

# When Fit Is Fundamental: Performance Evaluations and Promotions of Upper-Level Female and Male Managers

Karen S. Lyness

Baruch College, City University of New York

Madeline E. Heilman

New York University

Using archival organizational data, the authors examined relationships of gender and type of position (i.e., line or staff) to performance evaluations of 448 upper-level managers, and relationships of performance evaluations to promotions during the subsequent 2 years. Consistent with the idea that there is a greater perceived lack of fit between stereotypical attributes of women and requirements of line jobs than staff jobs, women in line jobs received lower performance ratings than women in staff jobs or men in either line or staff jobs. Moreover, promoted women had received higher performance ratings than promoted men and performance ratings were more strongly related to promotions for women than men, suggesting that women were held to stricter standards for promotion.

*Keywords:* gender bias, gender stereotypes, performance appraisal, promotion, glass ceiling

Although their presence in the management ranks is increasing, women continue to be underrepresented in senior management at many large private-sector companies (Lyness, 2002; Powell, 1999). One possible explanation is that there are gender differences in performance evaluations, and how women managers' performance is evaluated relative to men's performance influences women's subsequent career success. Yet, relatively little is known about whether or when performance evaluations of women in upper-level management jobs differ from those of their male counterparts, or if female managers' performance ratings differ depending on the types of jobs they hold. Moreover, it is not clear from prior research how performance evaluations are related to the actual career progress of men and women managers. The present study addresses these issues.

## Performance Evaluations

The perceived lack of person–job fit has been used to explain the occurrence of gender bias against women in organizational decisions about managers including performance evaluations. That is, the perceived incongruity between stereotypically based attributes ascribed to women (e.g., kind, caring, and relationship-oriented) and the attributes ascribed to men (e.g., tough, forceful, and achievement-oriented) believed to be necessary for success at

male gender-typed jobs, is thought to give rise to expectations that women will perform poorly in these positions, and the greater the perceived lack of fit, the more negative the expectations (Heilman, 1983, 1995, 2001). The lack-of-fit model further asserts that these expectations play a key role in evaluative processes because there is a tendency to perpetuate and confirm them. These expectations become the lens through which information is filtered, including what behavior is attended to, how that behavior is interpreted, and whether it is remembered when critical decisions are made. As a consequence, the negative expectations resulting from perceptions of lack of fit detrimentally affect how women are regarded and how their work is evaluated when they are in traditionally male jobs.

There is empirical evidence to support the lack-of-fit model as it relates to performance evaluations. A meta-analysis of laboratory studies indicated that there was greater gender bias against women in performance ratings on masculine tasks than on feminine tasks (Swim, Borgida, Maruyama, & Myers, 1989). Another meta-analysis that focused on evaluations of leaders indicated that the devaluation of female leaders was more pronounced for roles occupied mainly by men than for roles occupied more equally by both sexes (Eagly, Makhijani, & Klonsky, 1992).

Nevertheless, questions can be raised about whether these findings generalize to field settings and performance evaluations of managers. As Bartol (1999) has noted, there are relatively few prior field research studies about gender differences in performance evaluations, especially for managers, and a review of studies that have been done indicates that their results have been inconsistent. Moreover, because the prior field studies that have examined gender differences in managers' performance ratings either have been conducted in government settings (e.g., Powell & Butterfield, 1994) or have used self-reported ratings or ratings collected exclusively for research purposes for their analyses (e.g., Cannings & Montmarquette, 1991; Tsui & Gutek, 1984), they are not necessarily informative about gender differences in performance ratings that occur in business organizations. Therefore, our first objective for the present study was to test predictions based on

---

Karen S. Lyness, Department of Psychology, Baruch College, City University of New York; Madeline E. Heilman, Department of Psychology, New York University.

We thank Michael Judiesch for his helpful comments about an earlier version of the paper and Roger Millsap for his suggestions about analyses. In April 2002, portions of this study were presented at the Annual Meeting of the Society for Industrial and Organizational Psychology, Toronto, Canada.

Correspondence concerning this article should be addressed to Karen S. Lyness, Department of Psychology, Baruch College, City University of New York, One Bernard Baruch Way, Box B8-215, New York, NY 10010-5585. E-mail: Karen\_Lyness@baruch.cuny.edu

lack-of-fit ideas as they relate to actual performance evaluations of upper-level corporate managers.

Key to understanding the role of lack of fit in the different evaluations of male and female managers is the gender-typing of the job of manager. Jobs are thought to become gender-typed as *male* or *female* based on either (a.) job responsibilities that are believed to be gender-linked (Heilman, 1983, 1995) or (b.) the sex of the usual job-holder (Cejka & Eagly, 1999; Krefting, Berger, & Wallace, 1978). Managerial jobs, especially at upper levels, have traditionally been considered to be male gender-typed because the high levels of organizational authority, responsibility, and status that are characteristic of these jobs have typically been associated with men, rather than with women, in our society (e.g., Ragins & Sundstrom, 1989), thus adding to the manager-as-male image (e.g., Schein & Mueller, 1992; Schein, Mueller, Lituchy, & Liu, 1996). In addition, although women have made progress at moving into management positions in recent years, gender segregation of organizational hierarchies persists (Blau, Ferber, & Winkler, 1998; Jacobs, 1992), with women often concentrated in lower- and middle-level management positions rather than the more salient upper-level positions (Lyness, 2002; Powell, 1999). In fact, a recent study of the highest management levels found that women held less than 16% of the corporate officer positions in Fortune 500 companies (Catalyst, 2002). The view that managerial jobs are male gender-typed, first reported over 30 years ago (Schein, 1973, 1975), has persisted over time, with men and managers consistently being described more similarly than women and managers (Heilman, Block, Martell, & Simon, 1989; Powell, Butterfield, & Parent, 2002; Schein, 2001).

However, when superiors evaluate managers in actual organizational settings, we suspect that the category *manager* will not be so broadly drawn. According to recent research, people attend more to detail when they are close up than when they are psychologically remote from a target (Trope & Liberman, 2003). Therefore, it is likely that when direct supervisors are conducting performance evaluations, the global concept of manager will give way to more fine-grained conceptions. With their extensive knowledge about organizational job requirements, these superiors are likely to make distinctions among job definitions and, consequently, in their perceptions about the degree of male gender-typing for particular managerial positions.

In the case of upper-level management jobs, we expect that line positions are likely to be viewed as more strongly male gender-typed than staff positions. Not only is our reasoning based on the typical gender proportions of men and women in upper-level line and staff jobs—as recently as 2002, women corporate officers in Fortune 500 companies held less than 10% of the line jobs, and 70% of these women held staff jobs (Catalyst, 2002)—but also on the different work contents and organizational roles of line and staff jobs. Managers in line jobs direct and control essential organizational activities, such as producing or selling products or services, whereas managers in staff jobs provide support and expertise to line managers (Hellriegel, Jackson, & Slocum, 2002). Furthermore, managers in line positions typically have greater organizational power and influence than managers in staff positions (Ragins & Sundstrom, 1989). Thus, the responsibilities of line management jobs appear to be highly consistent with the forcefulness and achievement orientation associated with men, but

the responsibilities of staff jobs, even at managerial levels, appear to be less exclusively male in character.

Because line jobs are likely to be viewed as more strongly male gender-typed than staff jobs, they should give rise to a greater perceived lack of fit for women than staff jobs. Female managers in line jobs should therefore be more disadvantaged, receiving lower performance evaluations, not only when compared to male managers in line jobs, but also when compared to female managers in staff jobs. However, we would not anticipate similarly differential effects for evaluations of men in upper-level line jobs and staff jobs. Regardless of whether they are line roles or staff roles, upper-level managerial jobs are considered to be male in gender-type. Line roles and staff roles cause deviations in the degree of male gender-typing of a particular position, but they do not cause a managerial job to be perceived as *female*. As a result, there is always a fit between perceptions of male managers' attributes and perceived requirements of their jobs, and we do not expect that male managers' performance evaluations will differ depending on whether they hold line jobs or staff jobs.

Although we are not aware of any published research involving performance evaluations of women and men managers that has focused on line and staff jobs, there is a prior field study that provides some support for our ideas about the relationship of degree of male gender typing of jobs to performance ratings for men and women. Pazy and Oron (2001) examined relationships between gender composition of work groups and performance ratings of Israeli military officers in noncombat positions, and found that women's performance ratings varied but men's did not, depending on the proportion of women in the work unit. Women received lower performance ratings when they worked in groups with less than 10% women, and women's average performance ratings increased as the proportion of women in their work groups increased. These findings can be interpreted as showing that even in what the researchers termed "a male-gendered environment," (Pazy & Oron, 2001, p. 691) it was possible for jobs to differ in degree of perceived male gender-typing (as measured by the proportion of women in the work unit), which was reflected in women's performance ratings. In addition, Pazy and Oron's finding that male officers' average performance ratings were not related to gender composition of their work groups also is consistent with our contention that male jobs may differ in degree of perceived male gender-typing and yet still be perceived as congruent with male attributes.

Thus, based on theory and research about perceived job-gender incongruity, we predicted that performance ratings for upper-level managers would be related to the interaction of gender with type of managerial position (i.e., line or staff). Specifically, we expected that women in line jobs would receive lower performance ratings than all other groups:

*Hypothesis 1:* Female managers in line jobs will be evaluated more negatively than female managers in staff jobs and than male managers in both line and staff jobs.

#### Performance Evaluation and Promotion

The lack-of-fit model also suggests that differing levels of performance would be required of men and women if they were to be seen as worthy of a promotion to a male gender-typed position.

That is, greater evidence of competence is needed to overcome the negative performance expectations that besiege women (but not men) in male gender-typed job domains if they are to be judged as warranting advancement. Foschi's (1992, 1996, 2000) theory about double standards for competence is also relevant to this issue. According to Foschi, different standards are used in scrutinizing equal performance by a man and a woman to make judgments about competence, with performance by members of lower status groups (e.g., women) assessed by stricter standards than similar performance by members of higher status groups (e.g., men). Furthermore, double standards are thought to be particularly likely when the status characteristic (e.g., gender) is perceived to be relevant to performance, as in the case of male gender-typed jobs. These ideas suggest that in order to be promoted, women would have to outperform their male counterparts. Our second objective for the present study was to test this idea using a longitudinal design to examine the relationship between performance ratings and actual promotions received during the subsequent 2 years. We predicted:

*Hypothesis 2:* Women who have been promoted will have received higher performance ratings than men who have been promoted.

These ideas also suggest that men will be given more leeway when promotion decisions are made. That is, there will be more flexibility in standards and a greater tendency to give the benefit of the doubt to men than to women when their performance is not exemplary. Thus, we also expected that there would be a stronger relationship between performance evaluations and promotions for women than for men. Specifically, we predicted:

*Hypothesis 3:* Promotions will be related to the interaction of gender with performance ratings, such that performance ratings will be more strongly related to promotions for women than for men.

Because we did not have information about the types of subsequent jobs for which women and men were considered as promotion candidates, we could not use lack-of-fit principles to make predictions about job type as a factor in promotion decisions. We did, however, explore initial job type, that is, line or staff, as a potential moderator of the effects predicted in Hypotheses 2 and 3.

## Method

### Participants and Procedures

The sample included 489 upper-middle-level and senior-level managers from U.S. offices of a large multinational financial services corporation. Leaders of business units and staff functions evaluated these managers on nine performance dimensions as part of a succession planning process in the first quarter of 1998. There were 177 raters whose gender was identifiable, including 28 women and 149 men. (Rater gender was missing for 52 ratees.) Raters evaluated an average of two managers ( $M = 2.46$ , mode = 1). We computed subsequent promotions by comparing participants' levels in the management hierarchy on January 1, 1998 to their levels on January 1, 2000. We obtained all data from organizational databases.

Other than rater gender, complete 1998 data were available for 448 managers, and we used this sample to test our predictions. Twenty-two percent ( $n = 100$ ) of the sample was women, and the sample was rather

evenly divided between line managers ( $n = 225$ ) and staff managers ( $n = 223$ ). As expected, a larger proportion of the men (54%) than the women (37%) were in line positions,  $\chi^2(1, N = 448) = 9.00, p < .01$ . The average age of the women was 45 years versus 47 years for the men,  $t(446) = -2.44, p < .05$ , the women averaged about 11 years of organizational tenure versus about 14 years for the men,  $t(446) = -2.84, p < .01$ , and a larger proportion of the men (71%) than the women (61%) held graduate degrees,  $\chi^2(1, N = 448) = 3.60, p = .06$ . Forty-two percent of the women compared to 65% of the men held senior level management positions, and the others were in upper-middle-level management positions,  $\chi^2(1, N = 448) = 17.47, p < .001$ . We controlled for these demographic variables that covaried with gender by using age, organizational tenure, education, and whether managers were in senior-level management positions as control variables in the analyses.

### Measures

*Performance ratings.* Managers' performance was rated in 1998 on nine dimensions developed by the organization: operating results; customer effectiveness; personal, business, and technical proficiency; execution skills; leadership; professional standards; relationships; global effectiveness; and social responsibility. All performance ratings were made on three-point scales coded from 1 (*low*) to 3 (*high*). A confirmatory factor analysis with EQS for Windows 6.1 (Bentler & Wu, 1995) indicated that a one-factor model provided a good fit for the nine performance dimensions,  $\chi^2(27, N = 485) = 60.0, p < .001, \chi^2/df = 2.22$ , comparative fit index = .96, root mean square error of approximation = .05. A one-factor model is also consistent with recent meta-analytic research indicating that there is a general factor in ratings of job performance (Viswesvaran, Schmidt, & Ones, 2005). We created a composite performance scale by summing the nine dimension scores, and the alpha coefficient for the composite performance scale was .79.

*Job type.* Each participant's position was coded as *staff* (0) or *line* (1) based on information provided by organizational human resource officers. Examples of line functions included business management, operations management, and sales; examples of staff functions included human resources, administration, and external affairs.

*Gender.* Participant gender was coded *female* (0) or *male* (1).

*Human capital.* Participants' age and organizational tenure were measured in years, as of January 1, 1998. Education was measured as highest degree earned, coded *graduate degree* (1) and *bachelor's degree* (0). Thirty-seven participants had complete data except for education so we repeated our regression analyses without education as a control variable, using the larger sample ( $N = 485$ ), and reached the same conclusions.

*Organizational level.* Participants spanned five levels in the management hierarchy, coded from *low* (1) to *high* (5), with Level 1 considered upper-middle-level management and Levels 2 through 5 considered senior-level management, according to organizational human resource officers. For ease of presentation of results, we dichotomized organizational levels into whether or not participants held senior-level management positions on January 1, 1998, coded *yes* (1,  $n = 269$ ) or *no* (0,  $n = 179$ ). However, we also repeated our regression analyses with four dummy variables representing the five management levels (contrasting Levels 2 to 5 with Level 1), and reached the same conclusions about tests of our hypotheses.

*Promotions.* A promotion was defined as a move to a higher salary level in the management hierarchy. Promotions were calculated by subtracting each manager's organizational level (using the five original management levels) on January 1, 1998 from his or her level on January 1, 2000. Because only 2% of the sample moved up more than one level, we treated promotion as a dichotomous variable, coded *yes* (1,  $n = 77$ ) or *no* (0,  $n = 371$ ).

## Analyses

Preliminary regression analyses indicated that neither the main effect for rater gender,  $\beta = -.02, p = .62$ , nor the interaction between rater gender and ratee gender,  $\beta = -.06, p = .67$ , was significantly related to performance ratings so we combined the data for all raters in subsequent analyses. Similarly, a preliminary logistic regression analysis indicated that neither the main effect for rater gender,  $b = .21, p = .60$ , nor the interaction between rater gender and ratee gender,  $b = .46, p = .58$ , was significantly related to promotion decisions so we also combined those data for all raters.

We tested Hypothesis 1 by carrying out a linear regression analysis with the nine-dimension composite performance scale as the dependent variable, and controls for human capital (age, organizational tenure, education, and organizational level) in Step 1; gender in Step 2 (to see if there was a significant main effect); job type in Step 3; and the gender by job type interaction in Step 4. We then conducted three planned orthogonal contrasts, as suggested by Strube and Bobko (1989) for testing an ordinal interaction when it is predicted that differences are due to one cell, as in our case where we predicted that women in line jobs would receive lower performance ratings than the other three groups. To control for study-wise error, we set the significance level for each of the three contrasts at  $\alpha = .05/3 = .017$ .

We tested Hypothesis 2 with a linear regression analysis predicting composite performance ratings, with control variables and the main effects for job type in Step 1, and gender in Step 2, and we carried out an exploratory analysis of initial job type as a moderating variable by entering the gender by initial job type interaction in Step 3. We limited the sample for these analyses to the 77 managers who had received promotions. For all regression analyses we used simultaneous entry of all variables within each step, and significance of the results was determined from the significance of beta coefficients and change statistics.

Finally, we tested Hypothesis 3 with a logistic regression analysis because promotion was a dichotomous variable. We entered the previously used control variables and job type as an additional control variable (because promotional opportunities might vary in line and staff positions) in Step 1, gender in Step 2 (to see if there was a significant main effect), performance rating in Step 3, and the interaction of gender and performance rating in Step 4. We carried out an exploratory analysis of initial job type as a moderating variable by entering the additional two-way interactions in Step 5; and the three-way interaction of gender, performance ratings, and initial job type in Step 6. We centered the composite performance ratings in the logistic regression analyses to eliminate unnecessary multicollinearity between this variable and the interaction term (Cohen, Cohen, West, & Aiken, 2003). Then we conducted separate logistic regression analyses for female managers and male managers. Because some

managers left the organization during the 2 years after the performance evaluations, we conducted all analyses of promotion data first with the 1998 sample ( $N = 448$ ) and then with only managers who were employed for the entire 2-year period ( $n = 382$ ).

## Results

The means, standard deviations, and intercorrelations of the variables are shown in Table 1. Examination of these results shows that gender, in and of itself, was not significantly correlated with either the nine-dimension performance composite ( $r = -.01$ ) or promotion ( $r = -.03$ ).

### Performance Evaluations

The results of the linear regression analysis testing Hypothesis 1, predicting that the interaction of gender by job type would be related to composite performance ratings, are presented in Table 2. Among the control variables, there was a positive relationship of performance ratings to organizational tenure,  $\beta = .12, p < .05$ . The main effect for gender was not statistically significant, but we found a significant main effect for job type (i.e., line or staff),  $\beta = -.34, p < .01$ , indicating that managers in staff positions received higher ratings than managers in line positions. As we predicted, the gender by job type interaction was significantly related to performance ratings,  $\beta = .24, \Delta R^2 = .01, p < .05$ .

Figure 1 depicts the interaction of gender and job type in the mean composite performance ratings of female and male managers. Consistent with Hypothesis 1, this figure shows that women in line jobs received the lowest performance ratings of all four groups.

We conducted a set of three planned orthogonal contrasts to test our specific prediction that women managers in line jobs would be rated significantly lower than managers in the other three groups. The contrasts indicated that (a) performance ratings did not differ significantly for men in line jobs and men in staff jobs,  $F(1, 446) = 2.57, p = .110$ ; (b) performance ratings for women in staff jobs did not differ significantly from pooled ratings for men in line jobs and staff jobs,  $F(1, 446) = 2.64, p = .105$ ; and (c) performance ratings for women in line jobs were significantly lower than pooled ratings for the other three groups, including women in staff jobs and men in line jobs and staff jobs,  $F(1, 446) = 5.85, p =$

Table 1  
Means, Standard Deviations, and Correlations

Variable	<i>M</i>	<i>SD</i>	1	2	3	4	5	6	7	8
1. Performance rating composite	21.69	2.97	–							
2. Promotion <sup>a</sup>	0.17	0.38	.16**	–						
3. Age <sup>b</sup>	46.90	6.71	.03	–.12*	–					
4. Organizational tenure <sup>b</sup>	13.49	9.52	.09*	–.10*	.55**	–				
5. Education <sup>c</sup>	0.69	0.46	–.03	.01	.08	.02	–			
6. Senior management level <sup>a</sup>	0.60	0.49	.08	–.14**	.30**	.19**	.01	–		
7. Job type <sup>d</sup>	0.50	0.50	–.13**	–.06	–.15**	–.03	.07	.03	–	
8. Gender <sup>e</sup>	0.78	0.42	–.01	–.03	.11*	.12*	.09	.20**	.14**	–

Note.  $N = 448$ .

<sup>a</sup> 0 = no; 1 = yes. <sup>b</sup> In years. <sup>c</sup> 0 = bachelor's degree; 1 = graduate degree. <sup>d</sup> 0 = staff; 1 = line. <sup>e</sup> 0 = female; 1 = male.

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

Table 2  
Linear Regression Analyses Predicting Performance Ratings  
with Gender and Job Type

Step	Variable	Step $\beta$	Full model $\beta$	$\Delta R^2$	$F (df)$
1	Age	-.05	-.08		
	Organizational tenure	.11 <sup>†</sup>	.12*		
	Education <sup>a</sup>	-.03	-.03		
	Senior management level <sup>b</sup>	.08	.09 <sup>†</sup>	.02	1.76 (4,443)
2	Gender <sup>c</sup>	-.04	-.10	.00	.51 (1,442)
3	Job type <sup>d</sup>	-.14**	-.34**	.02**	8.82 (1,441)
4	Gender by job type <sup>1</sup>	.24*	.24*	.01*	4.51 (1,440)

Note.  $N = 448$ . Entries are standardized beta weights.

<sup>a</sup> 0 = bachelor's degree; 1 = graduate degree. <sup>b</sup> 0 = no; 1 = yes. <sup>c</sup> 0 = female; 1 = male. <sup>d</sup> 0 = staff; 1 = line.

<sup>†</sup>  $p < .10$ . \*  $p < .05$ . \*\*  $p < .01$ .

.016. These findings are therefore supportive of our reasoning that women in line jobs were more disadvantaged in performance ratings relative to women in staff jobs and men in both line jobs and staff jobs.

#### Performance Evaluation and Promotion

We tested Hypothesis 2, predicting that promoted women would have received higher performance ratings than promoted men, using performance rating as the dependent variable and limiting the analysis to managers who were promoted ( $n = 77$ ). These results, shown in Table 3, were supportive of our prediction. Gender was negatively related to performance ratings, indicating

that promoted female managers had higher performance ratings than promoted male managers, step  $\beta = -.26$ ,  $\Delta R^2 = .06$ ,  $p < .05$ , with gender explaining 6% of the variance in these ratings. When we restricted the sample to the 69 promoted managers who were employed for the entire 2-year period, we found similar results, with higher performance ratings among promoted women managers than promoted men managers, step  $\beta = -.29$ ,  $\Delta R^2 = .07$ ,  $p < .05$ . Our additional analyses showed that the gender by initial job type interactions were not statistically significant with either sample, indicating that initial job type did not have a moderating effect.

Figure 2 shows the mean composite performance ratings for female managers and male managers who were promoted and not promoted. Promoted women's composite performance ratings averaged over 1.5 points higher than promoted men's composite performance ratings ( $M = 23.89$  for women, 22.31 for men).

The results of the logistic regression analysis to test Hypothesis 3, predicting that subsequent promotions would be related more strongly to performance ratings for women than men, are presented in Table 4. Among the control variables, only senior-level management had a significant relationship to promotion, indicating that managers in senior management jobs were less likely to be promoted than those in lower-level jobs, step  $b = -.54$ , odds ratio = .58,  $p < .05$ . The main effect for gender was not significant, but (as would be expected) the main effect for performance ratings was positively related to promotion, step  $b = 1.60$ , odds ratio = 4.97,  $\chi^2(1, N = 448) = 13.34$ ,  $p < .001$ . The addition of the interaction of gender by performance ratings to the model in Step 4 resulted in a significant improvement to the model,  $\chi^2(1, N = 448) = 3.88$ ,  $p < .05$ , according to the likelihood ratio test, supporting our prediction. Although the Wald test indicated that the beta coeffi-

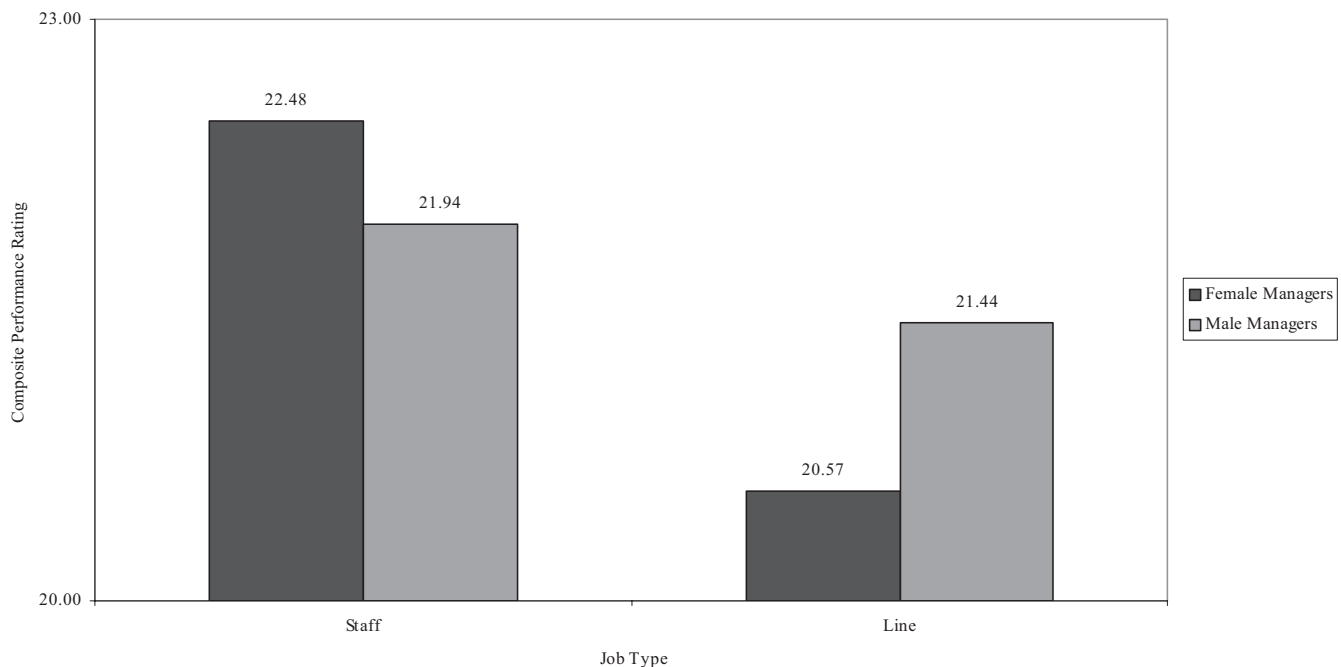


Figure 1. Mean composite ratings for female managers and male managers in staff jobs and line jobs are provided.

**Table 3**  
*Linear Regression Analysis Predicting Performance Ratings of Promoted Managers with Gender and Job Type*

Step	Variable	Step $\beta$	Full model $\beta$	$\Delta R^2$	$F (df)$
1	Age	-.15	-.14		
	Organizational tenure	.14	.14		
	Education <sup>a</sup>	-.09	-.14		
	Senior management level <sup>b</sup>	-.06	.00		
	Job type <sup>c</sup>	-.08	-.29	.05	.78 (5,71)
2	Gender <sup>d</sup>	-.26*	-.35*	.06*	4.95 (1,70)
3	Gender by job type	.32	.32	.02	1.29 (1,69)

*Note.*  $n = 77$  promoted managers. Entries are standardized beta weights.  
<sup>a</sup> 0 = bachelor's degree; 1 = graduate degree. <sup>b</sup> 0 = no; 1 = yes. <sup>c</sup> 0 = staff; 1 = line. <sup>d</sup> 0 = female; 1 = male.  
 \*  $p < .05$ .

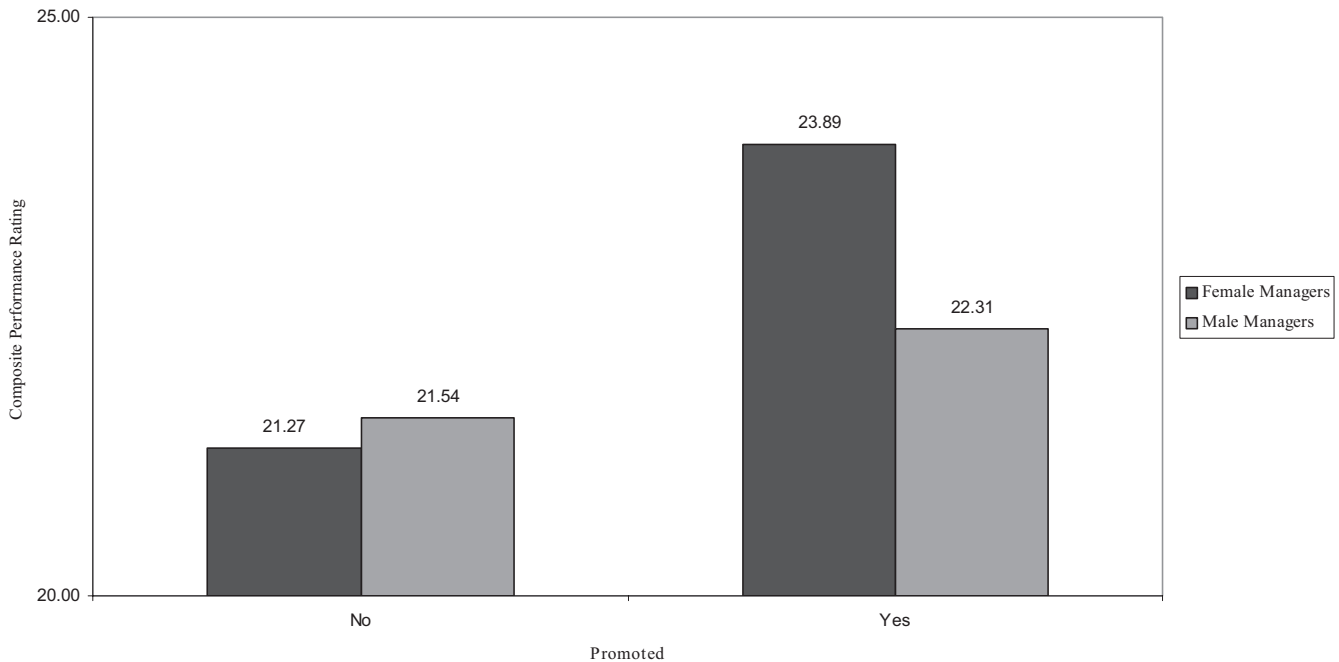
cient for the interaction term in this step was only marginally significant, step  $b = -2.28$ , odds ratio = .10,  $p = .07$ , the likelihood ratio test for change in model fit (which was statistically significant) is the preferred test for the impact of an individual predictor (Cohen, Cohen, West, & Aiken, 2003). We found even stronger support for our prediction when we repeated the analyses with the 382 managers who were employed for the entire 2-year period. With this sample, addition of the gender-by-performance ratings interaction resulted in a significant improvement to the model,  $\chi^2(1, N = 382) = 5.08, p < .05$ , according to the likelihood ratio test, and the Wald test also indicated that the beta coefficient for the interaction term was statistically significant, step  $b = -.33$ , odds ratio = .72,  $p < .05$ . Additional analyses indicated that

neither of the other two-way interactions involving initial job type nor the three-way interaction of gender, performance rating, and initial job type was statistically significant with either sample, indicating that the relationship of performance rating to promotion did not differ for managers who initially held line jobs or staff jobs.

We conducted separate logistic regression analyses for female managers and male managers (see Table 5), and found that, as predicted in Hypothesis 3, performance ratings were more strongly related to promotion for women,  $b = .43$ , odds ratio = 1.53,  $\chi^2(1, N = 100) = 13.57, p < .001$ , than for men,  $b = .12$ , odds ratio = 1.13,  $\chi^2(1, N = 348) = 4.73, p < .05$ . We found similar results for the managers who were employed for the entire 2 years, with a stronger relationship of performance ratings to promotion for women,  $b = .46$ , odds ratio = 1.59,  $\chi^2(1, N = 87) = 13.00, p < .001$ , than for men,  $b = .12$ , odds ratio = 1.12,  $\chi^2(1, N = 295) = 4.09, p < .05$ .

**Discussion**

These results suggest that gender bias can detrimentally affect the performance evaluations and promotional opportunities of upper-level women managers. The fact that our data were obtained from organizational archives rather than research instruments, and that they describe decisions made by senior managers who were accountable for their ratings and judgments, makes our findings particularly dramatic in their implications. Moreover, because the targets of the organizational evaluations and decisions about career advancement were upper-level managers, our results suggest that even when women are well on the way toward breaking through the glass ceiling, they face greater obstacles than their male counterparts.



*Figure 2.* Composite performance ratings for female managers and male managers who were promoted or not promoted are represented.

Table 4  
*Logistic Regression Analyses Predicting Promotion with Performance Ratings, Gender, and Job Type*

Step	Variable	Step			Full model <i>b</i>	$\chi^2$ ( <i>df</i> )
		<i>b</i>	<i>SE</i>	Odds ratio		
1	Age	-.03	.02	.97	-.02	14.07 (5)*
	Organizational tenure	-.01	.02	.99	-.02	
	Education <sup>a</sup>	.16	.28	1.18	.16	
	Senior management level <sup>b</sup>	-.54*	.27	.58	-.68*	
	Job type <sup>c</sup>	-.37	.26	.69	.18	
2	Gender <sup>d</sup>	.09	.31	1.10	.73	.09 (1)
3	Performance rating	1.60**	.47	4.97	4.06*	13.34 (1)***
4	Gender by performance rating	-2.28†	1.24	.10	-3.51*	3.88 (1)*
5	Performance rating by job type	.69	.95	2.00	-1.52	
	Gender by job type	-.13	.70	.88	-.48	.61 (2)
6	Gender by performance rating by job type	2.62	2.55	13.74	2.62	1.03 (1)

Note. *N* = 448. Entries are unstandardized beta weights.

<sup>a</sup> 0 = bachelor's degree; 1 = graduate degree. <sup>b</sup> 0 = no; 1 = yes. <sup>c</sup> 0 = staff; 1 = line. <sup>d</sup> 0 = female; 1 = male.

† *p* < .10. \* *p* < .05. \*\* *p* < .01. \*\*\* *p* < .001.

We found that performance evaluations were affected by the combination of the manager's gender and the type of position he or she held. Specifically, women in line jobs were the most negatively evaluated of all managers in our sample. Not only were women in line jobs rated lower than men in either line jobs or staff jobs, but they were also rated less favorably than other women holding staff jobs. Because line management positions are likely to be perceived as more strongly male gender-typed than staff management positions, these results support lack-of-fit ideas, which specify that more negatively biased evaluations of women managers will occur when there is a greater perceived lack of fit between job requirements and attributes of women. These data also support our contention that although management jobs are generally male gender-typed, especially at high levels in the organizational hierarchy, in actual work situations superiors make differentiations among these jobs in degree of perceived male gender-typing.

In addition, our data showed that performance evaluations had career consequences for the managers. Specifically we found that the composite performance ratings were related to actual promotions managers received during the 2-year period subsequent to the evaluations, after controlling for age, organizational tenure, edu-

cation, and organizational level. Moreover, we found that promoted women had received higher performance ratings than promoted men, suggesting that women had to be regarded as more stellar in their accomplishments than men if they were to be promoted. In addition, we found that performance ratings had more direct career consequences for female managers than male managers. Our data demonstrated that women's performance ratings were more strongly related to promotions than were men's, indicating that the standards for promotion of men were more flexible than the standards for promotion of women. These findings, which support our predictions, are consistent both with the lack-of-fit model (Heilman, 1983, 1995, 2001) and Foschi's (1992, 1996, 2000) theory about double standards for inferring competence, with lower status individuals (i.e., women) held to stricter standards and requiring more evidence of competence than higher status individuals (i.e., men). They also are consistent with anecdotal claims that women have to work harder to get to the same place, doing more and doing it better than men in similar positions.

The findings concerning line positions have important implications for female managers who aspire to advance in their organizations. Women are often advised to seek line positions in order to advance their careers. In fact, a recent survey of female executives

Table 5  
*Logistic Regression Analyses Predicting Promotion for Female Managers and Male Managers*

Step	Variable	Female managers					Male managers				
		Step	<i>SE</i>	Odds ratio	Full model <i>b</i>	$\chi^2$ ( <i>df</i> )	Step	<i>SE</i>	Odds ratio	Full model <i>b</i>	$\chi^2$ ( <i>df</i> )
1	Age	.00	.05	1.00	.02		-.04	.03	.96	-.04	
	Organizational tenure	-.01	.04	.99	-.04		-.01	.02	.99	-.01	
	Education <sup>a</sup>	1.07†	.63	2.92	1.20†		-.16	.32	.85	-.15	
	Senior management level <sup>b</sup>	-.88	.59	.41	-1.22†		-.46	.31	.63	-.54†	
	Job type <sup>c</sup>	-.41	.60	.67	.23	6.85 (5)	-.30	.30	.74	-.24	11.99 (5)*
2	Performance rating	.43**	.14	1.53	.43**	13.57 (1)***	.12*	.06	1.13	.12*	4.73 (1)*

Note. *n* = 100 female managers; 348 male managers. Entries are unstandardized beta weights.

<sup>a</sup> 0 = bachelor's degree; 1 = graduate degree. <sup>b</sup> 0 = no; 1 = yes. <sup>c</sup> 0 = staff; 1 = line.

† *p* < .10. \* *p* < .05. \*\* *p* < .01. \*\*\* *p* < .001.

and (mostly male) CEOs of Fortune 500 companies found that lack of line management experience was cited as the top barrier to advancement for women (Wellington, Kropf, & Gerkovich, 2003). However, our results have particularly troublesome implications for women in upper-level line management positions. Not only were they, along with staff women, subjected to more stringent standards than men when promotion decisions were made, both in terms of the level of performance required and the centrality of their ratings in determining whether they received promotions, but line women's performance was also evaluated more negatively than that of all other managers. Thus, the women in line management positions were especially vulnerable to gender-based bias. Given the importance of line management experience for career advancement, these findings sound a warning for women seeking to attain the highest corporate management positions. Line positions may be critical for advancement, but they also are rife with potential hazards for women.

This research has important implications for organizations. Our findings indicate that there can be gender bias in performance evaluations of managers even when there is not a significant main effect for gender as a predictor. Instead, the gender differences that we observed were not universal, and only were apparent when female managers held line jobs. This does not, however, mean that gender bias did not occur, or that its severity was not great. Because of the complexity of gender bias and what appears to be its basis in perceptions of lack of fit, its occurrence depends heavily on the nature of a position and its degree of perceived male gender typing. Thus, if organizations only concern themselves with aggregate ratings, they may overlook the incidence of gender bias, or conclude that it is absent when it is not. Similarly, our promotion findings showed that even though we did not find a significant main effect for gender, women were held to stricter standards than men for promotion. Such subtle indications of gender bias could easily be overlooked by organizations unless they conduct detailed analyses of gender differences in performance ratings required for promotion and examine relative strengths of relationships between performance ratings and promotions for women and men.

Organizations should also be interested in our finding that the relationship between performance ratings and promotions was weaker for men than for women. This finding raises important questions about why this relationship was weaker for men, and what factors, besides performance, might have influenced organizational decisions to promote men. Possible answers are provided by other research showing that male managers made greater use than female managers of informal means for career advancement, such as networks with powerful sponsors or mentors (Cannings & Montmarquette, 1991; Lyness & Thompson, 2000). We urge organizational decision makers to ensure that important rewards, such as promotions, are tightly linked to successful performance for men as well as women.

When interpreting these data and drawing conclusions, care should be taken to consider the limitations of the study. Although our use of archival organizational data is a strength of the research, an accompanying limitation is that we cannot rule out the possibility that the differences in performance ratings were based on real performance differences, not on biased assessments. The idea that women in line positions are in fact lower performers seems unlikely, however, given that our data show that in order to

advance, women managers generally had to perform better than their male counterparts. Moreover, although our data presented a unique opportunity to learn about ratings and promotion decisions concerning upper-level men and women managers, they are from a single organization, and that may limit generalizability of our results. However, this limitation may be offset to some extent by the fact that the focal managers held a wide variety of line and staff jobs in many different business units and locations, and 177 raters made the performance evaluations. In fact, given the diversity of our participants and the many uncontrollable factors that might influence actual ratings of upper-level managers, the consistency of our findings is noteworthy. Another limitation was the somewhat short 2-year time period for measuring promotions, suggesting that our study provided a relatively conservative test of predictions about promotions. Finally, more complete information about promotion candidacies and the types of jobs involved would have allowed us to better explore effects of gender by job type interactions in predicting promotion decisions.

Despite the limitations of our study, the results have important implications for how future research should be conducted about gender bias in performance evaluations. Our study underscores the need for future researchers to examine specific job content and demographic composition of jobholders in order to assess perceived gender typing of jobs. When these issues are not taken into account, as in some prior studies about gender differences in managers' performance evaluations, our findings provide a plausible explanation for the lack of pro-male bias that was reported (e.g., Tsui & Gutek, 1984). In fact, based on our results and lack-of-fit ideas, we would not expect pro-male bias in ratings of performance unless jobs are strongly male gender-typed. It is not the manager's gender alone that gives rise to bias but rather the perceived misfit of gendered attributes and job requirements. It is important, therefore, that future research about gender differences in performance evaluations take gender type of specific jobs into account rather than assuming that all managerial positions are perceived as equally male gender-typed.

Our findings also raise questions for future research. Although there have been some prior studies about the advancement strategies of successful women executives (e.g., Lyness & Thompson, 2000; Ragins, Townsend, & Mattis, 1998), we need to learn more about strategies that women in strongly male gender-typed positions have used to overcome the perceived gender incongruity with job requirements that appear to be so detrimental to women's career progress. Additionally, it would be interesting to explore the apparently different processes used in going from performance evaluations to promotion decisions for men and women managers, and to determine whether there are instances when the standards held for men are equally rigorous as those held for women.

The results of this study are important for those who are concerned about the advancement of women in the workforce. Our data strongly suggest that gender bias affects performance evaluations of managers in actual organizational settings, and illustrate the different consequences of performance evaluations for promotion of women and men in managerial roles. Although the results suggest that it is only women in line positions whose performance was evaluated in a biased manner, the results make evident that all women managers, whatever their jobs, were subjected to more strict and less flexible standards than their male counterparts when promotion decisions were made. The fact that this bias was evident



in decision making about upper-level managers implies that even with women's steady progress up the ranks of organizations, potential for their differential treatment persists.

### References

- Bartol, K. M. (1999). Gender influences on performance evaluations. In G. Powell (Ed.), *Handbook of gender and work* (pp. 165–178). Thousand Oaks, CA: Sage.
- Bentler, P., & Wu, E. J. C. (1995). *EQS for Windows user's guide* (Version 6.0). Encino, CA: Multivariate Software, Inc.
- Blau, F. D., Ferber, M. A., & Winkler, A. E. (1998). *The economics of women, men, and work* (3rd ed.). Saddle River, NJ: Prentice Hall.
- Cannings, K., & Montmarquette, C. (1991). Managerial momentum: A simultaneous model of the career progress of male and female managers. *Industrial and Labor Relations Review*, *44*, 212–228.
- Catalyst. (2002). *2002 Catalyst census of women corporate officers and top earners of the Fortune 500*. New York: Catalyst.
- Cejka, M. A., & Eagly, A. H. (1999). Gender-stereotypic images of occupations correspond to the sex segregation of employment. *Personality and Social Psychology Bulletin*, *25*, 413–423.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah, NJ: Erlbaum.
- Eagly, A. H., Makhijani, M. G., & Klonsky, B. G. (1992). Gender and the evaluation of leaders: A meta-analysis. *Psychological Bulletin*, *111*, 2–33.
- Foschi, M. (1992). Gender and double standards for competence. In C. L. Ridgeway (Ed.), *Gender, interaction, and inequality* (pp. 181–207). New York: Springer-Verlag.
- Foschi, M. (1996). Double standards in the evaluation of men and women. *Social Psychology Quarterly*, *59*, 237–254.
- Foschi, M. (2000). Double standards for competence: Theory and research. *Annual Review of Sociology*, *26*, 21–42.
- Heilman, M. E. (1983). Sex bias in work settings: The lack of fit model. In B. M. Staw & L. L. Cummings (Eds.), *Research in Organizational Behavior* (Vol. 5, pp. 269–298). Greenwich, CT: JAI Press.
- Heilman, M. E. (1995). Sex stereotypes and their effects in the workplace: What we know and what we don't know. *Journal of Social Behavior and Personality*, *10*, 3–26.
- Heilman, M. E. (2001). Description and prescription: How gender stereotypes prevent women's ascent up the organizational ladder. *Journal of Social Issues*, *57*, 657–674.
- Heilman, M. E., Block, C. J., Martell, R. F., & Simon, M. C. (1989). Has anything changed? Current characterizations of men, women, and managers. *Journal of Applied Psychology*, *74*, 935–942.
- Hellriegel, D., Jackson, S. E., & Slocum, J. W., Jr. (2002). *Management: A competency-based approach*. Cincinnati, OH: South-Western.
- Jacobs, J. A. (1992). Women's entry into management: Trends in earnings, authority, and values among salaried managers. *Administrative Science Quarterly*, *37*, 282–301.
- Krefting, L. A., Berger, P. K., & Wallace, M. J., Jr. (1978). The contribution of sex distribution, job content, and occupational classification to job sex typing: Two studies. *Journal of Vocational Behavior*, *13*, 181–191.
- Lyness, K. S. (2002). Finding the key to the executive suite: Challenges for women and people of color. In R. Silzer (Ed.), *The 21st century executive: Innovative practices for building leadership at the top* (pp. 229–273). San Francisco: Jossey-Bass.
- Lyness, K. S., & Thompson, D. E. (2000). Climbing the corporate ladder: Do female and male executives follow the same route? *Journal of Applied Psychology*, *85*, 86–101.
- Pazy, A., & Oron, I. (2001). Sex proportion and performance evaluation among high-ranking military officers. *Journal of Organizational Behavior*, *22*, 689–702.
- Powell, G. N. (1999). Reflections on the glass ceiling: Recent trends and future prospects. In G. N. Powell (Ed.), *Handbook of gender and work* (pp. 325–345). Thousand Oaks, CA: Sage.
- Powell, G. N., & Butterfield, D. A. (1994). Investigating the "Glass Ceiling" phenomenon: An empirical study of actual promotions to top management. *Academy of Management Journal*, *37*, 68–86.
- Powell, G. N., Butterfield, D. A., & Parent, J. D. (2002). Gender and managerial stereotypes: Have the times changed? *Journal of Management*, *28*, 177–193.
- Ragins, B. R., & Sundstrom, E. (1989). Gender and power in organizations. *Psychological Bulletin*, *105*, 51–88.
- Ragins, B. R., Townsend, B., & Mattis, M. (1998). Gender gap in the executive suite: CEOs and female executives report on breaking the glass ceiling. *Academy of Management Executive*, *12*(1), 28–42.
- Schein, V. E. (1973). The relationship between sex role stereotypes and requisite management characteristics. *Journal of Applied Psychology*, *57*, 95–100.
- Schein, V. E. (1975). Relations between sex role stereotypes and requisite management characteristics among female managers. *Journal of Applied Psychology*, *60*, 340–344.
- Schein, V. E. (2001). A global look at psychological barriers to women's progress in management. *Journal of Social Issues*, *57*, 675–688.
- Schein, V. E., & Mueller, R. (1992). Sex role stereotyping and requisite management characteristics: A cross cultural look. *Journal of Organizational Behavior*, *13*, 439–447.
- Schein, V. E., Mueller, R., Lituchy, T., & Liu, J. (1996). Think manager—think male: A global phenomenon? *Journal of Organizational Behavior*, *17*, 33–41.
- Strube, M. J., & Bobko, P. (1989). Testing hypotheses about ordinal interactions: Simulations and further comments. *Journal of Applied Psychology*, *74*, 247–252.
- Swim, J., Borgida, E., Maruyama, G., & Myers, D. G. (1989). Joan McKay versus John McKay: Do gender stereotypes bias evaluations? *Psychological Bulletin*, *105*, 409–429.
- Trope, Y., & Liberman, N. (2003). Temporal construal. *Psychological Review*, *110*, 403–421.
- Tsui, A. S., & Gutek, B. A. (1984). A role set analysis of gender differences in performance, affective relationships, and career success of industrial middle managers. *Academy of Management Journal*, *27*, 619–635.
- Viswesvaran, C., Schmidt, F. L., & Ones, D. S. (2005). Is there a general factor in ratings of job performance? A meta-analytic framework for disentangling substantive and error influences. *Journal of Applied Psychology*, *90*, 108–131.
- Wellington, W., Kropf, M. B., & Gerkovich, P. R. (2003). What's holding women back? *Harvard Business Review*, *81*(6), 18–19.

Received December 12, 2004

Revision received May 9, 2005

Accepted June 21, 2005 ■