

**C**

**E**



**Center for  
Effective  
Organizations**

---

**ASSESSING THE FACTORS INFLUENCING  
DIFFERENCES BETWEEN SUPERVISOR  
AND SUBORDINATE PERFORMANCE  
RATINGS: A MULTIPLE SAMPLE STUDY**

**CEO PUBLICATION  
G 95-8 (284)**

**K. KACMAR  
FLORIDA STATE UNIVERSITY**

**D. CARLSON  
UNIVERSITY OF UTAH**

**P. WRIGHT  
TEXAS A&M UNIVERSITY**

**G. McMAHAN  
UNIVERSITY OF SOUTHERN CALIFORNIA**

**June, 1995**

# **Assessing the Factors Influencing Differences Between Supervisor and Subordinate Performance Ratings: A Multiple Sample Study**

## **ABSTRACT**

The 360 degree feedback technique includes, among others, performance ratings by self and supervisor. However, past research suggests that these two types of ratings may not be comparable for a variety of reasons. For example, individuals tend to inflate self-ratings thus decreasing the correlation between these ratings and ratings provided by supervisors. This inflation can be even greater in older individuals and those with high self-esteem. Other research suggests that the type of job held (i.e., professional versus blue-collar) also will influence the correlation between supervisor and subordinate ratings of performance. Specifically, professionals will rate themselves higher than their supervisors more often than will blue-collar workers. All three of these contentions were tested using two different samples, one of professional employees (N=203) and one of blue-collar workers (N=233). Results indicated support for all but one test. That is, individuals rated themselves higher than did their bosses with this effect being intensified with older workers as well as those with high self-esteem. The one test that did not find support in the present study was the type of job held influencing the ratings. While past research would expect the correlation between supervisor and self ratings to be higher for the professional sample, the exact opposite was found. Rationale for each of these findings are discussed and future research directions are suggested.

## **Assessing the Factors Influencing Differences Between Supervisor and Subordinate Performance Ratings: A Multiple Sample Study**

The use of 360-degree feedback is increasing in organizations in order to allow employees to compare their perceptions of the situation with the perceptions of others (Nowack, 1993). This approach to assessment is advantageous in that it can be used in a number of different human resource situations and provides for a more comprehensive evaluation of the employee. However, the use of multiple raters may result in inconsistent results and the issue of how to explain these differences arises. In fact, past research has suggested that there is frequently a lack of agreement between self-rating and supervisor-ratings (Ferris, Yates, Gilmore, & Rowland, 1985; Harris & Schaubroeck, 1988; Klimoski & London, 1974; Mabe & West, 1982; Thornton, 1980). Given these findings, the present research was designed to isolate factors which influence the degree to which supervisor and subordinate performance appraisal ratings differ.

### **Self versus Supervisor Ratings**

Individuals in organizations, although not formally, self-assess their performance regularly. Some individuals generate "to-do" lists and feel good about their performance when they are able to cross every item off of the list by the end of the day. Others simply compare their productivity to those around them. In most instances, the assessor comes out on top. Reasons for the positive evaluation, which are many, can be found by examining the fundamentals of equity theory (Adams, 1963). For example, the person may have selected only poor performers as comparison others. Further, the information from which the comparer is working may be inaccurate or biased. It also may be the case that an individual will make external attributions about his or her own poor performances (e.g., the situation was a no-win one) but attribute a poor performance by the comparison other to internal reasons (e.g., did not possess the abilities needed) (De Vader, Bateson, & Lord, 1986; Harris & Schaubroeck, 1988; Jones, & Nisbett, 1972). Whatever the reason, the outcomes are the same: individuals generally overrate their own performance. Hence,

it is not difficult to predict what will happen when individuals are asked to evaluate their performance on a formal basis.

In order to determine when an individual is overvaluing his or her performance, a true measure of performance must be found. While this value rarely exists, the performance measure most frequently used as the truth is the performance rating of the supervisor. Past research has found that the relationship between an individual's rating of himself or herself and the supervisor's rating of this individual rarely produce more than a moderate correlation (Harris & Schaubroeck, 1988; Mabe & West, 1982). Hence, not only have self-ratings been found to be inflated, the correlations between self and supervisor ratings have not been strong (Fahr & Dobbins, 1989; Ferris et al., 1985; Thornton, 1980; Williams & Levy, 1992). For example, one of the major findings reported by Williams and Levy (1992) was that self-appraisals were affected by a leniency bias and the correlations between self and supervisor ratings were moderate and positive. Further, Ferris et al. noted that the tendency to rate one's performance more highly than what might be deemed accurate was moderated by age. That is, older workers rated themselves higher than did their supervisors, however, younger workers did not. Finally, Fahr and Dobbins (1989) reported that individuals with high self-esteem were more lenient raters of themselves than were their low self-esteem counterparts.

Another significant finding reported by Fahr and Dobbins (1989) was that when performance dimensions were more clearly defined, less leniency was found in the ratings. However, along these same lines, Harris and Schaubroeck (1988) reported differing results. Specifically, they noted that ratings scales (trait versus behavioral) and formats (dimensional versus global) had little impact on the relationship between self and supervisor ratings. The variable they viewed as a potential moderator was job type (professional versus blue collar).

While some of results regarding self versus supervisor evaluations are fairly clear (e.g., supervisors rate subordinates lower than subordinates rate themselves), potential moderators or factors that influence these differences have not been as clearly delineated. Thus, the focus of the

present paper is to address this issue. To begin, the data will be examined to verify that a leniency effect for self ratings was found. We define leniency in this study as significantly different (i.e., higher) than the midpoint on the rating scale. This examination will be performed across two different samples, one of professionals (sample 1) and one of blue-collar workers (sample 2). If Harris and Schaubroeck's (1988) finding, that the type of job influences the ratings employees awarded themselves, is replicated, then the correlations between self and supervisor ratings will be lower for the professional sample than for the blue-collar sample. Hence,

**Hypothesis 1:** Respondents from the blue-collar sample will rate themselves more similarly to their supervisors (i.e., have higher correlations between self and supervisors ratings) than will respondents from the professional sample.

Further, if an overall leniency effect is found for either or both samples, then an examination of possible reasons for this finding will be explored. Specifically, building upon past work, age and self-esteem will be examined as potential influences of this effect (Fahr & Dobbins, 1989; Ferris et al., 1985). If past research is found to hold, then older individuals and individuals with higher self-esteem will rate themselves higher than their respective counterparts. Thus,

**Hypothesis 2:** Older respondents will rate themselves higher than will younger respondents.

**Hypothesis 3:** Respondents with higher self-esteem will rate themselves higher than will respondents with lower self-esteem.

## **Method**

### **Sample 1**

As part of a larger research project, self ratings of performance were collected from 203 white-collar workers for a state agency in a southeastern state. In addition, each of these respondents' immediate supervisors also rated each respondents' performance. There were a total of 59 supervisors, and each rated no more than 5 subordinates (range 1-5;  $M = 2.96$ ). The demographic characteristics of the subordinates include an average age of 46.23 years (range 24-

66 years), 102 (50%) males, an average organizational tenure of 11.9 years, average tenure with supervisor of 7.5 years, and an average tenure in the current job of 8.6 years. The supervisors had an average age of 42.07 years (range 28-66 years) and 37 (63%) were male. The average organizational tenure for the supervisors was 11.4 years and the average tenure in the current job was 3.4 years. The racial composition of those responding (N=197), included 148 (73%) white, 29 (14%) were African-American, 15 (7%) were Hispanic and 1% who indicated either Asian and American Indian. Of the supervisors, 39 (66%) were white, 12 (20%) were African-American, 6 (10%) were Hispanic, and 3% indicated they were Asian.

### **Procedure**

During an off-site workshop for managers of the state agency, the 59 supervisors in attendance were asked to complete a survey about up to five of their direct reports. Specifically, each supervisor was given five blank surveys. The first step was to have the supervisor fill in the name of the subordinate that he or she was thinking about while completing the survey. One of the measures in the survey was a performance appraisal form. Once we received the completed surveys from the supervisors, we mailed each of the subordinates evaluated by the supervisor a survey and asked him or her to complete it. This resulted in 203 matched pairs of supervisors and subordinates.

## **Measures**

**Self-appraisal.** Respondents were asked to rate themselves on four dimensions: overall effectiveness, performance in required roles, performing as your supervisor wants, and overall level of performance. The ratings were made on a 5-point Likert type scale with the anchor for 1 being strongly disagree and the anchor for 5 being strongly agree. The overall internal consistency reliability estimate for these four items was .86.

**Supervisor appraisal.** The immediate supervisor for each subordinate was asked to rate the subordinate on the same four dimensions noted above. The same scale was used. The internal consistency reliability estimate was .91.

## **Analyses**

In order to determine the level of agreement between the two raters, correlations between the supervisors' and subordinates' rating for each item and the average for the overall scale were calculated. Further, a comparison of the means for each item and the overall scale also were made.

## **Sample 2**

As part of a needs assessment program for a large warehouse and shipping operation, subordinates and their immediate supervisors completed performance ratings. There were a total of 242 subordinates and 15 supervisors participating. There were 135 (56%) males and 107 (44%) females in the subordinate sample. The majority (203 - 84%) were white with 38 (16%) being African American and 1 indicating a race of Asian. The subordinate sample had an average age of 49 years (range 22-81 years) and an organizational tenure of 16 years (range 4-44 years). Of the 15 supervisors in the sample, 14 (93%) were male and 1 (7%) was female. The average age for the supervisors was 48 years (range 32-65 years) with an average tenure with the organization of 20 years (range 5-35 years). Three (20%) of the supervisors were African-American while the remaining 12 (80%) were white.

## **Procedure**

Prior to automating the work performed by the employees in the warehouse, a needs assessment was performed to determine if the workers possessed the skills required to use the technology to perform the new jobs, or if training would be required. As part of the assessment, the employees were asked to appraise their performance in their current jobs. This appraisal asked each worker to rate himself or herself on 10 items such as initiative and flexibility. Supervisors were given the same form to complete for each individual they supervised. This procedure resulted in 233 matched pairs of ratings.

## **Measures**

**Self-appraisal.** Respondents were asked to rate themselves on ten work related behaviors: motivation, initiative, enthusiasm, analytical ability, timeliness, quality of work, strong work habits, little need for supervision, trustworthy, and gets along with others (Wright, Kacmar, McMahan, & DeLeeuw, in press). Each rating was made on a 5-point Likert type scale with the anchor for 1 being strongly disagree and strongly agree for 5. The overall internal consistency estimate for these 10 items was .92.

**Supervisor appraisal.** Each subordinate who completed a self-appraisal also was evaluated by his or her immediate supervisor. The supervisors responded to the same 10 items denoted above, however, the scale used was a 7-point Likert-type scale with the anchor for 1 strong disagree and the anchor for 7 strongly agree. The Cronbach alpha internal consistency estimate was .92.

## **Analyses**

As in sample 1, interrater agreement was examined with correlations. That is, each item and the overall scale were correlated. Further, the individual means for the items and the overall scale were compared. Because the ratings to be compared and correlated were made on two different scales, one with anchors of 1 to 5 and one with anchors of 1 to 7, prior to any analyses being performed the two scales had to be made comparable. To do this, the items responded to on



a 5-point scale were subtracted from the midpoint and then divided by the high anchor (i.e.,  $(N-3)/5$ ). Similarly, for the 7-point scale, each item was subtracted from the midpoint on the scale and then divided by the high anchor (i.e.,  $(N-4)/7$ ). This transformation was consistent with our definition of leniency (i.e., significantly higher than the midpoint), and allowed for comparisons to be made between the supervisor and subordinate ratings. The range for the transformed scales is -.4 to +.4.

## Results

### Sample 1

**Overall effectiveness.** The correlation between the supervisors' and subordinates' ratings for the first item which tapped overall effectiveness was .22 ( $p < .05$ ). However, while the correlation was statistically significant, the agreement indicated by this figure is not high. An examination of the means indicated that the average rating of the subordinates of themselves was 6.0 while the supervisors rating was a more conservative 5.3. Further, these two means were significantly different from one another ( $t(159)=-7.22, p < .01$ ).

**Performance in required roles.** Subordinate ratings for performing required roles also indicated that they felt their performance was better than did their supervisors ( $M = 6.02$  vs  $5.46$ ;  $t(200)=-7.61, p < .01$ ). The correlation between these two average ratings was .24 ( $p < .01$ ). Once again, while the correlation was significant, the agreement rate represented was only moderate.

**Performing the way the supervisor wants.** The pattern denoted above was repeated for performing as your supervisor wants. Specifically, the correlation between the average ratings for this dimension of performance was .27 ( $p < .01$ ). The average scores were 5.80 for the subordinates and 5.20 their supervisors ( $t(198)= -6.67, p < .01$ ) again indicating that the incumbents felt they performed better than did their supervisors.

**Overall level of performance.** For the final performance dimension, overall level of performance, the correlation between the average ratings was once again statistically significant ( $r = .23, p < .01$ ), but lower than expected. The average ratings for the subordinates on this dimension was 5.80 while the supervisors recorded an average score of 5.22 ( $t(199)=-7.30, p < .01$ ).

**Overall Scale.** The same analyses were performed on the overall scale (i.e., the four items averaged together). The correlation between these two scales was .298 ( $p < .01$ ), once again showing a moderate rate of agreement. The averages for these two scales were 5.90 for the subordinates ratings and 5.33 for the supervisors. These figures were statistically different from one another ( $t(157)=-7.05, p < .01$ ).

## **Sample 2**

**Motivation.** The correlation between the supervisors' and subordinates' ratings for the first item which measured motivation was .124 ( $p = .058$ ) indicating that the correlation was not significant. The individual ratings for this item were .19 for the employee and .11 for the supervisor. These two means were significantly different from one another ( $t(232)=4.91, p < .000$ ).

**Initiative.** The average rating for the subordinates by the supervisor on initiative were .20 while the subordinates rated themselves an average score of .22. These two values were significantly different from one another ( $t(232)=5.42, p < .000$ ), and the correlation between these values was significant ( $r=.180, p < .01$ ), but not high.

**Enthusiasm.** The ratings provided by the subordinates for enthusiasm indicated that they felt they performed better on this dimension than did their supervisors (.16 vs. .03,  $t(232)=7.21, p < .000$ ). Although significant, the correlation for these two ratings was not extremely high ( $r=.183, p < .01$ ).

**Analytical ability.** The pattern denoted above for dimensions 2 and 3 was repeated for dimension 4, analytical ability. Specifically, the correlation between the two rating was

significant ( $r=.141$ ,  $p < .05$ ) and the means were significantly different (.14 for subordinates and .02 for supervisors,  $t(231)=7.08$ ,  $p < .000$ ).

**Timeliness.** The pattern of results changed for the dimension of timeliness. In this case, the ratings were not significantly different from one another (.11 vs. .11,  $t(226)=-.22$ , ns) and the correlation was not significant ( $r=.081$ , ns).

**Quality of work.** The correlation for the ratings for quality of work also was not significant ( $r=.07$ , ns). Similarly, the mean ratings were not significantly different either (.14 vs. .13,  $t(229)=-.30$ , ns).

**Work habits.** The mean ratings for strong work habits followed the same pattern found in dimensions 5 and 6. That is, they were not significantly different from one another (.09 vs. .08,  $t(228)=.51$ , ns). However, the correlation between these two ratings was significant ( $r=.224$ ,  $p < .01$ ).

**Need for supervision.** Subordinates judged their need for supervision to be much lower than did their immediate supervisors (.21 vs. .09,  $t(229)=6.75$ ,  $p < .000$ ). The correlation between these two ratings also were significant ( $r=.234$ ,  $p < .000$ ).

**Trustworthy.** For the dimension of trustworthiness, once again supervisors rated their subordinates lower than subordinates rated themselves (.21 vs. .03,  $t(229)=9.70$ ,  $p < .000$ ). Further, the correlation between these two values also was significant ( $r=.224$ ,  $p < .01$ ).

**Get along with others.** On the final dimension, getting along with others, once again both the correlation and the mean differences were significant ( $r=.191$ ,  $p < .01$ ; .19 vs. .11,  $t(231)=4.81$ ,  $p < .000$ ). These results indicate that the incumbents viewed their relations with others as more positive than did their supervisors.

**Overall Scale.** The same analyses were performed on the overall scale (i.e., the 10 items averaged together). The correlation between these two scales was .26 ( $p < .000$ ), once again showing a moderate rate of agreement. The averages for these two scales were .16 for the

subordinate ratings and .08 for the supervisors. These means were statistically different from one another ( $t(217)=6.68, p < .000$ ).

### **Leniency effect**

A leniency effect can be found by examining the differences between the ratings provided by individuals and their supervisors (Williams & Levy, 1992). In all cases, the supervisors rated the subordinates lower than they rated themselves. The mean discrepancy between self and supervisor ratings for the professional sample was .64 and the mean difference between the rating for the blue collar sample was 1.31. While three of the differences for the blue collar sample were not significantly different, they still were in the direction expected for a leniency effect. These results confirm that a leniency effect was found. Hence, additional tests can be performed to help isolate possible reasons for these differences.

### **Hypothesis 1: Comparison of Sample 1 and 2**

One goal of this study was to assess whether the contention made by Harris and Schaubroeck (1988), that correlations between self and supervisor ratings of professional individuals would be lower than correlations between self and supervisor ratings of blue-collar workers, was accurate. While the actual appraisal forms were not identical, they were comparable. Results indicated little support for this notion. First, on the whole, the correlations for the individual items were larger for the professional sample than the blue collar workers. Further, no support for Harris and Schaubroeck's findings were found in a comparison of the overall scale correlations. Once again, the correlation for the self and supervisor ratings for the professional sample was higher than the correlation for the blue-collar sample. Thus, no support was found for hypothesis 1.

### **Hypothesis 2: Age**

A correlation between age and self appraisal ratings was calculated for each sample. Results indicated that for the professional sample the correlation was positive and significant ( $r=.16, p < .05$ ) suggesting that as employees grow older, their self-performance ratings increase.

This age-performance link was not found for the supervisors' ratings. Specifically, the correlation between subordinate age and supervisor rating was not significant ( $r = -.007$ , ns). Further, for the blue collar sample, the correlation between age and self-ratings and age and supervisor ratings were both negative and nonsignificant ( $r = -.08$ , ns for self and  $r = -.006$ , ns for supervisor). Hence, hypothesis 2 received support from the professional sample, but not the blue-collar one.

### **Hypothesis 3: Self-esteem**

Self-esteem of the employee was only measured in the blue collar sample. The correlation between self-rating and self-esteem was  $.17$  ( $p < .01$ ). This denotes that for the blue collar sample, as self-esteem increases, so did self-ratings, providing support for hypothesis 3.

### **Discussion**

Several major findings which are consistent with past research emerged from the current study. First, a leniency effect, in that for each dimension the subordinates rated themselves higher, and in most cases significantly higher, than did their supervisors, was detected (Williams & Levy, 1992). While this finding is not new to the literature, it does reinforce earlier reports of similar results (Fahr & Dobbins, 1989; Williams & Levy, 1992). What these results do not tell the reader is why leniency errors occur. Several reasons are possible. For example, individuals may only recall positive events when rating their own performance. Blocking painful and negative events is a common coping mechanism (Averill & Rosenn, 1972). Another explanation may be that the standards used to evaluate oneself differ from those used by the supervisor. Understanding the standards applied by both raters may help to better understand the differences in the final ratings. Future research should attempt to explore these reasons and others to shed light on the issue of why individuals inflate their own performance ratings when compared to ratings provided by their supervisors.

Additionally, the agreement rate between the ratings provided by these two samples was only moderate as is consistent with past research (Fahr & Dobbins, 1989; Harris & Schaubroeck, 1988; Mabe & West, 1982). One of the arguments used to explain the leniency bias found in the

present research can be employed to explain these findings as well. That is, the performance criteria used by each set of raters may differ (Harris & Schaubroeck, 1988) causing a difference in ratings on both the individual dimensions and the overall ratings. This explanation also can be extended to note that even if the same criterion are used, the weights applied to each may differ (Klimoski & London, 1974). Only future research can verify if this explanation holds.

Finally, age in the professional sample and self-esteem in the blue collar sample had a significant and positive relationship with self appraisals as predicted (Fahr & Dobbins, 1989; Ferris et al., 1985). With respect to age, it may be that individuals are equating life learned wisdom with performance. For example, older individuals may view mistakes made with clients as learning experiences that will never be repeated, while younger employees view them as poor performance stemming from a lack of experience. Further, older employees may have developed their own techniques which substantially reduce the time needed to perform the job and hence view themselves as productive. However, since a positive correlation between age and self appraisal was not found for the blue collar sample, future research is needed to verify if these explanations are valid, or if there are other more appropriate explanations for these findings.

Since the current findings for self esteem mirror those of Fahr and Dobbins (1989), their explanations should be applicable to the present study. These authors suggested that individuals with high self esteem have a biased perception of their own behavior which inflates their ratings. They further noted that it may be that individuals with high self-esteem simply are unaffected by negative feedback, thus never lowering their opinions of themselves when faced with situations in which their performance was poor. Either of these explanations seem appropriate for the present results.

There also was one finding in the present study that differed from past findings. That is, the correlations between self and supervisor ratings for the professional sample were higher than those for the blue collar sample. This finding is exactly opposite of that which was predicted based upon past research (Harris & Schaubroeck, 1988). One explanation for this unexpected

finding may be the difference in the familiarity the respondents had with performance appraisals. The respondents in the professional sample were evaluated by their supervisors at least once a year as is common practice in this type of job. However, the blue-collar workers were unionized and never received a performance evaluation from their supervisors. Hence, the individuals in the professional sample may have been more aware of how their supervisors rated them due to the previous appraisals they had received.

### **Conclusion**

All of these findings suggest that the ratings from these two sets of raters consistently differ enough to make the reconciliation of the ratings for a 360 degree feedback exercise difficult. For example, higher ratings would be expected from self ratings as well as from older or more self-assured workers. These ratings may have to be tempered to more closely align with other ratings provided. Because the ratings provided by these individuals reflect their view of the reality of the situation, explanations given for why these values have to be modified must be clear and thoughtful. Further, objective measures of performance also must be incorporated in order to offset any subjective bias that may exist. While objective measures may be difficult to develop, it is imperative that attempts to do so be made.

As in any empirical study, there are limitations which must be noted. First, the performance appraisal forms used in the two samples were not identical. Since they were not, care needs to be taken when interpreting the comparisons made. However, these comparisons are not completely without merit. While the actual items in the performance appraisals differed, the concept of congruence between self and supervisor ratings was constant, and it was this concept that was of interest in this study, not the individual content of the items.

A second limitation is that only subjective ratings were used. That is, both the subordinates and supervisors simply reported their perceptions of performance. Obviously, a much stronger test would have included an objective measure of performance and compared the self and supervisor ratings to this value. While this may never be possible in the professional

sample due to the type of jobs performed, objective performance data was recently collected for the blue collar workers, and will be incorporated in future studies.

There also are strengths in this study which need to be mentioned. First, the two samples used in the current study were larger than many of the samples used in past research in this area (e.g., Fahr, Werbel, & Bedeian, 1988; London & Wohlers, 1991; Williams & Levy, 1992). Another advantage of the samples used was that they came from two different industries and types of workers, allowing us to explore a variety of issues with respect to the performance appraisal.

While the results presented are interesting and informative, there is still much to be learned in this area. First, it is important to incorporate objective measures of performance in future studies. Although supervisor ratings are often used as the "true" value of one's performance, this procedure is not without its limitations (Ashford, 1989). Hence, future studies should attempt to replicate the present findings by also including a comparison to a more objective measure of performance. Another study that needs to follow the present one is one that addresses why professional self ratings should be more highly correlated with supervisors ratings than blue collar worker self ratings. It could be argued that the tasks performed by blue collar workers may be more conducive to providing feedback to the workers than would professional jobs (Hackman & Oldham, 1975). If this were indeed the case, then both workers and supervisors could be trained to focus on the objective feedback provided by the job prior to evaluating performance. This type of training may help to improve the correlations between the two ratings.

In conclusion, this study provided confirmation of several findings previously presented in the literature. Given that virtually all research papers end with a request for replication, this is no small contribution. In addition to verifying past results, it is hoped that this paper also stimulated thought for future research in this area.



## References

- Adams, J.S. (1963). Toward an understanding of inequity. *Journal of Abnormal and Social Psychology*, 67, 422-436.
- Ashford, S.J. (1989). Self-assessments in organizations: A literature review and integrative model. In L.L. Cummings & B.M. Staw (Eds.), *Research in organizational behavior*, (Vol. 11, pp. 133-174). Greenwich, CT: JAI Press.
- Averill, J.R., & Rose, M. (1972). Vigilant and nonvigilant coping strategies and psychophysiological stress reactions during the anticipation of electric shock. *Journal of Personality and Social Psychology*, 23, 128-141.
- De Vader, C.L., Bateson, A.G., & Lord, R.G. (1986). Attribution theory: A meta-analysis of attribution hypotheses. In E. Locke (Ed.), *Generalizing from laboratory to field settings*, (pp. 63-81). Lexington, MA: Lexington Press.
- Fahr, J.L., & Dobbins, G.H. (1989). Effects of self-esteem on leniency bias in self-reports of performance: A structural equation model analysis. *Personnel Psychology*, 42, 835-850.
- Fahr, J.L., Werbel, J.D., & Bedeian, A.G. (1988). An empirical investigation of self-appraisal-based performance evaluation. *Personnel Psychology*, 41, 141-156.
- Ferris, G. R. Yates, V. L. Gilmore, D. C. & Rowland, K. M. (1985). The influence of subordinate age on performance ratings and causal attributions. *Personnel Psychology*, 38, 545-557.
- Hackman, J.R., & Oldham, G.R. (1975). Development of the Job Diagnostic Survey. *Journal of Applied Psychology*, 60, 159-170.
- Harris, M. M. & Schaubroeck, J. (1988). A meta-analysis of self-supervisor, self-peer, and peer-supervisor ratings. *Personnel Psychology*, 41, 43-62.

- Jones, E.E., & Nisbett, R.E. (1972). The actor and the observer: Divergent perceptions of the causes of behavior. In E.E. Jones, D.A. Kanouse, H.H. Kelly, R.E. Nisbett, S. Valines, B. Weiner (Eds.) *Attribution: Perceiving the causes of behavior*, (pp. 79-94). Morristown, NJ: General Learning Press.
- Klimoski, R.J., & London, M. (1974). Role of rater in performance appraisal. *Journal of Applied Psychology*, 59, 445-451.
- London, M., & Wohlers, A.J. (1991). Agreement between subordinate and self-ratings in upward feedback. *Personnel Psychology*, 44, 375-390.
- Mabe, P. A. & West, S. G. (1982). Validity of self-evaluation of ability: A review and meta-analysis. *Journal of Applied Psychology*, 67, 280-296.
- Nowack, K. M. (1993). 360-degree feedback: The whole story. *Training & Development*, 47, 69-72.
- Thornton, G. C. (1980). Psychometric properties of self-appraisals of job performance. *Personnel Psychology*, 33, 263-271.
- Williams, J.R., & Levy, P.E. (1992). The effects of perceived system knowledge on the agreement between self-ratings and supervisor ratings. *Personnel Psychology*, 45, 835-847.
- Wright, P.M., Kacmar, K.M., McMahan, G.C., & DeLeeuw, K. (in press). Ability as a moderator of the relationship between personality and job performance. *Journal of Management*.