Promotion Scores on Self- and Supervisor Ratings

	<u>Model 1</u>	<u>Model 2</u>
	Betas	Betas
Self rating	0.08*	0.07
Supervisor rating	0.06	0.29
Self X self		0.24
Supervisor X supervisor		-0.36
Self X supervisor		0.07
R	0.10	0.11
R <sup>2</sup>	0.01	0.01
F	3.95*	2.00
ΔR <sup>2</sup>		0.00
F <sub>Δ</sub>		0.71

\* p &lt; .05

Figure Caption

Figure 1. Contour plot of promotion score predicted from subordinate-ratings and self-ratings.

