JOURNAL OF ORGANIZATIONAL BEHAVIOR, Vol 10, 83–90 (1989)

Research Note

Employee performance evaluations: Effects of ratee age, rater age, and ratee gender

RODGER W. GRIFFETH

AND

ARTHUR G. BEDEIAN

Department of Management, Louisiana State University, Baton Rouge, LA. 70803-6312, U.S.A.

Introduction

Research into the effects of gender and race on employee performance evaluations has become quite common over the past decade (Rosenstein and Hitt, 1986). By contrast, research dealing with the effects of other nontask factors (e.g. ratee age) on evaluations of employee performance have been less frequently investigated (*cf.* Waldman and Avolio, 1986). Of concern to the present investigation is that a clear pattern of results has yet to emerge for either rater or ratee effects with respect to either gender or age differences in performance ratings (Williams and Walker, 1985). Consequently, questions about the possible influence of gender and age on employee performance evaluations have persisted. Predictably, calls for further research have mounted in intensity with at least one such call (i.e. Cleveland and Landy, 1981) stressing the need to investigate the *combined* influence of rater–ratee characteristics upon performance ratings.

Past research on the effects of gender and age on employee performance evaluations has largely involved laboratory simulations. This fact has been increasingly offered as at least one explanation for the current mixed and often contradictory results reported concerning the effects of nontask factors (e.g. gender and race) on performance ratings. Dipboye (1985) gives a review of this literature. Critics (e.g. Wendelken and Inn, 1981) argue that such simulations bear little resemblance to the typical performance appraisal process in real organizations. They contend that at least three major differences suggest caution in generalizing from traditional laboratory research to actual performance appraisal situations. First, most laboratory simulations involving performance appraisals have used hypothetical ratee profiles. Second, simulation studies generally present raters with a limited number of artificially created orthogonal performance evaluation dimensions on a one time basis. Finally, simulations largely focus on just one aspect of the total performance evaluation process (DeNisi, Cafferty and Meglino, 1984). That is, integration of information to form a rating.

The present study attempted to overcome these methodological problems by collecting actual (rather than simulated) performance ratings from a large number of superior–subordinate dyads

The helpful comments of Angelo J. Kiniki and Peter W. Hom on an earlier draft of this manuscript are gratefully acknowledged.

0894–3796/89/010083–08$05.00

Received 17 July 1987
Revised 17 November 1987

in a field setting. Hence, employee performance was observed and assessed within the context of frequent face-to-face interactions between raters and ratees. Moreover, as a consequence of being conducted in a real organizational environment, ratee performance was assessed as it naturally varied — it was not held constant or artificially manipulated.

The current study was undertaken to investigate the effects attributable to two nontask factors — gender and age — acting as main effects and in interaction on performance ratings. Five hypotheses were investigated. The first three were suggested by the previous field research of Cleveland and Landy (1981).

H_1: Younger raters would give significantly lower evaluations than older raters.

H_2: Younger ratees would receive significantly higher evaluations than older ratees.

H_3: There would be a significant Age of Rater × Age of Ratee interaction.

With respect to H_3, it was expected that older raters would give lower evaluations to older ratees than younger ratees, but that there would be a larger difference in the evaluations of older and younger ratees judged by younger raters.

H_4: Male ratees would receive significantly higher evaluations than female ratees.

This hypothesis was derived from a meta-analysis by Dobbins and Trahan (1986) which found that performance evaluations are biased against females occupying male-dominated occupations.

H_5: There would be a significant Age of Rater × Gender of Ratee interaction.

This hypothesis was suggested by Dipboye's (1985) analysis of neglected performance appraisal variables. On balance, this analysis suggests that both female and older ratees typically encounter greater sex and age bias than male or younger employees, respectively. Moreover, such biases may well vary by rater age. Thus, it was anticipated that in comparison to male ratees, raters of all ages would give lower evaluations to female ratees and that this difference would be greater for older raters. Rater gender (and race) was controlled by including only white, male supervisors in the sample described below. Additionally, this was a practical consideration given the small number of female supervisors satisfying the study's data requirements.

Method

Background

The present sample, 464 supervisor–subordinate dyads, was identified from a larger data set of 1080 respondents randomly selected (with a 63 per cent participation rate) from the membership lists of four U.S.-based professional accounting associations. Supervisor and subordinate age was collected from both respondent groups using an identical ordinal scale. All supervisors (raters) were white males. Approximately 43 per cent (n = 193) were between the ages of 30–39, with roughly 32 per cent (n = 147) and 27 per cent (n = 124) between the ages of 40–49 or over 50, respectively. Some 53 per cent of the subordinates (ratees) were male. Twenty-three per cent (n = 109) were between the ages of 20–29, with roughly 41 per cent (n = 191), 20 per cent (n = 92), and 16 per cent (n = 72) between the ages of 30–39, 40–49, or over 50, respectively.

Measures

Each subordinate was given a research packet containing a performance evaluation form to be completed by his or her supervisor. Evaluations were collected expressly for use by the researchers to minimize sources of possible rater bias (e.g. halo, leniency) affecting performance appraisals when they are used for human resource planning decisions (Bernardin and Villanova, 1986).

Accordingly, supervisors were instructed to rate their subordinates on 23 items, each with a seven-point scale ranging from unsatisfactory (1) to excellent (7). A principal component factor analysis with varimax rotation of these items supported three factors accounting for 64.1 per cent of the explained variance. The first factor, clearly defined by 12 items with substantial loadings (>0.61), accounted for 50 per cent of the explained variance and was labelled Overall Job Performance ($\bar{x}$ = 5.86; S.D. = 0.76; α = 0.94). The second factor, accounting for 9 per cent of the extracted variance, has six items with factor loadings greater than 0.59. It was designated Commitment ($\bar{x}$ = 6.09; *S.D.* = 0.75; α = 0.87). The third factor, accounting for some 5 per cent of the explained variance, included three items with loadings greater than 0.66. It was named Interpersonal Relations ($\bar{x}$ = 5.83; *S.D.* = 0.92; α = 0.80).

Composite scales for each of the three factors were constructed by item unit weighting. The scales were moderately intercorrelated (mean r = 0.65). This is not unexpected given the common evaluative theme running through the content of the scales. Given that the scales do tap conceptually distinct information and because the study was not meant to be prescriptive, the scales were each maintained as separate dimensions. The lack of complete independence among the scales, however, does suggest the need for multivariate statistical methods.

Predictably, the distributions of the responses to the three performance scales were each negatively skewed. Such responses have repeatedly been shown to be representative of normal performance ratings which almost invariably tend to be lenient (Bernadin and Beatty, 1984). Considering the restricted distribution of the performance ratings, were there significant effects, the results would argue even more strongly for the efficacy of the investigated relationships.

Statistical analyses

A 3 × 4 multivariate analysis of variance (MANOVA) was conducted to evaluate the overall effects of the study's independent variables on the focal dependent variables. This procedure takes into account correlations among multiple dependent variables while considering their simultaneous impact. MANOVA rather than canonical correlation was used since both rater and ratee age were collected using categorical scales. While this may have lead to a loss of statistical power, it was considered an acceptable constraint since gender is, by definition, a categorical variable. The first factor, rater age, consisted of three levels: (1) 30–39, (2) 40–49, and (3) 50 and over. The second factor, ratee age, consisted of four levels: (1) 20–29, (2) 30–39, (3) 40–49, and (4) 50 and over. A 3 × 2 MANOVA was also conducted on the dependent variables. This analysis consisted of the three levels of rater age as the first factor and ratee gender as the second factor.

To interpret significant MANOVA differences, discriminant analysis was used. Discriminant analysis is the method of choice when dependent variables are intercorrelated. Interpretation of the discriminant functions was based on structure coefficients, which are the correlations between respondents' scores on the study's original measures and derived discriminant functions. A variable was considered to define a function if it was strongly ($\geq$ 0.40) correlated with that function.

Results

Hypotheses 1, 2, and 3

Hypotheses 1, 2, and 3 involved the effects of rater age and ratee age on employee performance evaluations. Specifically, hypothesis 1 suggested that younger raters would give significantly lower

evaluations than older raters. This hypothesis was examined by inspecting the rater age main effect of the 3 × 4 MANOVA. The result indicates that there is a significant main effect attributable to rater age ($F = 5.9$, $df = 6/900$, $p \leq 0.05$).

Given this main effect, a three-group discriminant analysis was conducted using rater age as the group variable and the three dimensions of employee performance as the discriminator variables. We found one statistically significant function ($p \leq 0.05$), accounting for 99.4 per cent of the discriminable variance. The multivariate analogue of the omega squared was 0.08. That is 8 per cent of the variance in employee performance evaluations can be attributed to between-group differences.

The structure coefficients, discriminant weights, and group mean scores (i.e. the group centroids) on the single significant discriminant function are presented in Table 1. The structure coefficients indicate that overall job performance (0.93) and commitment (0.41) defined this function. In addition, the lower the group mean, the lower the rating on the structure coefficients defining the function. Using the group centroids and the structure coefficients as interpretative tools, it is clear that younger raters (age 30–39) gave lower evaluations of overall job performance and commitment than did either older rater group. Moreover, this trend seemed to increase as rater age increased. Thus, H_1 was substantially supported.

Table 1. Structure coefficients, standardized discriminant weights, and group means in a three-group (rater age) rotated discriminant function

Dependent measures	Structure coefficients	Standardized Discriminant Weights
	I	I
Overall job performance	0.93	0.88
Commitment	0.41	0.13
Interpersonal relations	0.29	0.05
Age group (raters)	Means (centroids)	
	I	
1. 30–39	–0.28	
2. 40–49	0.03	
3. ≥ 50	0.39	

Hypothesis 2 predicted that younger ratees would receive significantly higher evaluation than older ratees. This hypothesis was examined by inspecting the ratee age main effect of the 3 × 4 MANOVA. This revealed that variations in employee performance evaluations due to ratee age can be accounted for by chance alone ($F = 1.6$, $df = 9/1346$, n.s.).

Hypothesis 3 predicted a significant rater age × ratee age interaction. Results of the 3 × 4 MANOVA, however, indicate that there is no multivariate effect attributable to an interaction between rater age and ratee age. Multivariate F equals 1.3 ($df = 18/1346$, n.s.). Thus, H_3 was not tenable.

Hypotheses 4 and 5

Hypothesis 4 predicted a significant ratee gender main effect. No such effect was found, $F(3, 457) < 1.0$, *n.s.* Hypothesis 5 predicted a significant rater age × ratee gender interaction effect. The 3 × 2

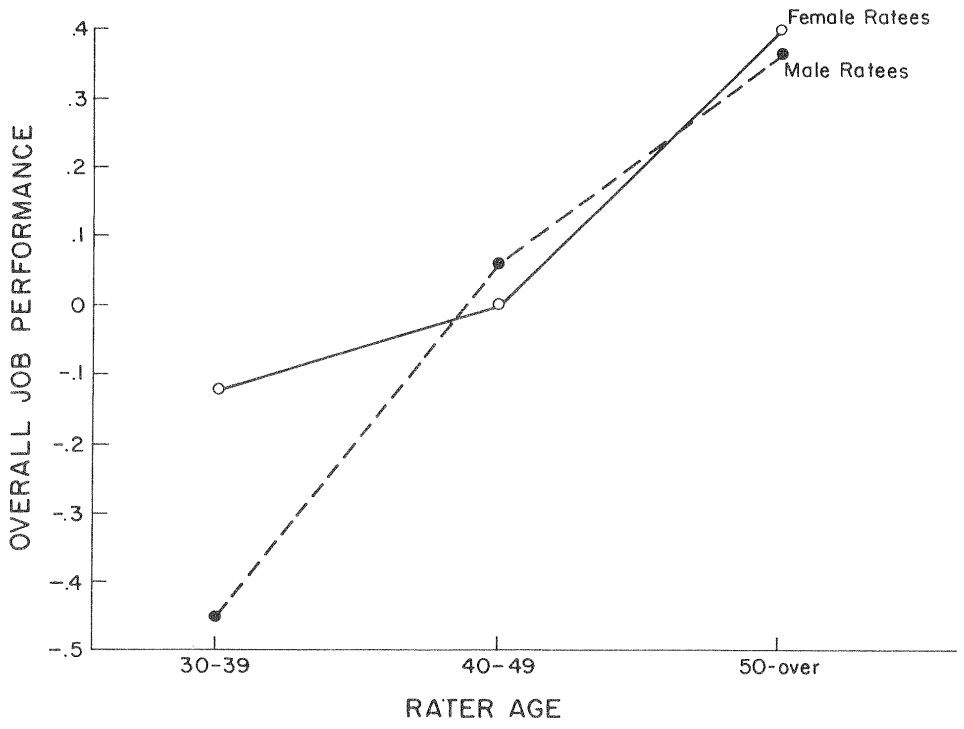

Figure 1. Group means in discriminant space defined by overall job performance

MANOVA for rater age by ratee gender did reveal a significant, F, (6, 912) = 7.7, $p \leq 0.05$, main effect for rater age. As predicted, a significant interaction was also found, $F(6, 912) = 2.1$, $p \leq 0.05$.

To interpret this significant interaction, a six-group (derived from the six cells of the MANOVA group classification) discriminant function analysis was performed. The results of this analysis found one significant ($p \leq 0.05$) function, which accounted for 87.6 per cent of the discriminable variance. The multivariate analogue of omega squared was 0.15. Thus, 15 per cent of the variance in employee performance evaluations was explained by between-group differences. The structure coefficients and discriminant weights for the single significant discriminant function are presented in Table 2. The structure coefficients indicate that the function was solely defined by overall job performance (0.91). Figure 1 presents a graphic representation of the magnitude of these means

Table 2. Structure coefficients and standardized discriminant weights for a six-group rotated discriminant function

Dependent measures	Structure coefficients	Standardized discriminant weights
	I	I
Overall job performance	0.91	0.82
Commitment	0.31	0.16
Interpersonal relations	0.28	0.10

for the six groups in discriminant space defined by overall job performance. It is apparent that the interaction was due to the relatively low evaluations given by 30–39 year old raters to male, as compared to female ratees. Moreover, there were no apparent differences in the evaluations given to either gender group by the 40–49 or 50 and over age raters. The similarity between the structure coefficients and discriminant weights reported in Tables 1 and 2 suggest that the relative importance of the study's dependent variables are the same in discriminating between groups.

Discussion

Our first hypothesis predicted that younger raters would give significantly lower evaluations than older raters. The reported findings demonstrate that younger raters did, in fact, give significantly lower evaluations on two of the three performance dimensions studied. While the amount of variance explained by rater age was relatively higher than previous field research (e.g. Cleveland and Landy, 1981, 2–3 per cent), it was still low (i.e. 8 per cent). Paradoxically, however, this result carries a positive implication. It suggests that nontask factors such as age and gender have a relatively small impact as biasing factors in performance evaluations. By default, other factors (hopefully task-related), largely determine employee performance evaluations.

Hypothesis 2 stated that younger raters would receive significantly higher evaluations than older raters. Our third hypothesis predicted that there would be a significant rater age × ratee age interaction. Neither hypothesis received support. No significant differences were found in the performance evaluations of older and younger ratees. Moreover, we failed to find a rater age × ratee age interaction.

Hypothesis 4 predicted that male ratees would receive significantly higher evaluations than female ratees. It failed to receive support. One possible explanation for this null finding may be the changing nature of the accounting profession. As noted, H_4 was drawn from results which indicate that performance evaluations are biased against females working in male-dominated occupations. These same results, however, also suggest that males and females are rated equally effective in neutral jobs (i.e. jobs held by an equal number of males and females). Field evidence suggests that as a result of the large numbers of females entering the profession over the last several years, accounting may no longer be 'male-dominated'. Rather, it may have become a 'neutral' occupation. Thus, if this is true, the present findings could be expected, especially when combined with what is known about the typically higher evaluations given to females portrayed as applying for traditionally male as compared to traditionally female jobs (Hartman, Harris, Crino and Griffeth, 1986).

Our fifth hypothesis stated that there would be a significant rater age × ratee gender interaction. Although a significant interaction was found, its exact nature was unexpected. Although prior research (e.g. Mobley, 1982) has reported females receiving higher performance evaluations than males, our findings introduce an important qualification. We found that evaluations varied not only as a function of ratee gender, but also as a consequence of rater age. Moreover, the nature of this interaction was unanticipated.

At least two alternatives may explain the finding that males in the 30–39 age group evaluated males lower than females on overall job performance. First, this result may be interpreted in light of Hall's (1976) career development model. Accordingly, male raters in the 30–39 age category would be on the threshold of the establishment/advancement phase of their careers. It is possible that they (consciously or unconsciously) view males as greater rivals for promotions than females. Women may not be seen as great a competitive threat since field interviews suggest that large numbers ultimately leave the accounting profession to establish a family. A similar phenomenon

has been documented in the nursing profession (e.g. Deets and Froebe, 1984). Raters 50 and over could conceivably be insulated from these feelings since most are likely to be at the final or maintenance phase in their careers. Likewise, raters age 40–49, being in the latter half of their establishment/advancement phase may feel less competition for promotions from females since historically so few females were in their beginning cohort. Verification of the speculation incorporated in this explanation represents an hypothesis for future research.

A second explanation for the finding that males in the 30–39 age group evaluated females higher than males on overall job performance may reside with the ratees themselves. It is quite likely that male raters age 30–39 may primarily be supervising entry level accountants. At this level, females may be perceived as more competent than males, possibly because females generally have higher undergraduatte grade point averages than males. Whether such an assessment reflects reality or is the result of a positive halo effect remains unknown.

Conclusion

Research on biases in performance evaluations have tended to focus on a limited number of factors (e.g. age, gender, race). The present results encourage two potential directions for future research in this area. First, in agreement with Dipboye (1985), they encourage development of a more holistic performance evaluation model which includes behavioral, cognitive, social, and affective factors. Second, the uncovering of a significant multivariate rater age × rater gender interaction suggests that only by taking an interactive perspective will it be possible to enrich our understanding of the complexity inherent in employee performance evaluations. Thus, research which combines nontask variables with behavioral, cognitive, social, and affective factors will likely remain in high demand.

References

Bernardin, H. J. and Beatty, R. W. (1984). *Performance Appraisal: Assessing Human Behavior at Work*, Kent, Boston.

Bernardin, H. J. and Villanova, P. (1986). 'Performance appraisal'. In: Locke, E. A. (Ed.) *Generalizing from Laboratory to Field Settings*, Lexington, Lexington, MA, pp. 43–62.

Cleveland, J. N. and Landy, F. J. (1981). 'The influence of rater and ratee age on two performance judgments', *Personnel Psychology*, **34**, 19–29.

Deets, C. and Froebe, D. (1984). 'Incentives for nurse employment', *Nursing Research*, **33**, 242–246.

DeNisi, A. S., Cafferty, T. P. and Meglino, B. M. (1984). 'A cognitive view of the performance appraisal process: A model and research propositions', *Organizational Behavior and Human Performance*, **33**, 360–390.

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

Dobbins, G. H. and Trahan, W. (1986). 'Sex bias in performance appraisal: A meta-analytic and critical review'. Paper presented at the meeting of the Academy of Management, Chicago, IL.

Hall, D. T. (1976). *Careers in Organizations*, Goodyear, Santa Monica, CA.

Hartman, S. J., Harris, O. J., Crino, M. and Griffeth, R. W. (1988). 'The effects of rater and canditate sex, candidate personal characteristics, and job type on assessment center evaluations'. Manuscript submitted for publication.

Mobley, W. H. (1982). 'Supervisor and employee race and sex effects on performance appraisals: A field study of adverse impact and generalizability', *Academy of Management Journal*, **25**, 598–606.

Rosenstein, J. and Hitt, M. A. (1986). 'Experimental research on race and sex discrimination: The record and the prospects', *Journal of Occupational Behaviour*, **7**, 215–226.

Waldman, D. A. and Avolio, B. J. (1986). 'A meta-analysis of age differences in job performance', *Journal of Applied Psychology*, **71**, 33–38.

Wendelken, D. J. and Inn, A. (1981). 'Nonperformance influences on performance evaluations: A laboratory phenomenon?' *Journal of Applied Psychology*, **66**, 149–158.

Williams, R. and Walker, J. (1985). 'Sex differences in performance rating: A research note; *Journal of Occupational Psychology,* **58**, 331–337.