1-2010

Understanding Performance Ratings: Dynamic Performance, Attributions, and Rating Purpose

Jochen REB
Singapore Management University, jochenreb@smu.edu.sg

Gary J. GREGURAS
Singapore Management University, garygreguras@smu.edu.sg

RUNNING HEAD: Dynamic Performance, Attributions, and Rating Purpose

UNDERSTANDING PERFORMANCE RATINGS:

DYNAMIC PERFORMANCE, ATTRIBUTIONS, AND RATING PURPOSE

Jochen Reb Gary J. Greguras

Singapore Management University

Correspondence concerning this article should be addressed to Jochen Reb, Lee Kong Chian School of Business, Singapore Management University, 50 Stamford Road, Singapore 178899. E-mail: jreb@smu.edu.sg. The authors would like to thank William Balzer for helpful comments on an earlier version of this manuscript. This research was funded in part by the Office of Research, Singapore Management University (Grant Number 08-C207-SMU-021).

ABSTRACT

The present two studies integrate and extend the literatures on dynamic performance, performance attributions, and rating purpose, making several important contributions. First, examining attributions of dynamic performance, Study 1 predicted and found that performance mean and trend affected judged ratee ability and effort, and performance variation affected locus of causality. Second, investigating the interaction between dynamic performance and rating purpose, Study 2 predicted and found that performance mean had a stronger impact on administrative than on developmental ratings, whereas performance trend and variation had a stronger impact on developmental than on administrative ratings. Third, both studies found that performance trend interacted with performance mean and variability to predict overall ratings. Fourth, both studies replicated main effects of dynamic performance characteristics on ratings in a different culture and more experienced sample of managers (Study 2).

KEYWORDS: Dynamic Performance, Performance Attributions; Performance Evaluation, Performance Ratings, Rating Purpose

Employee performance is dynamic and changes over time (Deadrick, Bennett, & Russell, 1997). These changes can be temporary, e.g., due to fluctuations in one's affective state (Beal, Weiss, Barros, & MacDermid, 2005) or permanent, e.g., due to learning (Deadrick et al., 1997). Considering both short-term and long-term changes in performance, employee performance profiles may differ on at least three performance characteristics: (a) mean, (b) trend, and (c) variation. When faced with the task of providing an overall performance rating for an appraisal period, raters must somehow integrate these dynamic features of performance into one summary evaluation. Unfortunately, little research on performance appraisal has considered the dynamic nature of performance (Reb & Greguras, 2008). To address this gap, we examine how raters interpret the information in dynamic performance profiles (Study 1), whether rating purpose interacts with dynamic performance characteristics (Study 2), and whether dynamic characteristics interact with each other to affect performance ratings (Study 1 & 2).

Study 1

Past research suggests that dynamic performance characteristics affect overall performance ratings. In particular, ratings are more favorable the higher the performance mean (Reb & Cropanzano, 2007) and the more upward sloping the performance trend (DeNisi & Stevens, 1981). Effects for performance variation typically have not been significant (e.g., Scott & Hamner, 1975). Given these findings, it is important to understand how raters make sense out of dynamic performance information. Fedor and Rowland (1989) point out that rater attributions may be especially useful in this respect. According to Campbell, McCloy, Oppler, and Sager's (1993) theory of performance, the determinants of employee performance are declarative and procedural knowledge and skills (which in part are a function of ability), motivation, and situational influences. Drawing on these distinctions, Study 1 examines how raters attribute

dynamic performance characteristics to employee ability, effort, and locus of causality (internal/external).

Ability: Influence of Performance Mean and Trend

A robust finding in the organizational sciences is that cognitive ability predicts employee performance across jobs (Hunter & Hunter, 1984). As such, attributions of ability are well suited to explain differences in employee mean performance. As expected, Porac, Ferris, and Fedor (1983) found that higher mean performance was associated with higher judged ability. Performance trend may also affect ability attributions. Specifically, over time employees with more experience are expected to perform higher because their experiences enable them to learn and develop skills (Campbell et al., 1993). As such, an upward trend may suggest that an employee is developing the ability to perform at a higher level. A flat or deteriorating trend, on the other hand, may indicate that an employee lacks the ability to learn or develop new skills.

H1a: Judged ability is greater the higher the performance mean.

H1b: Judged ability is higher for an improving trend than for a flat or deteriorating trend.

Effort: Influence of Performance Mean and Trend

Differences in performance often are attributed to effort (Russell, McAuley, & Tarico, 1987). Someone who consistently performs poorly (or favorably) may do so because of consistently low (or high) effort. Similarly, effort is a common explanation for systematic changes in behavior. When performance improves, it often is considered the result of effort (Dweck & Leggett, 1988). In contrast, when performance deteriorates, it may be attributed to a reduction of effort (Jones, Rock, Shaver, Goethals, & Ward, 1968).

H2a: Judged effort is greater the higher the performance mean.

H2b: Judged effort is highest for an improving trend, followed by a flat trend, followed by a deteriorating trend.

Locus of Causality: Influence of Performance Variation

Performance is determined both by internal and external factors and people's attributions reflect both loci of causality (Russell et al., 1987). Existing research indicates that consistent performance tends to be attributed to internal factors, whereas highly variable performance tends to be attributed to external factors (Fedor & Rowland, 1989). This could be because deviations from mean performance tend to be attributed to random external events (Slifkin & Newell, 1998). Thus we predict the following.

H3: Judged locus of causality is more external for large than for small performance variation.

Method

Participants

Fifty-four undergraduate students in Singapore participated in exchange for course credit. The majority of the participants were female (51.9%) with an average age of 21.0 years. Many participants indicated they had performed a peer (77.8%) or subordinate (48.1%) evaluation as part of their jobs at least once.

Design, Materials, and Manipulations

The experimental procedure and materials used in the current study were adapted from Reb and Cropanzano (2007). The experiment consisted of a 3 (performance mean: below average vs. average vs. above average) x 3 (performance trend: deteriorating vs. flat vs. improving) x 2 (performance variation: small vs. large) fully-crossed within-subjects design.

Participants assumed the role of "Regional Supervisor" whose task was to review and evaluate the junior-level sales personnel's performance over the past 26 weeks. The weekly performance data indicated how much money that salesperson contributed relative to a long-term company average. Following the instructions, the survey booklet contained the dynamic

performance profiles of 18 hypothetical employees (for examples, see Reb & Cropanzano, 2007). Each page in the booklet contained the performance profile of one employee on top and the dependent measures below. To minimize order effects, participants browsed through the entire booklet before starting with their evaluations.

Following the procedure of Reb and Cropanzano (2007), the profiles resulted from the within-subjects manipulation of the three dynamic performance characteristics of performance (i.e., mean, trend, and variation). Mean revenue contribution of an employee, relative to the company's long-term average for this job, was either negative $1,800, zero ($0), or positive $1,800. Performance variation was either small (standard deviation = $400) or large (standard deviation = $1200). To create the improving (deteriorating) trend conditions, $150 were added (subtracted) each week, while holding overall mean level constant.

Measures

Attributions. Attribution measures were adapted from Ronis, Hansen, and O'Leary (1983). All measures were taken on 7-point (1-7) scales. Judged ability was measured with "Over the 26 weeks, this employee showed [strong / poor] ability." We assessed judged effort with: "Over the 26 weeks, this employee invested [a lot of / little] effort." Finally, we assessed judged locus of causality with: "Over the 26 weeks, this employee's performance was determined largely by factors [inside / outside] his/her control."

Performance. For several supplementary analyses (see below), we measured overall employee performance with one item on a 7-point (1-7) scale: "Over the 26 weeks, this employee had [very poor / very good] overall performance."

Results and Discussion

Analyses consisted of a series of 3 (trend) x 3 (mean) x 2 (variation) repeated-measures

ANOVAs on the different dependent variables. All pairwise comparisons reported below are significant, $p < .05$, unless otherwise noted.

Attributions of Dynamic Performance Characteristics

Ability. Consistent with Hypothesis 1a, mean performance influenced judged ability, $F(2, 106) = 119.71$, $p < .001$, $\eta_p^2 = .69$. Below average performance resulted in lowest ratings of ability ($M = 3.00$), followed by average performance ($M = 4.09$), and above average performance ($M = 4.80$). As predicted (Hypothesis 1b), performance trend also affected judged ability, such that an improving trend led to higher judged ability ($M = 4.57$) than either a flat trend ($M = 3.69$) or a deteriorating trend ($M = 3.63$), which did not differ, $F(2, 106) = 37.37$, $p < .001$, $\eta_p^2 = .41$. The effect of performance variation was not significant, $F(1, 53) = .07$.

Effort. As predicted (Hypothesis 2a), mean performance influenced judged effort, such that below average mean performance resulted in the lowest ratings of effort ($M = 3.36$), average performance in medium ratings ($M = 3.99$), and above average performance in highest ratings ($M = 4.40$), $F(2, 106) = 57.66$, $p < .001$, $\eta_p^2 = .52$. Consistent with Hypothesis 2b, performance trend also affected judged effort, $F(2, 106) = 90.38$, $p < .001$, $\eta_p^2 = .63$. Raters judged effort to be highest when trend was improving ($M = 4.98$), with the flat trend falling in the middle ($M = 3.64$), and the deteriorating trend receiving the lowest ratings ($M = 3.12$). The effect of performance variation was not significant, $F(1, 53) = .65$.

Locus of causality. Consistent with Hypothesis 3, performance variation affected judged locus of causality, $F(1, 53) = 40.01$, $p < .001$, $\eta_p^2 = .48$. Specifically, large variation was attributed more to factors outside the ratee's control ($M = 3.95$) than was small variation ($M = 3.10$). Although not predicted, performance trend also affected judged locus of causality such that an improving trend was attributed more to internal factors ($M = 3.34$) than either a flat ($M =$

3.55) or deteriorating trend (M = 3.68), which did not differ, $F(2, 106) = 6.79$, $p < .01$, $\eta_p^2 = .11$. The effect of performance mean was not significant, $F(2, 106) = 2.37$.

Supplementary Analyses: Overall Performance Ratings

To assess the robustness of past findings, we examined the effects of performance mean, trend, and variation on overall performance ratings. Consistent with past findings, the higher the mean performance, the higher the performance ratings (below average, M = 2.61, average, M = 3.99, above average, M = 5.13), $F(2, 106) = 297.61$, $p < .001$, $\eta_p^2 = .85$. Also consistent with past findings, performance trend affected ratings (deteriorating, M = 3.36, flat, M = 3.67, improving, M = 4.69), $F(2, 106) = 134.82$, $p < .001$, $\eta_p^2 = .72$. Thus, the second contribution of this study is to replicate past findings (e.g., Reb & Cropanzano, 2007) regarding the effects of performance mean and trend on overall performance ratings and doing so in a different cultural context (i.e., Singapore). In contrast to previous findings, performance variation affected ratings such that small variation led to more favorable ratings (M = 4.06) than large variation (M = 3.76), $F(1, 53) = 18.27$, $p < .001$, $\eta_p^2 = .26$. A possible explanation lies in the strength of our variation manipulation. Specifically, the standard deviation was double in both the small and large variance conditions than that of Reb and Cropanzano.

On an exploratory basis, we analyzed whether dynamic performance characteristics interacted to predict ratings. First, performance mean interacted with performance trend, $F(4, 212) = 12.75$, $p < .001$, $\eta_p^2 = .19$. Analysis of cell means suggests that below average performance coupled with a deteriorating (M = 2.09) or flat trend (M = 2.05) led to particularly negative evaluations, but coupled with an improving trend protected ratees somewhat from poor evaluations (M = 3.70). Indeed, the difference in ratings between the average and below average condition was smaller when trend was improving (M = 3.70 to M = 4.66, diff = .96) than when it

was deteriorating (M = 2.09 to M = 3.52, diff = 1.43) or flat (M = 2.05 to M = 3.78, diff = 1.73). Second, performance trend interacted with performance variation, $F(2, 106) = 30.33$, $p < .001$, $\eta_p^2 = .36$. Ratings increased significantly more from a deteriorating (M = 3.36) to an improving trend (M = 5.13, diff = 1.77) when variation was small rather than large (M = 3.37 to M = 4.25, diff = .88) possibly due to the increased salience of trend when variation was small. Overall, these findings highlight the need to examine interactions among dynamic performance characteristics. Study 2 will develop formal hypotheses and attempt to replicate these findings.

Study 1 has several limitations. First, using Campbell et al.'s theory as a basis for identifying attributions, we focused on ability, effort, and locus of causality. Future research should assess additional attributions, such as mood, and causal dimensions, such as stability, drawing on attribution theory (e.g., Weiner, 1979). Second, to achieve control over the nature of the performance profiles, raters evaluated hypothetical employees, which raises concerns about the external validity of our results. However, research indicates that there is a high degree of correspondence between findings from laboratory and field research (Anderson, Lindsay, & Bushman, 1999), and if anything, the effects are probably larger in the field than in the laboratory for performance appraisal research (Murphy & Cleveland, 1995).

Study 1 contributes to the dynamic performance literature by linking dynamic performance characteristics (i.e., mean, trend, and variation) to attributions of ability, effort, and locus of causality. Study 2 builds on Study 1 by drawing on the results regarding rater attributions to derive hypotheses predicting that rating purpose moderates the influence of mean, trend, and variation on overall performance ratings.

Study 2

The importance of rating purpose in the performance appraisal process is widely

acknowledged (e.g., Greguras & Robie, 1998). Research suggests that ratings made for administrative purposes are generally more lenient (Jawahar & Williams, 1997), less variable (Farh, Cannella, & Bedeian, 1991), less reliable (Greguras, Robie, Schleicher, & Goff, 2003), and less accurate (McIntyre, Smith, & Hassett, 1984) than ratings made for developmental purposes. Given these effects, rating purpose is considered one of the most important factors when designing and implementing a performance management system (e.g., Balzer, Greguras, & Raymark, 2004). Study 2 integrates research on rating purpose and dynamic performance that has so far developed independently.

Performance Mean and Rating Purpose

We expect rating purpose to interact with performance mean such that administrative ratings will be more influenced than developmental ratings by ratee mean performance. Because administrative ratings are often used to make compensation decisions, mean level of performance should be especially salient to raters. As noted by Fossum and Fitch (1985), "Usually espoused organizational policy indicates that pay increases should be based on performance…" (p. 590). Indeed, many companies advocate merit-based pay or pay-for-performance policies. Consistent with this view, research indicates that mean performance level significantly impacts compensation decisions (Deshpande & Schoderbek, 1993).

Further, as observed in Study 1, high mean performance tends to be attributed to high ability. Organizations desire to retain highly skilled employees in order to remain competitive. Past research found a curvilinear relation between performance and turnover, such that high and low performers were more likely to turnover (Williams & Livingstone, 1994). Moreover, this relationship is moderated such that turnover for high performers is reduced (exacerbated) by high (low) salary growth (Trevor, Gerhart, & Boudreau, 1997). As such, administrative ratings and

decisions (e.g., employee raises) that are primarily influenced by mean level (Fossum & Fitch, 1985) may help organizations to actively influence turnover in the desired direction, enticing skilled, high-performing employees to stay and low-performing employees to leave.

H4: Performance mean and rating purpose interact to predict performance ratings such that the difference between ratings in the above average and the below average mean condition is larger when rating purpose is administrative than when rating purpose is developmental.

Performance Trend and Rating Purpose

We expect rating purpose to interact with performance trend such that developmental ratings will be more influenced than administrative ratings by performance trend. As the results of Study 1 suggest, when performance improves, employees may be perceived as learning and developing skills or abilities. Moreover, Study 1 found that judged effort was higher the more positive the trend. Employees showing an improving trend may be perceived as putting in the effort required to learn and acquire necessary skills, whereas those with a deteriorating trend may be perceived as lacking motivation given that they had previously performed at higher levels.

In attribution theory (Weiner, 1985), effort is considered an internally controllable cause of performance. Because effort is seen as controllable, it suits itself for feedback interventions. As such, ratings that reflect such changes may be especially useful for employee development. For example, improving employees may be evaluated positively as a sign of acknowledgment or to motivate further improvements. Deteriorating employees may be evaluated negatively to send a strong signal that performance is decreasing and that change is expected. Such a strong consideration of performance trend in the assignment of ratings may be considered less appropriate for administrative purposes which may more appropriately capture average performance across the evaluation period (Fossum & Fitch, 1985).

H5: Performance trend and rating purpose interact to predict performance ratings such that the difference between ratings in the improving and the deteriorating trend condition is

larger when rating purpose is developmental than when rating purpose is administrative.

Performance Variation and Rating Purpose

We expect rating purpose to interact with performance variation such that developmental ratings will be more influenced than administrative ratings by performance fluctuations. Results from Study 1 suggest that raters attribute large performance variability to situational factors. These situational factors present employees with opportunities or constraints that can influence performance variability such that employees who are more able to adapt perform more favorably thereby reducing the variability of their performance (Hesketh & Neal, 1999).

Because employees may be able to be coached on how to respond to situational constraints or opportunities, developmental ratings are expected to be more influenced by performance variability than are administrative ratings. For example, managers may use negative developmental ratings for inconsistent performance to highlight a developmental need. In contrast, stable performance may be acknowledged with positive ratings. Indeed, people who perform consistently receive positive evaluations (Fox, Bizman, Hoffman & Oren, 1995). Taken together, these arguments suggest that large variation will lead to more negative developmental ratings. However, when ratings are made for administrative purposes, a strong consideration of performance variation may be considered less appropriate because this variation does not affect the mean level of performance.

H6: Performance variation and rating purpose interact to predict performance ratings such that the difference between ratings in the small and the large variation condition is larger when rating purpose is developmental than when rating purpose is administrative.

Interactions among Dynamic Performance Characteristics

Study 1 reported that performance mean interacted with trend such that when mean performance over the evaluation period was below average, a deteriorating and a flat trend led to

particularly unfavorable evaluations, whereas an improving protected ratees to some extent from the negative effects of the below average mean. A possible interpretation is that supervisors expect at least average performance from their subordinates over the longer run. When performance shows an improving trend, showing about average performance at the end of the observation period, raters may assume that the employee is able and willing to perform at least at an average level in the near future. No such expectations may exist towards employees with below average performance a flat or deteriorating trend.

H7: Performance mean and performance trend interact to predict performance ratings such that the difference between ratings in the below average and average mean condition is larger when trend is flat or deteriorating than when trend is improving.

In Study 1, performance trend interacted with variation such that ratings increased significantly more from a deteriorating to an improving trend when variation was small rather than large. This finding is consistent with Reb and Cropanzano's (2007) theorizing that salient Gestalt characteristics of a performance profile influence ratings. When variation is small, trend becomes more recognizable and salient and, thus, should have a stronger influence on ratings.

H8: Performance trend and performance variation interact to predict performance ratings such that the difference between ratings in the deteriorating and improving trend condition is larger when performance variation is small than when variation is large.

Method

Participants and Procedure

Managers (N = 176) from a variety of industries participated on a voluntary basis (response rate = 87%). Of these, 32 (18%) were excluded because they did not answer a manipulation check correctly (i.e., they were unable to identify the correct rating purpose). Three additional participants (2%) were excluded because they indicated that they worked in a non-managerial position. The final sample consisted of 141 participants of which 73% were male. The average

age was 42.2 years (*SD* = 11.1). About 87% indicated that they currently formally evaluate others as part of their jobs, with an average of 15.1 evaluees (*SD* = 26.4). On average, participants had supervised employees for about 10.8 years (*SD* = 10.0) and had worked for their present company 11.2 years (*SD* = 10.6) and in their current position 5.6 years (*SD* = 6.2).

Trained undergraduate students were given extra credit for distributing survey packets to full-time working adults with supervisory experience. Each participant was given a survey packet containing an instruction letter, the survey, and a self-addressed stamped return envelope. To check the data quality of responses, about 11% (*n* = 20 of 176) of respondents were called. All verified participation in this study and one demographic piece of information (e.g., age).

Materials and Manipulations

The design, materials, and task instructions were the same as in Study 1 but also included a between-subjects manipulation of rating purpose. Thus, the experiment consisted of a mixed 3 (performance mean) x 3 (performance trend) x 2 (performance variation) x 2 (rating purpose: administrative vs. developmental) design. Order of presentation of the performance profiles was randomized for each participant. We also repeated three profiles at the end of the study in order to estimate reliability (these repeated profiles were excluded from substantive analyses).

Within the instructions, a paragraph titled "Purpose of Performance Ratings" manipulated rating purpose. Participants in the administrative decision condition were told that "Your evaluations are going to be used to make administrative decisions about the salespeople." In addition, examples of administrative decisions were given (e.g., pay increase, promotion, termination). Participants in the developmental feedback condition read that "Your evaluations are going to be used to provide developmental feedback to the salespeople." Examples of developmental feedback were given (e.g., performance feedback, identifying areas for

improvement). To further improve the salience of the ratings purpose manipulation, participants were asked to write down "some advantages and disadvantages of evaluating subordinate performance in order to [make these administrative decisions / give developmental feedback]."

Measures

Overall performance was assessed with two items. The first item had participants evaluate the employee on a 5-point scale (1: very poor, 5: very good): "Overall, how would you rate this employee's performance?" The second item asked participants to evaluate employee performance on an 11-point scale anchored at -100: worst performance and 100: best performance. These two ratings were combined ($\alpha = .94$) by first transforming the first rating to range from -100 to 100 and then averaging this measure with the second rating. A small number of missing values on the two items ($n = 13$; .005% of the complete data) was replaced with cell means. Test-retest reliability estimates for the three repeated performance profiles were on average $r = .67$, consistent with past findings (Sturman, Cheramie & Cashen, 2005).

Results and Discussion

We conducted a 3 (trend) x 3 (mean) x 2 (variation) x 2 (rating purpose) mixed-measures ANOVA on performance ratings. All pairwise comparisons reported below are significant, $p < .05$, unless otherwise noted. Replicating Study 1, performance mean affected performance ratings (below average, $M = -40.34$, average, $M = 1.92$, above average, $M = 41.36$), $F(2, 278) = 715.05$, $p < .001$, $\eta_p^2 = .84$; performance trend affected ratings (deteriorating, $M = -25.95$, flat, $M = 3.36$, improving, $M = 32.26$), $F(2, 278) = 348.77$, $p < .001$, $\eta_p^2 = .72$; and performance variation affected ratings (small, $M = 2.14$, large, $M = -.18$), $F(1, 139) = 4.75$, $p < .05$, $\eta_p^2 = .03$. Thus Study 2 replicates the results of Study 1. Importantly, whereas past studies used relatively inexperienced participants, Study 2 is the first to use a sample of managers with considerable

supervisory experience.

Dynamic Performance Characteristics and Rating Purpose

Performance mean. Consistent with Hypothesis 4, rating purpose interacted with mean performance such that performance mean influenced administrative ratings more strongly than developmental ratings to affect overall performance ratings, $F(2, 278) = 5.31$, $p < .01$, $\eta_p^2 = .04$. (see Figure 1). As predicted, the difference between the ratings in the above average mean condition ($M = 43.46$) and in the below average mean condition ($M = -44.70$) was larger when rating purpose was administrative (diff = 88.16) rather than developmental (below average, $M = -35.97$, above average, $M = 39.26$, diff = 75.23), $F(1, 139) = 5.13$, $p < .05$.

Performance trend. Consistent with Hypothesis 5, rating purpose interacted with performance trend to predict overall ratings, $F(2, 278) = 5.35$, $p < .01$, $\eta_p^2 = .04$. Trend had a stronger impact on developmental than on administrative ratings (see Figure 2). As predicted, the difference between developmental ratings for an improving ($M = 36.21$) and a deteriorating trend ($M = -29.27$) was larger (diff = 65.48) than this difference for administrative ratings (improving, $M = 28.30$, deteriorating, $M = -22.64$, diff = 50.94), $F(1, 139) = 6.45$, $p < .05$.

Performance variation. Consistent with Hypothesis 6, rating purpose interacted with performance variation such that variation affected developmental ratings more than administrative ratings, $F(1, 139) = 6.56$, $p = .01$, $\eta_p^2 = .05$ (see Figure 3). As expected, when rating purpose was developmental, the difference between ratings in the small ($M = 3.83$) and the large ($M = -1.19$) variation condition was larger (diff = 5.02) than when purpose was administrative (small, $M = .44$, large, $M = .87$, diff = -.43), $F(1, 139) = 6.56$, $p = .01$.

Following calls to examine the influence of contextual factors on performance ratings (Murphy & Cleveland, 1995), this study observed that dynamic performance characteristics

interacted with rating purpose (administrative or developmental) to predict overall performance ratings. These findings are important as they suggest that supervisors may consciously or unconsciously consider dynamic performance information to fit the specific appraisal purpose. For feedback purposes, the goals are for ratings to highlight future-oriented developmental needs and to acknowledge performance strengths, which may be inferred from trends and variation in performance. In contrast, for administrative ratings (e.g., merit pay decisions) the goal likely is for ratings to closely reflect average performance across a given period (Fossum & Fitch, 1985).

Interactions among Dynamic Performance Characteristics

Consistent with Hypothesis 7, performance mean interacted with performance trend, $F(4, 556) = 31.96$, $p < .001$, $\eta_p^2 = .19$ (see Figure 4). As predicted, the difference in ratings between the below average and the average performance mean condition was significantly larger when trend was deteriorating ($M = -70.17$ to $M = -19.93$, diff = 50.24) or flat ($M = -52.27$ to $M = -.77$, diff = 51.50) than when trend was improving ($M = .23$ to $M = 26.82$, diff = 26.59).

Consistent with Hypothesis 8, performance trend interacted with performance variation such that trend had a stronger influence when variation was small, $F(2, 278) = 45.20$, $p < .001$, $\eta_p^2 = .25$ (see Figure 5). As predicted, ratings increased significantly more from a deteriorating ($M = -30.38$) to an improving trend ($M = 38.69$, diff = 69.07) when variation was small rather than large ($M = -21.53$ to $M = 25.82$, diff = 47.35).

Although not predicted, performance mean interacted with performance variation, $F(2, 278) = 11.25$, $p < .001$, $\eta_p^2 = .08$. Below average mean performance led to more unfavorable ratings when variation was large ($M = -44.67$) rather than small ($M = -36.00$), $p < .001$. A possible explanation is that when variation was large, the minimum values ("downside peaks") were lower than when variation was small, which may have led to the perception that the employee

was a particularly poor performer. Such effects of peaks on the evaluation of dynamic sequences have been previously observed (Ariely & Carmon, 2000).

Study 2 has several strengths and limitations. The contributions include (a) theoretically hypothesizing and observing several interactions between rating purpose and performance mean, trend, and variation, (b) theoretically hypothesizing and replicating several interactions among dynamic performance characteristics, and (c) replicating the main effects of performance mean, trend, and variation on performance ratings with a sample of experienced managers. A limitation of Study 2 is that each rater provided evaluations for a singly purpose only (administrative or developmental). Although research on rating purpose has typically only investigated one purpose at a time, raters often evaluate employees for multiple purposes simultaneously (Murphy & Cleveland, 1995). Future research should explore how dynamic performance characteristics affect ratings when ratings are used for multiple purposes. As with Study 1, the scenario-based nature of our study also raises concerns about the external validity of our results. To overcome this limitation and to extend our findings, future research should explore how dynamic performance characteristics influence a variety of outcomes in a field context (e.g., satisfaction with coworkers, working relationships, perceptions of dependability).

Conclusion

Fisher (2008) recently suggested that organizations should acknowledge the importance of dynamic performance. In response, we conducted two experiments in which we empirically examined the complex interplay between dynamic performance characteristics, rater attributions, and rating purpose in predicting overall performance ratings. Consistent with past research, we found that dynamic performance features significantly influenced performance ratings. As a result, two employees having the same mean performance may receive substantially different

ratings (cf. Reb & Cropanzano, 2007). Based on results from Study 1, they also may be considered as substantially differing in terms of ability, effort, and locus of causality.

Further complicating matters, dynamic performance characteristics interacted to predict ratings. For example, an improving (deteriorating) trend led to more favorable (unfavorable) ratings when variation was small rather than large, presumably due to the increased salience of trend when variation is small (cf. Reb & Cropanzano, 2007). Study 2 showed that dynamic characteristics interacted not only with each other, but also with rating purpose. Thus, depending on the rating purpose, two employees showing the same profile of performance over time may receive different ratings. For example, given the same improving trend, developmental ratings were higher than administrative ratings.

At first, this seems to introduce additional, and unwanted, complexities into the already challenging task of evaluating employee performance. However, from a practical perspective rating systems that can capture dynamic performance information may provide extensive practical value (cf. Kane, 1986, 2000). In particular, such systems can provide information specifically tailored to particular rating purposes. For example, if one is attempting to award end-of-year bonus money (i.e., an administrative purpose), then mean performance might be most relevant. If the bonus is a reward for past behavior, then it would seem reasonable to "partial out" the effects of other distributional characteristics. On the other hand, if one is attempting to discover developmental needs, then trend and variation data would seem especially valuable. Our findings suggest that supervisors already make use of dynamic performance information in this manner, at least implicitly. Designing performance appraisal systems to take into account the dynamic nature of performance would allow organizations the ability to make even better and more explicit use of this information.

REFERENCES

Anderson, C. A., Lindsay, J. J., & Bushman, B. J. (1999). Research in the psychological laboratory: Truth or triviality? *Current Directions in Psychological Science, 8*, 3- 9.

Ariely, D., & Carmon, Z. (2000). Gestalt characteristics of experiences: The defining features of summarized events. *Journal of Behavioral Decision Making, 13,* 191-201.

Balzer, W. K., Greguras, G. J., & Raymark, P. (2004). Multisource feedback. In Michel Hersen and Jay Thomas (Eds.), *Comprehensive Handbook of Psychological Assessment* (Vol. 4, pp. 390-411). New York, NY: John Wiley & Son.

Beal, D. J., Weiss, H. M., Barros, E., & MacDermid, S. M. (2005). An episodic process model of affective influences on performance. *Journal of Applied Psychology, 90,* 1054-1068.

Campbell, J. P., McCloy, R. A., Oppler, S. H., & Sager, C. E. (1993). A theory of performance. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations* (pp. 35-70). San Francisco: Jossey-Bass Publishers.

Deadrick, D. L., Bennett, N., & Russell, C. J. (1997). Using hierarchical linear modeling to examine dynamic performance criteria over time. *Journal of Management, 23,* 745-757.

DeNisi, A., & Stevens, G. E. (1981). Profiles of performance, performance evaluations, and personnel decisions. *Academy of Management Journal, 24,* 592-602.

Deshpande, S. P., & Schoderbek, P. P. (1993). Pay-allocations by managers: A policy-capturing approach. *Human Relations, 46,* 465-479.

Dweck, C. S., & Leggett, E. L. (1988). A social-cognitive approach to motivation and personality. *Psychological Review, 95,* 256-273.

Farh, J., Cannella, A. A., Jr., & Bedeian, A. G. (1991). Peer ratings: The impact of purpose on rating quality and user acceptance. *Group and Organizational Studies, 16,* 367-386.

Fedor, D. B., & Rowland, K. M. (1989). Investigating supervisor attributions of subordinate performance. *Journal of Management, 15,* 405-416.

Fisher, C. D. (2008). What if we took within-person performance variability seriously? *Industrial and Organizational Psychology, 1,* 185-189.

Fossum, J. A., & Fitch, M. K. (1985). The effects of individual and contextual attributes on the sizes of recommended salary increases. *Personnel Psychology, 38,* 587-602.

Fox, S., Bizman, A., Hoffman, M., & Oren, L. (1995). The impact of variability in candidate profiles on rater confidence and judgments regarding stability and job suitability. *Journal of Occupational and Organizational Psychology, 68,* 13-23.

Greguras, G. J., & Robie, C. (1998). A new look at within-source interrater reliability of 360-degree feedback ratings. *Journal of Applied Psychology, 83*, 960-968.

Greguras, G. J., Robie, C., Schleicher, D. J., & Goff, M. III (2003). A field study of the effects of rating purpose on the quality of multisource ratings. *Personnel Psychology, 56,* 1-21.

Hesketh, B., & Neal, A. 1999. Technology and performance. In D. R. Ilgen, & D. E. (Eds.), *The changing nature of performance: Implication for motivation, and development* (pp. 21-55). San Francisco, CA: Jossey-Bass.

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72-98.

Jawahar, I. M., & Williams, C. R. (1997). Where all the children are above average: The performance appraisal purpose effect. *Personnel Psychology, 50,* 905-925.

Jones, E. E., Rock, L., Shaver, K. G., Goethals, G. R., & Ward, L. M. (1968). Pattern of performance and ability attribution: An unexpected primacy effect. *Journal of Personality and Social Psychology, 10,* 317-340.

Kane, J. S. (1986). Performance distribution assessment. In R. A. Berk (Ed.), *Performance assessment: Methods and applications* (pp. 237-273). Baltimore: Johns Hopkins University Press.

Kane, J. S. (2000). Accuracy and its determinants in distributional assessment. *Human Performance, 13*, 47-84.

McIntyre, R. M., Smith, D. E., & Hassett, C. E. (1984). Accuracy of performance ratings as affected by rater training and perceived purpose of rating. *Journal of Applied Psychology, 69,* 147-156.

Murphy, K. R., & Cleveland, J. N. (1995). *Understanding performance appraisal: Social, organizational, and goal-based perspectives.* Thousand Oaks, CA: Sage Publications.

Porac, J. F., Ferris, G. R., & Fedor, D. B. (1983). Causal attributions, affect, and expectations for a day's work performance. *Academy of Management Journal, 26,* 285-296.

Reb, J., & Cropanzano, R. (2007). Evaluating dynamic performance: The influence of salient gestalt characteristics on performance ratings. *Journal of Applied Psychology, 92*, 490-499.

Reb, J., & Greguras, G. J. (2008). Dynamic performance and the performance-performance rating relation. *Industrial and Organizational Psychology, 1*, 194-196.

Ronis, D. L., Hansen, R. D., & O'Leary, V. E. (1983). Understanding the meaning of achievement attributions: A test of derived locus and stability scores. *Journal of Personality and Social Psychology, 44,* 702-711.

Russell, D. W., McAuley, E., & Tarico, V. (1987). Measuring causal attributions for success and failure: A comparison of methodologies for assessing causal dimensions. *Journal of Personality and Social Psychology, 52,* 1248-1257.

Scott, W. E. Jr., & Hamner, W. C. (1975). The influence of variations in performance profiles on

the performance evaluation process: An examination of the validity of the criterion. *Organizational Behavior and Human Performance, 14,* 360-370.

Slifkin, A. B., & Newell, K. M. (1998). Is variability in human performance a reflection of system noise? *Current Directions in Psychological Science, 7,* 170-177.

Sturman, M. C., Cheramie, R. A., & Cashen, L. H. (2005). The impact of job complexity and performance measurement on the temporal consistency, stability, and test--retest reliability of employee job performance ratings. *Journal of Applied Psychology, 90*, 269-283.

Trevor, C. O., Gerhart, B., & Boudreau, J. W. (1997). Voluntary turnover and job performance: Curvilinearity and the moderating influences of salary growth and promotions. *Journal of Applied Psychology, 82,* 44-61.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71,* 3-25.

Weiner, B. (1985). An attributional theory of achievement motivation and emotion. *Psychological Review, 92,* 548-573.

Williams, C.R., & Livingstone, L.P. (1994) Another look at the relationship between performance and voluntary turnover. *Academy of Management Journal, 37,* 269-298.

Figure 1: Interaction between Performance Mean and Rating Purpose on Performance Ratings, Study 2

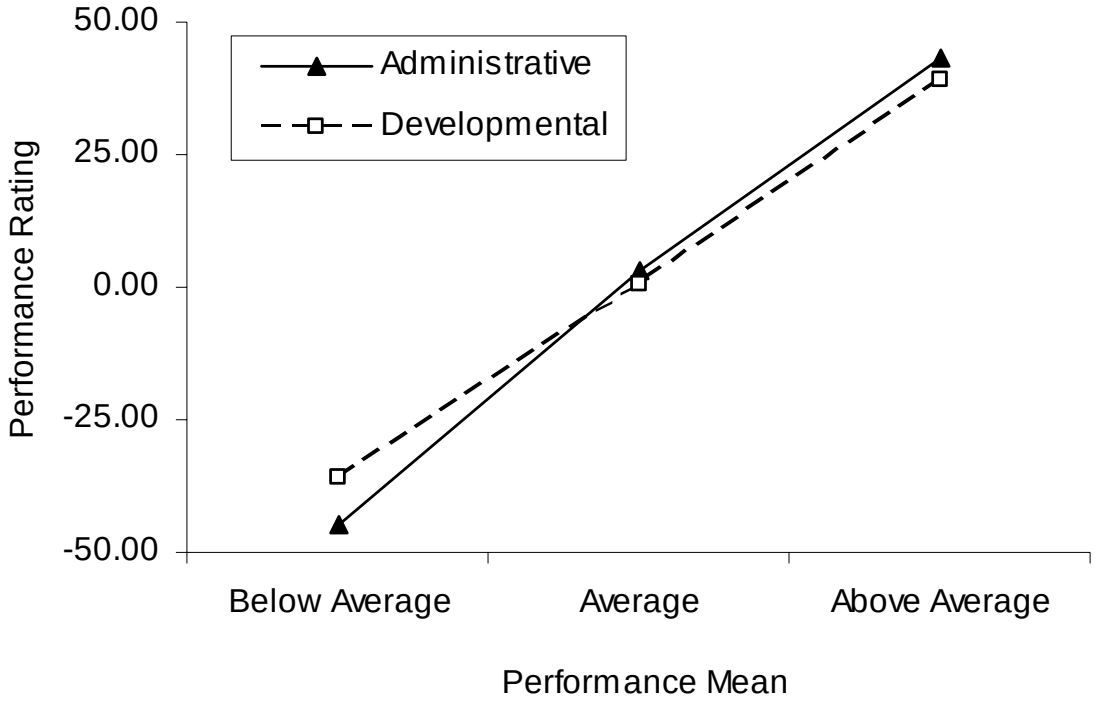

Figure 2: Interaction between Performance Trend and Rating Purpose on Performance Ratings, Study 2

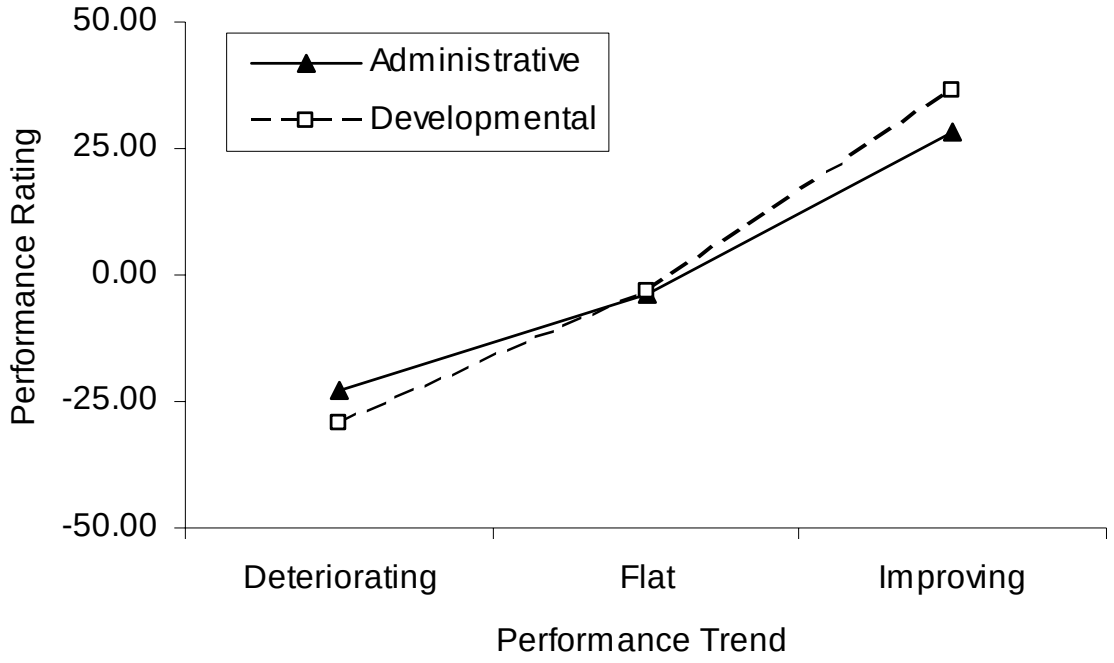

Figure 3: Interaction between Performance Variation and Rating Purpose on Performance Ratings, Study 2

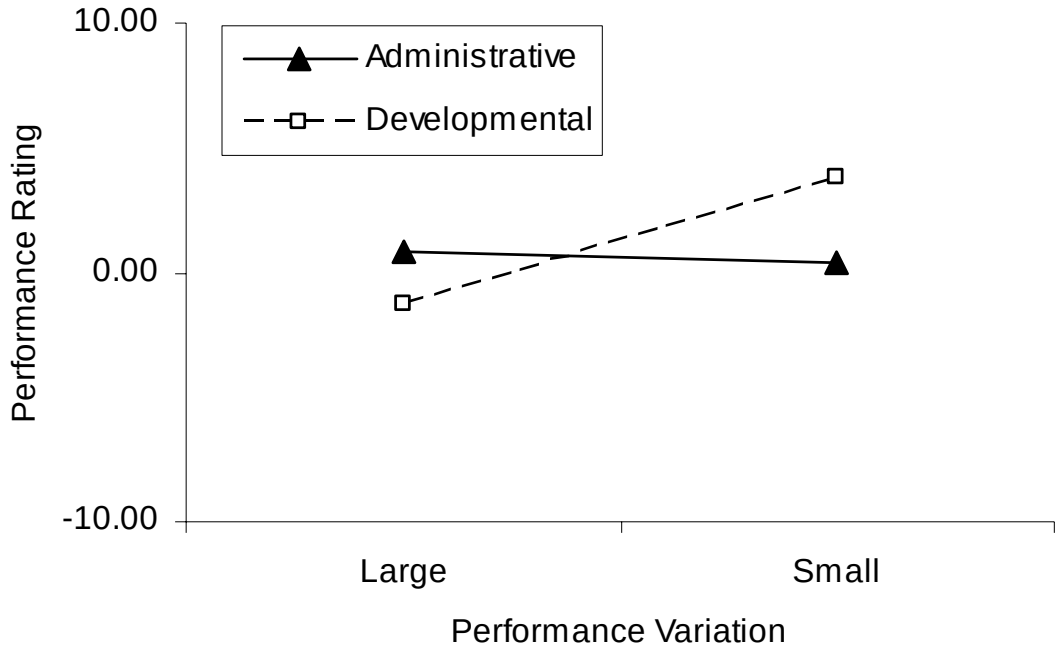

Figure 4: Interaction between Performance Trend and Performance Mean on Performance Ratings, Study 2

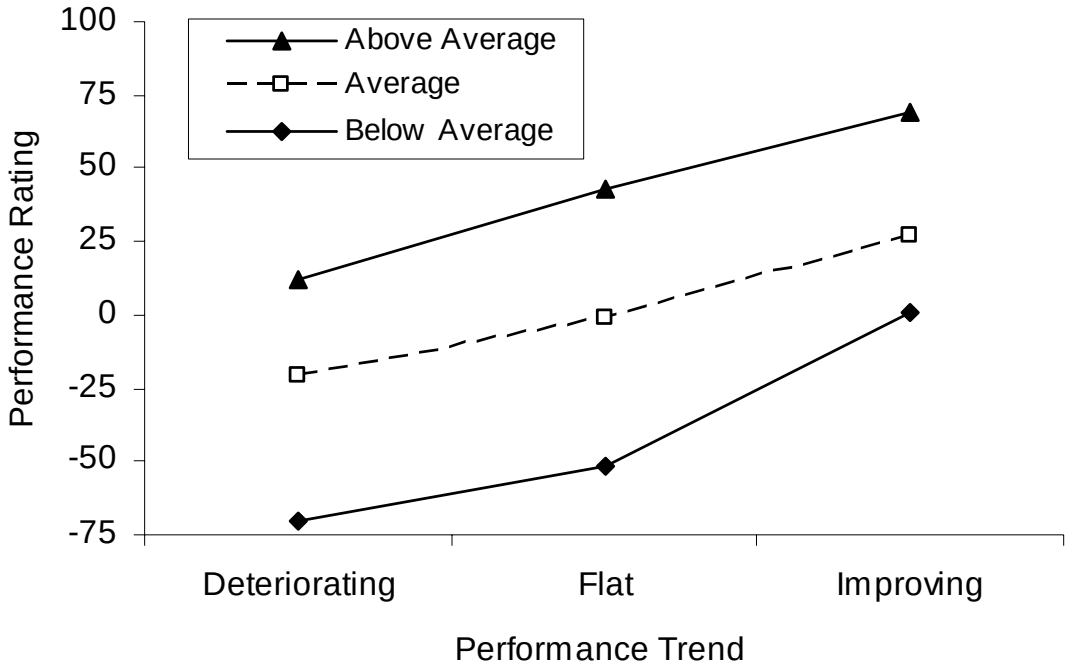

Figure 5: Interaction between Performance Trend and Performance Variation on Performance Ratings, Study 2