

Effects Of Peer Ratings On Supervisor Ratings Of Job Performance

Ted Shore, Ph.D., California State University, San Marcos, USA
Armen Tashchian, Ph.D., Kennesaw State University, USA
William R. Forrester, Ph.D., Kennesaw State University, USA

ABSTRACT

The influence of peer rating information on supervisor performance ratings was investigated using a 2 x 2 x 2 crossed factorial experimental design. Participants were undergraduate business students who assumed the role of “supervisor” and evaluated a fictitious “subordinate” whose performance was either good or poor. Participants were given fictitious information about peer ratings (high or low) from either one or three co-workers. As expected, peer rating information resulted in inflated performance ratings for poor performers and deflated ratings for good performers. No differences in performance ratings were found for a single versus multiple peer ratings. Contrary to expectations, rater personality characteristics (conscientiousness and agreeableness) did not moderate the influence of peer rating information on performance ratings.

Keywords: Peer Rating; Supervisor Ratings; Job Performance

INTRODUCTION

The use of multi-source performance appraisal systems that employ two or more sources of ratings of the same job incumbent have gained in popularity in recent years (cf. Beehr, T.A., Ivanitskaya, L., Hansen, C.P., Erofeev, D., & Gudanowski, D.M, 2001; London & Smither, 1995; Timmreck & Bracken, 1997). For example, 360 degree feedback systems employ ratings from supervisors, co-workers (peers), customers and the job incumbent (self-ratings). The rationale for this approach is that employee performance can be more accurately assessed and is more likely to improve when the employee receives feedback from multiple rating sources rather than only from their immediate supervisor or manager. Although initially used primarily for developmental purposes (London & Smither, 1995), the current trend is for multi-source ratings systems to be used in administrative decisions (Bohl, 1996; Greguras, Robie, Schleicher & Goff, 2003).

The widespread use of multi-source performance appraisal systems raises questions about the potential influence among the various rating sources. Ideally, performance ratings made by multiple sources should be independent of one another. Otherwise, there is little advantage of using multiple rating sources. For example, several studies have shown that when supervisors are given knowledge of (high) self-ratings prior to completing their own performance evaluation, the supervisor ratings are inflated (Klimoski & Inks, 1990; Shore & Tachshian, 2007). This occurs even in the absence of an expectation to provide feedback to the subordinate, suggesting that supervisors are influenced by self-rating information when conducting performance evaluations. Very little is known about the influence of peer ratings on supervisor ratings. Thus, the purpose of this study is to extend knowledge of multi-source rating influence by investigating the effects of peer rating information on supervisor performance ratings.

For several reasons we expect that peer ratings will have an influence upon supervisor performance evaluations. In many work situations co-workers have more frequent (i.e., daily) contact with one another than with their supervisor. This is particularly true when work processes require that individuals work as a team or when work procedures are interdependent. Since co-workers often have greater opportunity to observe (and experience the effects of) one another's performance than the supervisor, the supervisor is likely to rely on peer ratings as an information source when completing their performance evaluation.

Although it is possible that peer ratings are influenced by friendship bias, it might be argued that self-ratings are even more biased than peer ratings. That is, a job incumbent is likely to be more motivated to inflate a self-rating in order to maximize potential personal gains than to inflate a peer rating. In a resource-constrained environment (e.g., limited promotional opportunities, finite pool of money for raises), the motivation to inflate self-ratings would appear to be greater than for peer ratings. Since research has shown that self-rating information influences supervisor ratings (Klimoski & Inks, 1990; Shore & Tashchian, 2007), we expect peer rating information to also impact on supervisor ratings.

Thus, we expect that when raters (i.e., supervisors) are given knowledge of ratings made by peers (i.e., co-workers who evaluated the same subordinate), they will assign performance ratings consistent with the peer ratings. This should occur regardless of the level of actual job performance. That is, when raters are given knowledge that a poorly performing employee has been rated high by his/her peer(s), this should influence the raters to inflate their rating. Conversely, when raters are given knowledge that a good performer has been given a low rating by a peer(s), this should influence the raters to deflate their rating.

Research on conformity has the potential to provide insight into the effects of peer ratings on supervisor performance ratings. Pioneering studies by Ashe (1956) found that individuals often conform to the opinions of a group by making erroneous decisions in a line judgment task to a much greater extent than individuals working alone. The generality of the conformity effect has been demonstrated in recent studies in the U.S. and other countries (e.g., Amir, 1984; Neto, 1995; Nickolson, Cole & Rocklin, 1985) suggesting that conformity is a universal and robust phenomenon.

One explanation for conformity, called informational social influence, is that individuals look to the opinions of others to determine the correct behavior, particularly in novel or ambiguous situations. In the current study, participants (college students) were asked to evaluate the performance of another individual subsequent to being given knowledge of how either several peers or a single peer evaluated the same “subordinate”, creating the potential for information social influence. Unlike in the Asch paradigm where participants made judgments about line lengths (in which the correct answer was obvious), making judgments about job performance is a more subjective task which should increase the effects of information social influence.

Furthermore, consistent with informational social influence theory, we would expect multiple peer ratings to have a greater impact on supervisor ratings than a single peer rating. For example, if a group of peers all rate a co-worker high (or low), the supervisor should be likely to assume these ratings have greater validity (i.e., less bias) and be given more weight than a single peer rating.

Thus, we hypothesized that:

- H1a:** Performance ratings for a *poorly* performing subordinate will be more lenient when raters (i.e., supervisors) are given knowledge of high peer ratings than when peer ratings are low.
- H1b:** Performance ratings for a good performing subordinate will be lower when raters (i.e., supervisors) are given knowledge of low peer ratings than when peer ratings are high.
- H2:** Performance ratings will be more influenced by multiple peer ratings than by a single peer rating.

A final purpose of the current study is to examine how certain personality characteristics influence the degree to which performance ratings are influenced by peer ratings. One study reported that individuals high on agreeableness and low on conscientiousness were likely to provide lenient performance ratings (Bernardin, Cooke & Villanova, 2000). By extension, we would expect that individuals who are more agreeable (i.e., cooperative, sympathetic), less conscientiousness (i.e., less thorough and careful) to be particularly susceptible to normative social influence. For example, in the case of a *poor* performer, we would expect highly agreeable and less conscientious raters to be more likely to conform to high peer ratings and produce lenient (i.e., inflated) performance ratings.

Thus we hypothesized that:

- H3:** Agreeableness will moderate the relationship between peer rating information and performance ratings such that highly agreeable raters will demonstrate greater conformity with peer ratings than less agreeable raters.
- H4:** Conscientiousness will moderate the relationship between peer rating information and performance ratings such that highly conscientious raters will demonstrate less conformity with peer ratings than less agreeable raters.

METHOD

Participants

Participants were 120 men (47%) and women students enrolled in business courses at a regional state university in the West. The mean age of participants was 22 years and had an average of 2 years of work experience.

Design

A 2 x 2 x 2 crossed factorial design was employed. The independent variables were peer evaluation (low, high), and performance level (poor, good) and number of peer ratings (1 or 3). The dependent variable consisted of the participant's performance evaluation of the "subordinate".

Procedure

Data were gathered during class sessions. Approximately the same number of participants were randomly assigned to the 8 experimental conditions. Each participant was asked to assume the role of "supervisor" and evaluate the performance of another individual ("subordinate") described as an undergraduate student who purportedly completed a clerical task. The gender of the "subordinate" was held constant (male) so as not to introduce additional gender effects into the design. In actuality there was no "subordinate"; the clerical task the participant evaluated was fictitious.

The scenario utilized was similar to the one used in prior research on self-ratings (e.g., Klimoski & Inks; 1990; Shore & Tashchian, 2007). The participants were told that another student had completed a clerical task that involved looking up catalogue numbers and suggested retail prices for 40 items from a merchandise catalogue. In addition, the "subordinate" was to calculate a 15%-off sale price for each item. Participants were provided with an answer key to score the "subordinate's" work. After determining the number of incorrect responses, participants were given additional information about peer ratings (described below) and then completed their performance evaluation. Participants then completed a post-experiment questionnaire and were debriefed.

Peer Ratings

Prior to completing the performance evaluation of their "subordinate", participants were given fictitious information about how either one (male) or three other co-workers (two males, one female) rated that same "subordinate". In the low peer rating conditions, the peers assigned performance ratings of "2" (single rater condition) or "2", "2" and "3" (multiple rater condition) on a scale of "1" (extremely poor) to "9" (extremely good). In the high peer rating conditions, peers assigned performance ratings of "8" (single rater condition) or "7", "8" and "8" (multiple rater condition). In addition to assigning numerical ratings, peer raters also listed reasons to support their ratings. A sample reason was "he was mostly very accurate." or "he did well on a difficult task".

Task Performance

Two levels of task performance were depicted. In the good performance condition, the “subordinate” made 5 errors out of 40 total items, an accuracy rate of 88%. In the poor performance condition, the “subordinate” made 12 errors out of 40 total items, an accuracy rate of 70%.

Dependent Measure

Subsequent to reviewing the peer rating information, each participant completed a performance evaluation form on which he/she rated their “subordinate’s performance on a scale of “1” (extremely poor) to “9” (extremely good).

Personality Measures

Participants completed measures of conscientiousness and agreeableness (9 items each) from the Big Five Inventory (John, Donahue & Kentle, 1991). A sample item from the agreeableness scale is “I see myself as someone who likes to cooperate with others”, and a sample item from the conscientious scale is “I see myself as someone who does a thorough job”. All measures used a scale ranging from 1 (strongly disagree) to 7 (strongly agree), and an average of the responses was computed for each measure. Coefficient alphas were .77 and .71 for conscientiousness and agreeableness, respectively.

RESULTS

Manipulation Checks

All manipulation check items were rated on a 7-point scale ranging from 1 (strongly disagree) to 7 (strongly agree). The stimulus materials were intended to depict two levels of task performance by the “subordinate”. The mean response to the item “My subordinate’s performance was above average,” was 3.40 for the poor performance conditions, and 5.30 for good performance conditions ($F=43.67$; $p<.001$). This suggests that participants clearly viewed “subordinate” performance as significantly better in the good than in the poor performance conditions. Another goal of the experimental manipulation was that participants would consider the peer rating information when completing their performance appraisals. Participants responses to the statement “I considered the peer evaluations(s) when evaluating my “subordinate” (mean=5.60), and to the statement “The peer evaluations had little influence on how I rated my “subordinate” (mean=3.20) suggests that participants did indeed utilize the peer rating information.

The means and standard deviations for all experimental conditions are shown in Table 1. ANOVA results are shown in Table 2. Contrary to expectations (Hypothesis 2), the main effect for the number of peer raters (one versus three) was not significant. Thus, results for the single and multiple peer rating conditions were combined for subsequent analyses. A main effect for peer rating information was significant ($F=61.65$; $p<.001$). Knowledge of high peer ratings resulted in higher performance appraisal ratings (mean=5.86) than knowledge of low peer ratings (mean=4.06). A main effect was also significant for task performance, such that ratings were higher in the good performance condition (mean=5.58) than in the poor performance condition (mean=4.24), which further attests to the success of the experimental manipulation for task performance.

Table 1: Mean Performance Rating for Experimental Conditions

	Task Performance			
	Poor Peer Ratings		Good Peer Ratings	
	Low	High	Low	High
Single Peer Rating	3.57 (1.09)	5.13 (1.32)	4.67 (1.30)	7.11 (0.76)
Multiple Peer Ratings	3.38 (0.72)	4.88 (1.20)	4.63 (1.26)	6.33 (1.35)

Note: Standard deviations are shown in parentheses

Table 2: Analysis of Variance for Dependent Measure

Source	df	F	Probability
Main Effects			
Peer Ratings	1	61.65	0.000
Task Performance	1	39.79	0.000
Number of Peers	1	1.90	0.171
2-way Interaction			
Peer Ratings x Task Performance	1	1.43	0.234
Number of Peers x Task Performance	1	.17	0.685
Number of Peers x Peer Ratings	1	.74	0.391
3-way Interaction			
Peer Ratings x Task Performance x Number of Peers	1	0.55	0.459

In support of Hypothesis 1a, in the poor performance condition, ratings assigned by raters given knowledge of high peer ratings were significantly higher (mean=5.00) than when (purported) peer ratings were low (mean=3.47; $F=22.89$; $p<.001$). In support of Hypothesis 1b, in the good performance condition, ratings assigned when raters were given knowledge of high peer ratings were significantly higher (mean=6.80) than when purported peer ratings were low (mean=4.64; $F=45.34$; $p<.001$).

We also hypothesized that more agreeable and less conscientious individuals would be more influenced by peer rating information (hypotheses 3 and 4). Moderated regression analyses showed that results for these hypotheses were not significant.

DISCUSSION

The purpose of this study was to investigate the influence of peer ratings on supervisor performance appraisals. Despite the increasing use of peer ratings in multi-source performance appraisal systems (e.g., Greguras, Robie, Schleicher & Goff, 2003), no prior empirical studies have evaluated the influence of peer ratings on supervisor performance ratings. We found that poor performers received significantly higher supervisor performance ratings when raters were told that either a single or multiple peers had rated the subordinate high than when peers rated the subordinate low. This finding parallels past research on the effects of self-ratings on supervisor performance ratings showing that high self-appraisals results in inflated supervisor ratings (Klimoski & Inks, 1990; Shore & Tashchian, 2007).

It is noteworthy that peer ratings also influenced supervisor ratings for good performers. Performance ratings were significantly lower when raters had knowledge that peers had purportedly assigned low ratings to the subordinate than when peer ratings were high. In fact, even a single low rating by a peer significantly reduced the performance ratings for good performers.

Past researchers have speculated that various political motives may cause a supervisor or manager to inflate or deflate performance ratings. Reasons for providing lenient ratings include avoiding conflict with a subordinate, helping the subordinate (e.g., gain a promotion) and maintaining a harmonious relationship with subordinates. Possible reasons for deflated performance ratings include building a record in preparation for terminating an

employee, setting of high standards, and sending a message that the employee needs to improve (Longenecker, et al., 1987).

The current study shows that the mere knowledge of peer ratings is sufficient to inflate or deflate supervisor ratings. This result is consistent with research showing that most people have a tendency to conform to majority group opinions even when the group opinion is obviously wrong (e.g., Amir, 1984; Asch, 1956). Unlike studies of conformity to group opinion, in our study raters were influenced by the opinion of even a single peer (co-worker) rating which was clearly discrepant from actual task performance. Our raters inflated performance ratings when the subordinate's performance was moderately poor (70% accuracy) when a peer assigned a high rating ("8"), and deflated ratings for good performance (88% accuracy) when a peer assigned a low rating ("2"). While leniency in performance ratings has been well documented (e.g., Murphy & Cleveland, 1995), rating deflation is less well understood. Unlike inflated (i.e., lenient) performance ratings which results in employees receiving greater rewards than deserved, when employees receive undeserved harsh ratings, they may be denied benefits they deserved and/or receive penalties not warranted.

We also expected that individuals who were more agreeable and less conscientious would be more likely to be influenced by peer ratings. Contrary to our expectations, neither of these personality characteristics were related to the influence of peer ratings. These expectations were based on very limited research showing a tendency for job performance ratings to be more lenient when made by more agreeable and less conscientious raters (Bernardin, et al., 2000).

A limitation of the current study is use of an undergraduate student sample. However, there is much precedent for the use of students in performance appraisal research (Murphy & Cleveland, 1995). Further, since a high percentage of our sample had full-time work experience, many participants could readily relate to the issues investigated in this study.

In summary, this is the first known empirical study to demonstrate that supervisor performance appraisals are distorted when raters are provided with peer (co-worker) ratings. Peer ratings were found to result in both inflated and deflated supervisor performance ratings. Further, these effects were similar when raters were provided with either a single or multiple peer ratings. Given the growing trend in the use of multi-source performance appraisals, it is important to understand the potential influence among rating sources.

AUTHOR INFORMATION

Ted Shore received his Ph.D. in industrial and organizational psychology from Colorado State University. He is Professor of Management at California State University, San Marcos, USA. His work has been published in a variety of journals including *Academy of Management Journal*, *Journal of Applied Psychology*, *Personnel Psychology*, *Journal of Organizational Behavior*, *Journal of Vocational Behavior*, and *Journal of Applied Social Psychology*. Ted Shore, Ph.D., Department of Management and Marketing, College of Business Administration, California State University, San Marcos, San Marcos, CA 92096-0001 USA. E-mail: tshore@csusm.edu.

Armen Tashchian is Professor of Marketing in the Coles College of Business at Kennesaw State University. He received his Ph.D. in Marketing from the University of Texas at Austin. He has published in numerous academic journals including the *Journal of Marketing*, the *Journal of Marketing Research*, and the *Journal of Academy of Marketing Science*. Armen Tashchian, Ph.D., Department of Marketing and Professional Sales, Coles College of Business, Kennesaw State University, 1000 Chastain Road, Kennesaw, GA 30144-5591 USA. E-mail: atashchi@kennesaw.edu.

William Forrester received his Ph.D. in Marketing in 1986 from the University of Tennessee. He currently works as Professor of Marketing in the Coles College of Business at Kennesaw State University. He has published in the *Journal of Academy of Marketing Science*, the *Journal of Economic Psychology*, and the *Journal of Applied Business Research*, and numerous other outlets. William R. Forrester, Ph.D., Department of Marketing and Professional Sales, Coles College of Business, Kennesaw State University, 1000 Chastain Road, Kennesaw, GA 30144-5591 USA. E-mail: wforrest@kennesaw.edu.

REFERENCES

1. Amir, T. (1984). The Asch conformity effect: A study in Kuwait. *Social Behavior and Personality*, 12, 187-190.
2. Asch, S.E. (1956). Studies of independence and conformity: A minority of one. *Psychological Monographs*, 70.
3. Beehr, T.A., Ivanitskaya, L., Hansen, C.P., Erofeev, D. & Gudanowski, D.M. (2001). Evaluation of 360 degree feedback ratings: Relationships with each other and with performance and selection predictors. *Journal of Organizational Behavior*, 22, 775-796.
4. Bernardin, H.H., Cooke, D.K., & Villanova, P. (2000). Conscientiousness and agreeableness as predictors of rating leniency. *Journal of Applied Psychology*, 85, 232-236.
5. Bohl, D.L. (1996). Mini-survey: 360-degree appraisals yield superior results, survey shows. *Compensation and Benefits Review*, 28, 16-20.
6. Greguras, G.J., Robie, C., Schleicher, D.J., & Goff, M. (2003). A field study of the effects of rating purpose on the quality of multi-source ratings. *Personnel Psychology*, 56, 1-21.
7. John, O.P., Donahue, E.M., & Kentle, R.L. (1991). The "Big Five" Inventory. Berkeley: University of California, Berkeley, Institute of Personality and Social Research.
8. Klimoski, R., & Inks, L. (1990). Accountability forces in performance appraisal. *Organizational Behavior and Human Decision Processes*, 45, 194-208.
9. London, M., & Smither, J.W. (1995). Can multi-source feedback change perceptions of goal accomplishment, self-evaluations, and performance-related outcomes? Theory-based applications and directions for research. *Personnel Psychology*, 48, 803-839.
10. Longenecker, C.O., Sims, H.P., & Gioia, D.A. (1987). Behind the mask: The politics of employee appraisal. *Academy of Management Executive*, 1, 183-193.
11. Murphy, K.R., & Cleveland, J.N. (1995). *Understanding performance appraisal: Social, organizational and goal-based perspectives*. Thousand Oaks, CA: Sage.
12. Neto, F. (1995). Conformity and independence revisited. *Social Behavior and Personality*, 23, 217-222.
13. Nicholson, N., Cole, S., & Rocklin, T. (1985). Conformity in the Asch situation: A comparison between contemporary British and U/S. university students. *British Journal of Social Psychology*, 24, 59-63.
14. Shore, T.H., & Tashchian, A. (2007). Effects of feedback accountability and self-rating information on employee appraisals: A replication and extension. *Psychological Reports*, 100, 1091-1100.
15. Timmreck, C.W., & Bracken, D.W. (1997). Multi-source feedback: A study of its use in decision-making. *Employment Relations Today*, 12, 86-94.

NOTES