ABSTRACT
        Relying on estimates of variance explained to assess
the practical significance of research findings is now common
practice. It is suggested that such estimates can offer an inaccurate
picture, often underestimating the practical significance of
statistically small effects. As an example, research on employment
discrimination indicates that in non-traditional work settings women
are generally judged less favorable than men. However, because the
amount of variance due to sex is typically quite modest, the actual
importance of sex bias effects has been questioned. Two numerical
examples with 600 men and 600 women, and with 100 men and 100 women,
respectively, demonstrate that even very small amounts of sex bias in
hiring decisions and performance evaluations can have profoundly
negative consequences for women. Computer simulations confirm this
conclusion. It is hoped that researchers will not automatically
discount the practical consequences of statistically small effects.
Five tables and one figure illustrate the analyses. (Contains 32
references.) (SLD)

ED 361 398

A Little Sex Bias Can Hurt Women A Lot

Richard F. Martell

Teachers College

Columbia University

David M. Lane and Cynthia E. Willis

Rice University

Running Head: SEX BIAS

ABSTRACT

Relying on estimates of "variance explained" (e.g., r^2, ω^2, ϕ^2) to assess the practical significance of one's research findings is now common practice. We believe, however, that such estimates can offer an inaccurate picture--often underestimating the practical significance of statistically small effects. As just one example, research on employment discrimination indicates that in nontraditional work settings women are generally judged less favorably than men; however, because the amount of variance due to sex is typically quite modest, the actual importance of sex bias effects has been questioned. In this paper we demonstrate that even very small amounts of sex bias in hiring decisions and performance evaluations can have profoundly negative consequences for women. In so doing, we hope to discourage researchers from automatically discounting the practical consequences of statistically small effects.

A Little Sex Bias Can Hurt Women A Lot

Over the past two to three decades, psychologists have become more and more concerned with assessing the <u>practical</u> significance of their research findings. Accordingly, researchers now turn to a variety of effect-size measures (e.g., r^2, ω^2, ϕ^2) which estimate the magnitude of an experimental effect in terms of "variance explained." It is important to recognize, however, that these measures do not in and of themselves reveal the practical significance of an effect. Precisely how much variance must be explained by an independent variable before it can qualify as practically significant is not at all obvious. John Campbell (1990, pp. 56-57) put the issue best when he recently asked: " ... by what metric or measurement model is this estimate [of variance explained] deemed to have meaning? What's high and what's low, and for what research issues?"

The confusion caused by a lack of a criterion for determining the importance of an effect was documented by Abelson (1985) who demonstrated that the proportion of variance explained by a variable does not necessarily mesh with people's intuition of the importance of the effect. Specifically, he found that batting average does not explain much variance in whether or not a batter gets a hit. In one of his scenarios, a baseball manager scans the bench to choose a pinch hitter. There are two choices: a 320 hitter and a 220 hitter. Abelson showed that only 1.3% of the variance in the outcome (making the simplifying assumption that the batter does not walk, get hit by a pitch, etc.) can be explained by the choice of hitters. Does this mean that it does not make much difference which batter is chosen? Clearly not, since the 320 hitter has almost a 50% greater probability of getting a hit.

Abelson argued that the cumulative importance of events must be taken into account in interpreting the estimate of effect size as follows:

In the present context, attitude toward explained variance ought to be conditional

on the degree to which the effects of the explanatory variable cumulate in practice (p. 133).

Although cumulative effects can be important, an appeal to cumulative effects of variables does not appear to provide a solution to this paradox. In the example, assume that there are two outs in the

bottom of the ninth inning and, for simplicity, that the home team will win if the batter gets a hit and will lose if the batter does not. Under these circumstances, the chances of winning this particular game are almost 50% higher with the 320 hitter than with the 220 hitter. However, there is nothing cumulative in this example; the outcome depends on a single event. Our resolution to the paradox is that in many contexts, including this one, the percentage of variance explained is simply a very misleading measure of the importance of the effect. We believe that many researchers are currently being misled by their estimates of effect size and, as a result, reaching incorrect conclusions about the importance of their independent variables. In this paper we focus on a research topic that has substantial public policy implications--employment discrimination--as an example.

Much research on employment discrimination indicates that in nontraditional work settings women are generally hired less frequently and their work performance judged less favorably than men. However, the amount of variance accounted for by sex is modest, typically less than 10%. These small effects have given way to the belief that the effects of an individual's sex on personnel decisions are of little or no practical significance (Borman, White, Pulakos, & Oppler, 1991; Latham, 1986; Olian, Schwab, & Haberfeld, 1988; Peters, O'Connor, Weekly, Pooyan, Frank, & Erenkrantz, 1984; Pulakos, White, Oppler, & Borman, 1989). Most recently, this issue was raised in response to the American Psychological Association Amicus Curiae Brief (APA, 1988; also, see Fiske, Bersoff, Borgida, Deaux, & Heilman, 1991) which reviewed research on sex stereotyping and discrimination and was used to support a claim of sex discrimination in the Price Waterhouse v Hopkins (1989) Supreme Court case. In a paper highly critical of the APA brief, Barrett and Morris (in press) pointed out, among other things, that the effects of sex on personnel decisions are generally quite small, a fact not mentioned in the brief. In overlooking the small magnitude of sex effects, Barrett and Morris argued that the problem of sex discrimination in the work place may have been exaggerated. In this paper we will provide several graphic demonstrations of how even "small amounts of sex bias" in hiring decisions and performance evaluations can have profoundly negative consequences for women. In so doing, we hope to

discourage researchers from automatically discounting the practical consequences of statistically

small effects. This is especially important insofar as effect sizes in many areas of psychology tend

to be rather small (see O'Grady, 1982; Sechrest & Yeaton, 1982 for an excellent discussion of the

factors that limit the magnitude of effects in psychological research).

Sex bias in hiring decisions

There has been a great deal of research investigating the treatment of women seeking entry

into nontradiiional occupations. According to the results of a recent meta-analysis of this literature,

which included 19 studies and 1842 subjects, male applicants were preferred over identically

qualified female applicants (Olian, Schwab, & Haberfeld, 1988). Yet, the magnitude of the effect

was quite small. Overall, applicant sex accounted for between only 4% to 9% of the variance in

hiring recommendations. The larger mean estimate was obtained using within-subject designs

which, because hiring decisions are usually made from a pool of applicants, is probably the more

appropriate design. Nonetheless, even a mean estimate of 9% is still considered small. In

contrast, the mean effect of applicant qualifications (e.g., education, experience, test scores) on

hiring recommendations accounted for 35% of the variance. Contemplating the practical

significance of the effects of applicant sex versus objective qualifications on hiring

recommendations, Olian et al. (1988, p. 180) concluded that " ... there is marginal evidence of

employment discrimination against females in experimental studies of hiring decisions." It is our

contention that effects of this magnitude can lead to substantial differences in the hiring rates of

men versus women and, thus, should not be so easily dismissed.

When characteristics of the hiring situation are taken into account, statistically significant

effects that explain only a small percentage of variance in hiring decisions can have enormous

practical consequences. For instance, it is well known that when the selection ratio (the proportion

of applicants to be hired) is low, as is often the case, selection tests with only a modest degree of

validity can still have salutary effects on hiring decisions (Cascio,1991; Hunter & Hunter, 1984;

Hunter & Schmidt, 1982; Schmidt, Hunter, McKenzie, & Muldrow, 1979). Analagously, very

small effects--in this case, a bias against women--can have enormous consequences when selection

ratios are taken into account. In a series of demonstrations outlined below we show how even small sex effects can cause women to be hired at substantially lower rates than men.

Demonstration 1

Begin with a pool of 1200 identically qualified applicants, 600 men and 600 women, and suppose 480 men (80%) and 310 women (52%) are hired. What are we to make of this difference in hiring rates? A chi-square test reveals that significantly more men than women were hired ($\chi^2 = 107.06$, p < .001). Indeed, because the hiring rate for women is less than 80% of the hiring rate for men, the "4/5ths" rule (U.S. EEOC, 1978) has been violated. This demonstration of "adverse impact" could quite properly expose our hypothetical organization to charges of sexual discrimination. Who could argue with the practical significance of this difference in hiring rates of men and women? Yet, how much variance in hiring decisions is due to applicant sex? A simple calculation of ϕ (a product-moment correlation used when both the independent and dependent variables are dichotomous, see Rosenthal & Rosnow, 1991) reveals a correlation between sex and hiring decisions of .30. That is, only 9% of the variance in hiring decisions is due to applicant sex. Now, suppose 480 men (80%) and 370 women (62%) are hired. A chi-square test reveals that significantly more men than women were hired ($\chi^2 = 48.79$, p < .001). Again, the "4/5ths" rule has been violated. Yet, a calculation of ϕ reveals a .20 correlation between sex and hiring decisions, only 4% of the variance is due to applicant sex. In fact, it can be seen in Table 1 that, as the selection ratio decreases, exceedingly tiny sex bias effects can violate the "4/5ths" rule.

Demonstration 2

Now, consider a much smaller pool of 200 identically qualified applicants, 100 men and 100 women. Suppose 80 men (80%) and 62 women (62%) are hired. A chi-square test reveals that significantly more men than women were hired ($\chi^2 = 7.86$, p < .01). Again, the "4/5ths" rule has been violated. Yet, a calculation of ϕ reveals a .20 correlation between sex and hiring decisions, only 4% of the variance is due to applicant sex. Even with this smaller applicant pool, it can be seen in Table 2 that small sex bias effects can violate the "4/5ths" rule, especially as the selection ratio decreases.

7

Sex bias in performance ratings

Research on the treatment of women who have gained entry into traditionally male
occupations reveals that their work performance is often judged less favorably then that of men,
even when performance is held constant (Heilman, 1983; Martell, 1991; Martell, in press; Sackett,
Dubois, & Noe, 1991). Yet, as was true with hiring decisions, the amount of variance in
performance ratings due to ratee sex is usually less than 10 percent, most often ranging from only
1 to 5 percent. Here too, these small effects have given way to the suggestion that sex
discrimination in performance appraisals are of little or no practical concern (Borman, White,
Pulakos, & Oppler, 1991; Latham, 1986; Peters, O'Connor, Weekly, Pooyan, Frank, &
Erenkrantz, 1984; Pulakos, White, Oppler, & Borman, 1989). For example, Latham's (1986, p.
133) review of the performance appraisal literature concluded that: " ... bias does not appear to be
a function of [ratee] sex ... Further research on this subject would appear unproductive in light of
the small criterion variance accounted for in the appraisal decision ... " In contrast, we will
demonstrate that such small sex effects are not trivial, and that a systematic bias of this magnitude
can severely hamper the upward mobility of women in organizations.

To appreciate how this can be, it is important to consider the structure of most work
organizations and the long-term consequences of early career performance assessments. First,
organizations usually are "pyramid" shape and, thus, there are increasingly fewer positions
available as one attempts to climb to the top. Consequently, as the promotion rate decreases at each
higher level, resulting in only the very best being promoted to the next level, even statistically small
sex effects in performance ratings can have large practical consequences. Second, most
organizations rely on a "tournament model" of career mobility in which early career success is a
precondition for future advancement (Schein, 1978; Van Maanen, 1977). Not surprisingly, early
career performance assessments have been found to strongly predict whether one reaches a top
management position (Rosenbaum, 1979). Accordingly, judging a woman's work performance
less favorably than a man's early on in her career (even just a little) is likely to serve as a constant
impediment, drastically limiting her upward progress. Thus, both an increasingly lower promotion

rate and the long-term consequences of early career performance assessments--factors overlooked when interpreting the results of "single-shot" studies--can exacerbate the effects of a small but systematic sex bias in performance evaluations. Several computer simulations were conducted to demonstrate the harmful effect of even small amounts of sex bias on the promotion rates of women.

Computer simulations

The simulations begin with an equal number of men and women awaiting promotion to the next level in the organization. Each person is assigned a performance evaluation score. We make the simple assumption that incumbents with the highest performance scores become eligible for promotion once a position arises. There are 8 levels in the organization and at each successive level there are fewer positions, ranging from 500 incumbents at the bottom to only 10 at the top level (See Table 3). The simulation begins by randomly removing 15% of the present incumbents from throughout the 8 levels. These positions are then filled from within the organization. Eligible individuals (those with the highest performance evaluation scores) are promoted into the position. The simulation continues to apply the 15% attrition rule until the organization is staffed entirely with "new" employees. That is, all incumbents present within the organization at the start of the simulation have now been replaced with men and women drawn from the initial pool of 1200. For each simulation, 20 computer runs were conducted to ensure an adequate degree of reliability.

The population distributions of performance evaluation scores of men and women were normal and identical ($\mu=50$, $\sigma=10$), with one exception. In Simulation 1, 6.29 "bias points" were added to the score of each man. After the bias was added, sex differences explained 9% of the variance in performance evaluation scores. In Simulation 2, 4.58 "bias points" (equivalent to an effect size of 5% of the variance) were added to the score of each man; in Simulation 3, 2.01 "bias points" (equivalent to an effect size of 1% of the variance) were added.

Detailed results are shown in Tables 3 to 5; the main findings are highlighted in Figure 1. An inspection of Figure 1 reveals that a very high percentage of upper-level jobs were filled by men. With 9% of the variance in performance evaluations due to sex, only 19% of the incumbents

9

at the top level of the organization were women. Even more dramatic is the finding that when sex differences explain only a "trivial" 1% of the variance, only 35% of the highest-level jobs were filled by women.

Discussion

These demonstrations reveal that effects that would normally be judged as trivial based on current interpretations of effect size measures can have dramatic consequences: sex accounted for very little variance, yet women were both hired and promoted at substantially lower rates than men.

The demonstrations presented here were not meant to model the complexity of actual hiring and promotion decisions; instead, they make the point that statistical indices of variance explained can be misleading. Nonetheless, it is reasonable to ask whether the sex composition of our hypothetical organization reflects reality to any reasonable degree. Recent research on the so-called "Glass Ceiling" phenomenon indicates that, indeed, women have been largely unsuccessful in their attempts to break into executive-level positions (Morrison, White, & Van Velsour, 1987). The proportion of top management positions held by women is less than 5% (U.S. Department of Labor, 1989). Thus, the computer simulation results point to biased performance evaluations as at least one reason why women remain underrepresented at upper levels of management.

Also, our demonstration that even small sex effects can lead to substantially lower hiring rates of women is entirely consistent with the results of a laboratory investigation of the hiring of men versus women managers (Dipboye, Fromkin, & Wiback, 1975). Although women in this study were rated only a little less favorably than men (accounting for but 1% of the variance in evaluative ratings), a man was choosen for the one available position 72% of the time. This supports our point that when selection ratios are low even a small bias against women can greatly reduce the probability of a woman being hired.

As we have demonstrated, it can be misleading to assess the the magnitude of an experimental effect apart from the "context" in which it occurs. Researchers who do so risk underestimating the practical significance of their findings. This message holds true not only in sex bias research but in other areas of study as well. For example, Rosenthal (1990; 1983) has

10

shown that whether looking at the effects of aspirin on heart attack rates or of psychotherapy on mental health even quite small effects yield substantial differences in the number of people who are affected versus those who are not. A similar point can be made regarding the effects of teacher expectations on students' academic performance--although the effect is only a modest one, r^2s range from .04 to .09 (see Jussim, 1990 for a recent review)--this small effect may still bear important consequences.

In summary, assessing the practical significance of one's research findings is an important endeavor. However, the current manner in which measures of "variance explained" are used can obscure effects of great practical significance; and thus, these measures should be interpreted with caution. Our best advice is this: Given that there is no table of critical effect-size values to consult to determine practical significance, researchers should ask themselves whether there are any factors at work that might render even statistically small effects practically important? In a number of research areas the answer is surely yes.

References

Abelson, R. (1985). A variance explanation paradox: When a little is a lot. Psychological Bulletin, 97, 129-133.

Brief for amicus curiae American Psychological Association in support of respondent, Price Waterhouse v. Ann B. Hopkins (1988). Washington, DC: American Psychological Association, Executive Office.

Barrett, G.V., & Morris, S.B. (in press). The American Psychological Association's Amicus Curiae Brief in Price Waterhouse v. Hopkins: The values of science versus the values of the law. Law and Human Behavior.

Borman, W.C., White, L.A., Pulakos, E.D., & Oppler, S.H. (1991). Models of supervisory job performance ratings. Journal of Applied Psychology, 76, 863-872.

Campbell, J.P. (1990). The role of theory in industrial-organizational psychology. In M.D. Dunnette and L.M. Hough (Eds.), Handbook of industrial organizational psychology (Vol.1, pp. 39-73). Palo Alto, CA.: Consulting Psychologists Press.

Cascio, W.F. (1991). Applied psychology in personnel management. NJ: Prentice-Hall.

Dipboye, R.L., Fromkin, H.L., & Wiback, K. (1975). Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. Journal of Applied Psychology, 60, 39-43.

Fiske, S.T., Bersoff, D.N., Borgida, E., Deaux, K., & Heilman, M.E. (1991). Social science research on trial: Use of sex stereotyping research in Price Waterhouse v. Hopkins. American Psychologist, 46, 1049-1060.

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

Hunter, J.E., & Schmidt, F.L. (1982). Fitting people to jobs: The impact of personnel selection on national productivity. In M.D. Dunnette & E.D. Fleishman (Eds.), Human

performance and productivity, Vol 1: Human capacity assessment (pp. 233-284). Hillsdale, NJ: Erlbaum.

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

Jussim, L. (1990). Social reality and social problems: The role of expectancies. Journal of Social Issues, 46, 9-34.

Latham, G.P. (1986). Job performance and appraisal. In C.L. Cooper and I. Robertson, (Eds.), International review of industrial and organizational psychology. New York: Wiley.

Martell, R.F. (1991). Sex bias at work: The effects of attentional and memory demands on performance ratings of men and women. Journal of Applied Social Psychology, 21, 1939-1960.

Martell, R.F. (in press). Toward eliminating sex discrimination in work settings. In K.M. Borman and P. Dubeck (Eds.), Encyclopedia of women and work. New York: Garland.

Morrison, A.M., White, R.P., & Van Velsour, E. (1987). Breaking the glass ceiling: Can women reach the top of America's largest corporations? Reading, Mass. Addison-Wesley.

O'Grady, K.E. (1982). Measures of explained variance: Cautions and limitations. Psychological Bulletin, 92, 766-777.

Olian, J.D., Schwab, D.P., & Haberfeld, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations. Organizational Behavior and Human Decision Processes, 41, 180-195.

Peters, L. H., O'Connor, E.J., Weekly, J., Pooyan, A., Frank, B., & Erenkrantz, B. (1984). Sex bias and managerial evaluations: A replication and extension. Journal of Applied Psychology, 69, 349-352.

Price Waterhouse v. Ann B. Hopkins, 109 S. Ct. 1775, (1989).

Pulakos, E.D., White, L.A., Oppler, S.H., & Borman, W.C. (1989). Examination of race and sex effects on performance ratings. Journal of Applied Psychology, 74, 770-780.

Rosenbaum, J.E. (1979). Tournament mobility: Career patterns in a corporation. Administrative Science Quarterly, 24, 220-241.

Rosenthal, R. (1990). How are we doing in soft psychology? American Psychologist, 45, 775-777.

Rosenthal, R. (1983). Assessing the statistical and social importance of the effects of psychotherapy. Journal of Consulting and Clinical Psychology, 51, 4-13.

Rosenthal, R., & Rosnow, R.L. (1991). Essentials of behavioral research: Methods and data analysis. New York: McGraw-Hill.

Sackett, P.R., Dubois, C.L.Z., & Noe, A.W. (1991). Tokenism in performance evaluation: The effects of work group representation on male-female and white-black differences in performance ratings. Journal of Applied Psychology, 76, 263-267.

Schein, E. (1978). Career dynamics: Matching individual and organizational needs. Reading, MA: Addison-Wesley.

Schmidt, F.L., Hunter, J.E., McKenzie, R.C., & Muldrow, T. (1979). The impact of valid selection procedures on work force productivity. Journal of Applied Psychology, 64, 609-626.

Sechrest, L., & Yeaton, W.H. (1982). Magnitudes of experimental effects in social science research. Evaluation Review, 6, 579-600.

U.S. Department of Labor, Women's Bureau (1989). Facts on working women, No. 89-4.

U.S. Equal Employment Opportunity Commission, et al (1978). Guidelines on employment selection procedures. Federal Register, 43, 38290-38315.

Van Maanen, J. (1977). Organizational careers: Some new perspectives. London: Wiley.

Author Notes

This research was conducted when the senior author was at Rice University.

Correspondence concerning this article should be sent to Dr. Richard F. Martell Department

of Social and Organizational Psychology Teachers College, Columbia University P. O. Box 6

New York, NY 10027

Table 1

Applicant Pool Of 600 Men and 600 Women.

Number of Men Hired	SRmen	Number of Women Hired	SRwomen	Chi-Square	Minimum Effect Size To Violate 4/5ths Rule
540	.90	428	.71	67.02	$r^2 = .056$
480	.80	380	.63	41.04	$r^2 = .034$
300	.50	236	.39	13.81	$r^2 = .012$
120	.20	92	.15	4.49	$r^2 = .004$

NOTE: SR (selection ratio)

$$\phi = \sqrt{\frac{\chi^2}{n}} = \text{Pearson's } r$$

All chi-square tests are significant at p < .05 to .0001.

Table 2

Applicant Pool Of 100 Men and 100 Women.

Number of Men Hired	SRmen	Number of Women Hired	SRwomen	Chi-Square	Minimum Effect Size To Violate 4/5ths Rule
90	.90	71	.71	11.48	$r^2 = .057$
80	.80	63	.63	7.09	$r^2 = .036$
50	.50	36	.36	3.99	$r^2 = .020$
20	.20	10	.10	3.90	$r^2 = .019$

NOTE: SR (selection ratio)

$$\phi = \sqrt{\frac{\chi^2}{n}} = \text{Pearson's r}$$

All chi-square tests are significant at $p < .05$ to .0001.

Table 3

Results of Computer Simulation 1: Effect Size 9% of The Variance

Level	Mean Score	Positions	Percentage of Women
8	77.37	10	19
7	70.20	40	24
6	65.08	75	31
5	61.80	100	36
4	58.87	150	42
3	56.22	200	46
2	51.79	350	51
1	45.90	500	60

Table 4

Results of Computer Simulation 2: Effect Size 5% of The Variance

Level	Mean Score	Positions	Percentage of Women
8	76.95	10	29
7	68.80	40	31
6	63.79	75	38
5	60.80	100	39
4	57.85	150	43
3	55.06	200	47
2	50.93	350	52
1	45.00	500	58

19

Table 5

Results of Computer Simulation 3: Effect Size 1% of The Variance

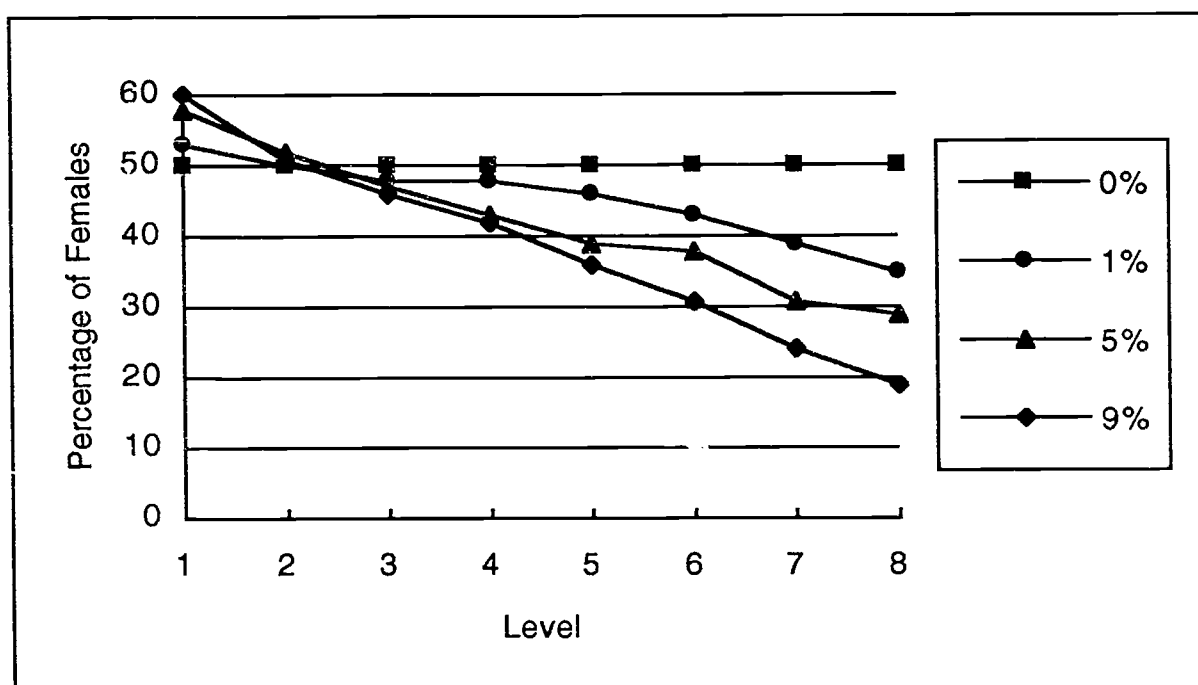| Level	Mean Score	Positions	Percentage of Women
8	74.08	10	35
7	67.14	40	39
6	62.16	75	43
5	59.15	100	46
4	56.03	150	48
3	53.64	200	48
2	49.77	350	50
1	44.02	500	53

Figure Captions.

Figure 1. Percentage of females at each job level as a function of the percentage of variance in

performance scores explained by sex.