

2012

Minority Bias in Supervisor Ratings: Comparing Subjective Ratings and Objective Measures of Job Performance

William Hernandez
Wright State University

Follow this and additional works at: https://corescholar.libraries.wright.edu/etd_all



Part of the [Industrial and Organizational Psychology Commons](#)

Repository Citation

Hernandez, William, "Minority Bias in Supervisor Ratings: Comparing Subjective Ratings and Objective Measures of Job Performance" (2012). *Browse all Theses and Dissertations*. 680.
https://corescholar.libraries.wright.edu/etd_all/680

This Thesis is brought to you for free and open access by the Theses and Dissertations at CORE Scholar. It has been accepted for inclusion in Browse all Theses and Dissertations by an authorized administrator of CORE Scholar. For more information, please contact library-corescholar@wright.edu.

**Minority Bias in Supervisor Ratings:
Comparing Subjective Ratings and Objective Measures of Job Performance**

A thesis submitted in partial fulfillment
Of the requirements for the degree of
Master of Science

By

William Hernandez
B.S., Syracuse University, 2008

2012
Wright State University

WRIGHT STATE UNIVERSITY
SCHOOL OF GRADUATE STUDIES

DECEMBER 4, 2012

I HEREBY RECOMMEND THAT THE THESIS PREPARED UNDER MY SUPERVISION BY William Hernandez ENTITLED Minority Bias in Supervisor Ratings: Comparing Subjective Ratings and Objective Measures of Job Performance BE ACCEPTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF Master of Science.

Corey Miller, Ph.D.
Thesis Director

Debra Steele-Johnson, Ph.D.
Graduate Program Director

John Flach, Ph.D.
Chair, Department of Psychology

Committee on Final Examination

Corey Miller, Ph.D.

Nathan Bowling, Ph.D.

Debra Steele-Johnson, Ph.D.

Andrew T. Hsu, Ph.D.
Dean, School of Graduate Studies

MINORITY BIAS IN SUPERVISOR RATINGS

ABSTRACT

Hernandez, William. M.S., Industrial/Organizational Psychology Program, Dept. of Psychology, Wright State University, 2012. Minority Bias in Supervisor Ratings: Comparing Subjective Ratings and Objective Measures of Job Performance

Supervisor ratings of job performance and objective sales performance were examined to better understand the causes of observed differences in performance ratings between Men and Women and Caucasians and African-Americans. Sex and race did not significantly predict subjective ratings of job performance. Ratee sex and race accounted for less than 2% of the variance in subjective ratings of ratee job performance. However, it was found that Women performed significantly lower than Men, but the difference disappeared when women comprised greater than 30% of the workforce, suggesting a real difference in performance. No tokenism effect was found for Women or racial minorities. Non-significant effects were found for the direct effects of race and sex, as well as sex congruency between rater and ratee.

MINORITY BIAS IN SUPERVISOR RATINGS

Table of Contents

Minority Bias in Supervisor Ratings: Comparing Subjective Ratings and Objective Measures of Job Performance	
Performance Appraisals 3	
Subjective and Objective Measures of Job Performance 5	
Performance Ratings 6	
Cognitive Processes Within Raters..... 8	
Race Effects on Performance Ratings 10	
Sex Effects on Performance Ratings 13	
Method 17	
Participants 17	
Objective Job Performance 17	
Subjective Job Performance 17	
Predictor Variables 18	
Control Variables 18	
Data Analyses 18	
Results..... 20	
Race Effects on Performance Ratings 21	

MINORITY BIAS IN SUPERVISOR RATINGS

Sex Effects on Performance Ratings.....	22
Tokenism Effects on Performance Ratings	22
Rater-Ratee Congruency Effects on Performance Ratings	25
Discussion	25
Potential Causes of Observed Performance Decrements	27
Theoretical Implications.....	28
Practical Implications.	30
Limitations.....	30
Future Directions	31
Conclusion.....	32
References.....	33

MINORITY BIAS IN SUPERVISOR RATINGS

LIST OF FIGURES

Figure

1. Predicted Performance Ratings for Women in Workgroups Comprised of Less than
20% Women Compared to Workgroups Comprised of Greater than 20%. Women 50

MINORITY BIAS IN SUPERVISOR RATINGS

LIST OF TABLES

Table

1. Summary of Intercorrelations, Means, & Standard Deviations of Study Variables	38
2. Multiple Regression Analyses of Effects of Ratee Race (Caucasian-African-American) on Subjective Supervisor Ratings of Ratee Job Performance.....	39
3. Multiple Regression Analyses of Effects of Ratee Race (Caucasian-Hispanic) on Subjective Supervisor Ratings of Ratee Job Performance.....	40
4. Multiple Regression Analyses of Effects of Ratee Race (Caucasian-Asian) on Subjective Supervisor Ratings of Ratee Job Performance.....	41
5. Multiple Regression Analyses of Effects of Ratee Race (African-American-Hispanic) on Subjective Supervisor Ratings of Ratee Job Performance.....	42
6. Multiple Regression Analyses of Effects of Ratee Race (Any Racial Minority) on Subjective Supervisor Ratings of Ratee Job Performance.....	43
7. Multiple Regression Analyses Testing the Effects of Ratee Sex on Subjective Supervisor Ratings of Ratee Job Performance.....	44
8. Multiple Regression Analysis Testing the Effects of Minority Status (Race or Sex Minority) on Subjective Supervisor Ratings of Ratee Job Performance	45
9. Multiple Regression Analysis Testing the Effects of Race Tokenism-AA on Subjective Supervisor Ratings of Ratee Job Performance.....	46

MINORITY BIAS IN SUPERVISOR RATINGS

10. Multiple Regression Analysis Testing the Effects of Race Tokenism-Hispanic on Subjective Supervisor Ratings of Ratee Job Performance..... 47

11. Multiple Regression Analysis Testing the Effects of Sex Tokenism on Subjective Supervisor Ratings of Ratee Job Performance..... 48

12. Multiple Regression Analysis Testing the Effects of Sex Congruency on Subjective Supervisor Ratings of Ratee Job Performance..... 49

MINORITY BIAS IN SUPERVISOR RATINGS: COMPARING SUBJECTIVE RATINGS AND OBJECTIVE MEASURES OF JOB PERFORMANCE

Supervisor ratings of job performance are the primary criteria for a number of personnel selection decisions (Arvey & Murphy, 1998; Lent, Aurbach, & Levin, 1971). Organizations typically use supervisor ratings as the criteria for decisions regarding employee terminations and promotions as well as identifying those employees in need of further training and development. The importance of supervisor ratings and their accuracy is readily apparent when considering the number of decisions organizations make based on the ratings. Given that supervisor ratings are subjective in nature, they may be particularly susceptible to rater bias. Critical to organizations is the ability to identify rater biases in performance evaluation procedures and correct any inaccuracies that exist. This is critical to organizations as such bias may result in legal ramifications or adverse impact. Discrimination regarding race and sex is of particular concern. Previous research has shown that members of a particular race tend to rate members of their own race higher than members of other races (Kraiger & Ford, 1985). However, dependence on group mean differences to detect significant effects is severely limiting. The most common explanations for group differences in performance ratings have been either that true differences in performance exist between groups or rater bias is evident towards particular groups. Use of a single measure of performance does not allow an explanatory distinction between these two rationalizations. Additionally, use of the same

MINORITY BIAS IN SUPERVISOR RATINGS

types of measures of performance may be unable to determine cause as well. For example, multiple subjective measures of performance (e.g. 360-degree feedback) are susceptible to the same concerns. As such, it is critical to examine multiple sources of performance using different methods of measurement.

Comparing subjective performance ratings with some metric of objective performance allows us to better assess the issue of potential bias in performance evaluation procedures, particularly supervisor ratings. Some jobs may lend themselves to such an assessment better than others, as objective measures are more readily available for certain jobs. For example, administrative positions where typing is essential to the job may readily allow for measurement of objective performance such as words typed per minute, calls answered per day, etc. Metrics such as calls taken per day, average handle time per call, or number of insufficiently answered calls (calls redirected to a supervisor) provide objective measures of job performance, assessed with relative ease.

Sales occupations provide an opportunity to measure objective financial performance as well as other potential metrics, such as sales efficiency (sales per day, week, month, etc.), sales quality (e.g. satisfaction of customers via maintained or severed customer-provider relationships), and magnitude of sales (e.g. sales per customer compared to total sales). Additionally, the objective measures of job performance for these sales positions directly link to the goals of the organization, profitability. The objective measures of performance for the other jobs previously mentioned would constitute measures of secondary organizational goals. For example, secondary organizational goals would include performance efficiency (e.g., words typed per minute,

MINORITY BIAS IN SUPERVISOR RATINGS

average handle time per call) which would aim to support the primary goal of organizational profit.

Comparing subjective supervisor ratings of performance with objective sales performance would allow for a more accurate determination of whether bias exists in the rating of members of different groups. For example, if a particular group performs similarly or better than another group on objective measures of job performance while simultaneously being rated lower by their supervisor, one can infer bias exists. Such a study allows for the ability to more accurately determine if group differences in supervisor ratings are a result of actual differences in job performance or merely bias.

Performance Appraisal

Performance appraisals are an effective way to increase organizational productivity through the identification of development opportunities for employees (Tangen, 2003). The two most common organizational uses of performance appraisal systems are for administrative purposes (i.e., employee performance feedback or identifying employees in need of training) or personnel decisions (i.e., criteria for selecting employees for promotions, terminations, or bonus allocation). Unfortunately, it is difficult to implement a performance appraisal system, which serves both administrative and developmental purposes. Raters may differ with regard to their motivation and desired purpose of the performance appraisal (Tziner & Kopelman, 2002). For example, a lenient supervisor may rate subordinates more highly when the appraisal system serves as criteria for personnel decisions, as the supervisor might not want to be responsible for an employees' termination. Similarly, that same supervisor may rate subordinates more accurately if the appraisal system serves a developmental

MINORITY BIAS IN SUPERVISOR RATINGS

purpose, as the supervisor may wish to identify areas for potential improvement (Den Hartog, Boselie, & Paauwe, 2004). As different goals (administrative or developmental) for appraisal systems might lead to different ratings, resulting in low agreement, raters must address the issue of showing agreement across both instances of ratings if the organization uses a performance appraisal system for both administrative and developmental goals. If a single appraisal system is used for both purposes, a supervisor might rate with his/her own goals in mind, as he/she might view employee development as more/less important than determining the allocation of employee bonuses (performance appraisal for personnel decisions).

Given the subjective nature of performance ratings, performance appraisal systems are often the focus of discrimination concerns involving terminations and promotions (Tangen, 2003). Raters are in fact human and, thus, subject to the same errors and bias as any other individual. Potential sources of bias or error in performance appraisal can result from halo, leniency, or stereotyping (Kraiger, Ford, & Schechtman, 1986). Additionally, rater goals can affect rating accuracy. For example, some supervisors may wish to maintain harmony in a work group, and thus rate all employees positively. Another supervisor may wish to motivate a work group by inflating the ratings of lower performers to instill confidence in those employees (Wong & Kwong, 2007).

When considering the potential biases that may enter into the performance evaluation of subordinates, it is important to consider the many possible sources from which they may originate. A source of bias is any influential force affecting the accurate measurement of job performance. As such, bias may not only result from the rater alone,

MINORITY BIAS IN SUPERVISOR RATINGS

but may also originate from seemingly distal sources such as the customers. Due to the direct influence of the customer-employee interaction on performance output, any potential bias, unrelated to job performance (i.e., race, sex, national origin), on the part of the customers towards an employee can result in artificially deflated ratings of performance. Of course, larger effects on the ratings of employees are logically more likely in occupations with direct contact with customers (i.e., sales, clerical, service), as compared to occupations with less contact (i.e., factory line workers, some technical jobs). A suitable explanation for such a bias is the occurrence of the actor-observer bias, in which the subordinate (the actor) attributes performance to situational factors (i.e., sales region, difficult customers) whereas the observer (the supervisor) attributes performance to some perceived stable traits within the subordinate (i.e., laziness, lack of ability) (Martinko & Gardner, 1987).

Subjective and Objective Measures of Job Performance

Job performance is a common criteria used to validate selection measures, as these measures intend to predict future job performance of potential job candidates. However, there is a variety of ways to measure job performance, such as via ratings of job performance, be it from supervisors, peers, subordinates, or self; or measures of output, such as sales performance, organization profitability, or time-based productivity, to name a few (Tangen, 2003). Thus, objective and subjective measures of job performance constitute the two categories of job performance measures.

Use of one type of measure over the other may result in different results with regard to personnel decisions. Previous research has found relatively low mean correlations between subjective and objective measures of job performance (Bommer et

MINORITY BIAS IN SUPERVISOR RATINGS

al., 1995; Heneman, 1986; Rich et al., 1999). Bommer et al. (1995) conducted a meta-analysis of studies containing both objective and subjective measures of job performance to test their interchangeability. Bommer et al. (1995) found a corrected mean correlation of just .389 between objective and subjective measures of job performance. Controlling for potential moderators such as job type, rating method (absolute vs. relative ratings), and rating format (overall or composite scores), they found no support for convergent validity, with subjective and objective measures of job performance sharing less than 25% of the explained variance in every situation. Similarly, Heneman (1986) found an even lower mean correlation ($r = .27$) between the subjective and objective performance measures. The aforementioned studies suggest that objective and subjective measures of job performance are not interchangeable.

Potential explanations for the lack of convergent validity of objective and subjective measures of job performance include but are not limited to error, rater bias, and measurement of differing components of overall job performance. Whereas objective measures of job performance may be free of rater bias, situational factors (i.e., a sales representative sells more product in an area with high demand for said product compared to a similarly performing sales representative in another area) remain a large concern. Additionally, if we examine job performance as a multi-dimensional construct (Chockalingam, Schmidt, & Ones, 2005; Viswesvaran & Ones, 2000), objective measures of performance may capture only a portion of the variance in total job performance. In comparison, subjective measures such as supervisor ratings may capture multiple dimensions of job performance, assuming bias is not present.

Performance Ratings

MINORITY BIAS IN SUPERVISOR RATINGS

Despite the potential entry of various sources of error in the rating process, the potential of performance ratings to capture aspects of job performance not explained by objective measures of job performance highlights the importance of their continued use and need for greater accuracy. Unfortunately, many supervisors must interpret rating scales on their own, with little direction in terms of how to complete subordinate ratings. Additionally, making supervisors aware of the organizational goals of appraisal system might still leave supervisors unaware of the potential errors likely made during the rating process. As such, there is general agreement on the recommended use of training to improve rater accuracy (Hedge & Kavanagh, 1988; Lee, 1985; Pulakos, 1984; Woehr & Huffcutt, 1994).

Woehr and Huffcutt (1994) evaluated four rating training strategies in terms of halo error, leniency, rating accuracy, and observational accuracy. The first training strategy, rater error training, seeks to make raters aware of issues with negatively skewed ratings, range restriction, and rater biases such as halo error. Results of the meta-analysis suggest that rater error training reduced rater errors. A second type of training strategy is performance dimension training, in which greater emphasis on the specific dimensions of performance are given. This strategy seeks to direct raters away from the use of a global performance score and reduce issues with halo, while increasing rating accuracy. A third strategy, frame-of-reference training, aims to train raters on both the dimensions of job performance evaluated and the standards for providing accurate ratings. Such training attempts to increase rating accuracy by training all raters to rate according to a similar standard of evaluation. The final training strategy, behavioral observation training, attempts to train raters to rate observations of subordinate behavior as opposed to

MINORITY BIAS IN SUPERVISOR RATINGS

subjective evaluations of subordinate behaviors. Positive effects on the outcome variables (rater accuracy, rater agreement, etc.) were evident with all four strategies.

In addition to rater training, appropriately constructed measures can increase rating accuracy. There are several performance rating format options to consider within a performance appraisal system, such as graphic rating scales (GRS), behavioral observation scales (BOS), and behaviorally anchored response scales (BARS). Each of these formats utilizes a numbered scale with anchors of some sort. GRS use abstract anchors such as 'agree', 'strongly disagree', etc. BARS are those scales that provide examples of behaviors that would constitute a given rating. Finally, BOS require raters to recall instances of the behaviors in question and make responses referencing the frequency of those behaviors. Not only do BOS's have greater reliability and validity than BARS or GRS (Latham & Wexley, 1981) but also result in improved attitudinal effects from raters and ratees with respect to their perceived fairness and satisfaction (Tziner & Kopelman, 2002).

Cognitive Processes within Raters

Even with adequate rater training, there are a number of sources of potential bias to consider. Perhaps the most common source of bias affecting supervisor ratings of subordinate job performance is within the cognitive processes of the supervisor. A general sequence of cognitive processes would include the presentation of information, attention, encoding, storage, retrieval, integration, and the act of providing ratings (Wexley & Klimoski, 1984). Loss or misinterpretation of the information presented at the beginning of such a sequence can occur at any of these processes. Feldman (1981) has suggested that this sequence is comprised of automatic and controlled processes.

MINORITY BIAS IN SUPERVISOR RATINGS

Feldman posited that the determination of which process is enacted, automatic or controlled, depends on the cognitive demand placed upon the rater. When cognitive demand is low and no problems have arisen, automatic processes are more likely. These automatic processes can include categorization or prototyping for example. On the other hand, when cognitive demands are high, such as when readily available prototypes or schemas are unavailable, controlled processes are more likely. An example of a controlled process includes the use of attributions in decision-making.

The need for such automatic and controlled processes in performance evaluation is due to the simple nature of human memory. As raters are required to provide ratings of subordinate performance, typically spanning twelve months, it is difficult to recall every instance of performance, and particularly difficult when the number of subordinates a rater is required to provide ratings for is large. Several studies have supported the notion that rater accuracy is dependent on the method raters utilized to encode information (DeNisi & Peters, 1996; Kinicki et al., 1995; Sanchez & De La Torre, 1996). DeNisi and Peters (1996) examined the effects of structured diary keeping and structured recall on rating accuracy. They found that those raters that used these methods of recall when rating subordinate performance provided less elevated ratings than those raters using traditional recall did.

As most raters do not use alternative recall methods such as structured diaries, they must rely on their own cognitive processes. For example, raters can defer to preconceived prototypes when providing ratings of subordinates. Additionally, a typical rater has only limited information about the performance of a subordinate, from indirect observations, and the rater may bias the ratings of one subordinate due to the rater's

MINORITY BIAS IN SUPERVISOR RATINGS

preconceived notion of that subordinate (Feldman, 1981). Similarly, stereotypes of race and sex may bias ratings of performance as a rater can have a prototype of a subordinate representing a given race or sex. Accompanying stereotypes are attributions of the respective prototype placed upon the subordinate. Supervisors are likely to form certain attributions about their subordinates, many of these attributions traced to affect, or how much a supervisor likes or dislikes a subordinate (Ferris et al., 1994). Isen and Baron (1991) have suggested that positive affect towards subordinates will result in greater recall of positive information. Similarly, negative affect facilitates the recall of negative information, such as instances of poor performance. Additionally, other research (Sinclair, 1988) has suggested that positive affect results in increased halo error.

Race Effects on Performance Ratings

Despite the various methods of training raters and use of ideal rating formats for respective positions, biases still enter into the subjective process of rating performance. Even with performance measures of the highest fidelity, discrepancy between actual performance and ratings of performance are often evident. Whereas a number of potential variables directly relate to performance unaccounted for by performance measures, there are a number of potential contaminating variables. Possible contamination includes rater bias; particularly, bias based on some factor unrelated to actual job performance. One such bias that has large legal and social implications is rater bias due to ratee race.

Landy and Farr (1980) conducted a review of race of ratee effects on performance ratings, highlighting the mixed results collected by various studies up until that point. There are a number of studies supporting the notion of a race of ratee effect on supervisor

MINORITY BIAS IN SUPERVISOR RATINGS

ratings but only in specific circumstances. Hamner et al. (1974) found that ratees received higher ratings from same race raters, whereas Schmidt and Johnson (1973) found no such results. The only conclusive result Landy and Farr (1980) forwarded is the notion that ratees receive higher ratings from same race raters, but Landy and Farr denoted that this is not necessarily true in highly integrated situations.

Attempting to reach more substantive conclusions about the effects of ratee race on performance ratings, Kraiger and Ford (1985) conducted a meta-analysis, which included 74 studies. They found that the correlation between ratee race and performance ratings were .183 for Caucasian raters and -.220 for African-American raters respectively; however, neither correlation was statistically significant. A more substantial finding in this study was that African-Americans received significantly lower performance ratings than their Caucasian counterparts when the workgroup was comprised predominately of Caucasians. A separate meta-analysis by Kraiger and Ford (1990) examined the correlations between job performance, job knowledge, and performance ratings. Results of their meta-analysis suggest that raters may base ratings of African-Americans ($r = .217$) more closely on job performance than for Caucasians ($r = .109$). They proposed that these differences are attributable to leniency towards same race ratees.

Recent studies have continued to produce conflicting results. Waldman and Avolio (1991) conducted a study of 21,547 individuals across 10 job categories, examining the effects of ratee race on performance ratings. They concluded that there was a significant effect of ratee race on performance ratings, but this effect disappeared when controlling for cognitive ability, education, and job experience. Such a finding

MINORITY BIAS IN SUPERVISOR RATINGS

suggests the possibility for a real ability difference between African-Americans and Caucasians. Additionally, they found no support for Kraiger and Ford's (1985) conclusion of a rater-ratee interaction. Contrary to the findings of Waldman and Avolio (1991), Sackett, DuBois, and Noe (1991) conducted a study examining tokenism in performance ratings and found support for the notion that African-Americans receive lower performance ratings than Caucasians when they constitute a small portion of the workforce. Interestingly, the only positive difference found between ratings of African-Americans and Caucasians occurred when African-Americans comprised 91-99% of the workforce.

Results from these meta-analyses and studies indicate that further investigation is necessary to understand the effects of ratee race on performance ratings. Each of the previously cited studies admits to several limitations. Meta-analyses typically have not included adequate information concerning the types of rating forms used during performance evaluation or may analyze objective and subjective measures of performance from different jobs. Many of the individual studies are similarly limited in that readily available objective and subjective measures of performance are rarely available for analysis. Additionally, previous studies have been unable to provide conclusive evidence that any differences found between ratings of African-Americans and Caucasians is a result rater bias or true differences in performance. There is the potential concern that any perceived differences in true performance differences between Caucasians and African-Americans could be a result of culturally biased evaluation measures as well.

MINORITY BIAS IN SUPERVISOR RATINGS

Despite the conflicting results across studies, there is reasonable support to expect rater bias in situations where African-Americans comprise a relatively small portions of the workforce if rater bias does exist. Given that previous literature (Sackett, DuBois, & Noe, 1991) suggested that tokenism does not have a reverse effect on Caucasians, where they receive lower ratings when comprising a relatively small portion of the workforce, it is reasonable to expect such a relationship will not exist in the current study as well. To a lesser extent, I may still find support for a race of ratee effect when Caucasians must rate African-American subordinates.

Sex Effects on Performance Ratings

Similar to issues of race-based rater bias, concerns over potential sex discrimination are an important topic for organizations to address. However, much of the previous research (Pulakos et al., 1989; Landy & Farr, 1980; Schmidt & Lippin, 1980) has provided even more inconsistent results for sex effects than race effects on performance ratings. Schmidt and Lippin (1980) concluded that although sex effects contributed a significant amount of variance in performance ratings, these effects were trivially small. On the other hand, Landy and Farr (1980) found support for sex effects on performance ratings when considering type of job. Landy and Farr concluded that ratee sex has a significant impact on performance ratings dependent on stereotype of the job (predominately masculine or feminine jobs). For example, women in jobs stereotypically dominated by men received less favorable ratings. Similarly, men in predominately-feminine jobs also received less favorable performance ratings. Landy and Farr (1980) conducted a review of sex of ratee effects on performance ratings, highlighting the mixed results collected by various studies up until that point. However,

MINORITY BIAS IN SUPERVISOR RATINGS

the large proportion of laboratory studies used in experiments looking at ratee sex effects on performance evaluation, severely limits their review. As such, previous research through 1980 has not supported the notion of a ratee sex effect on performance ratings or a ratee-rater sex interaction.

Sackett, DuBois, and Noe (1991) found evidence of a sex effect of ratee on performance ratings; stronger evidence when compared to any race effects. They found that women received significantly lower ratings than men did, when women comprised a significantly smaller portion of the workforce. Specifically, they found negative effect sizes once the proportion of women dropped below 50% of the total workforce. They found significant and large negative effect sizes when the proportion women in the workforce dropped below 20%. Support for rater bias as an explanation for this difference is evident in the lack of a tokenism effect for men. However, Sackett, DuBois, and Noe (1991) identified their inability to separate rater bias and true differences in performance ability as an explanation for the apparent differences in ratings of performance. In fact, such a differentiation remains a hypothetical explanation, with previous studies providing incomplete evidence for the idea that rater bias contributes to some degree in explaining the differences in male and female performance ratings.

Given the history of conflicting results concerning effects of race and sex of ratee on ratings of job performance, I propose that examining the effects of race and sex of ratee on ratings of performance using both subjective and objective measures of performance from the same sample will provide a more definitive conclusion. Whereas a number of studies have supported a race and sex effect, even those that provided non-significant results found results in their hypothesized direction (e.g., majority group,

MINORITY BIAS IN SUPERVISOR RATINGS

Caucasians or men, rated higher than the minority group, African-Americans or women).

With the use of objective and subjective data from the same source and the presentation of objective data to the raters prior to their completion of subjective measures, I hypothesize that race of ratee will account for significant variance in supervisor ratings, controlling for objective performance measures and demographic variables.

Additionally, I hypothesize that sex of ratee will account for significant variance in supervisor ratings, controlling for objective performance measures and demographic variables.

Hypothesis 1A: Race of ratee will predict unique variance in supervisor ratings over and above objective sales performance, controlling for demographic variables.

Hypothesis 1B: Sex of ratee will predict unique variance in supervisor ratings over and above objective sales performance, controlling for demographic variables.

Rosabeth Moss Kanter (1977) published a case study of 20 upper management saleswomen in a 300-person sales force in a fortune 500 company. Her results indicated that these twenty women experienced negative consequences due to their “token” status, their membership in a subgroup (being female) comprising less than 15% of the total workgroup. Recent studies (Sackett, DuBois, & Noe, 1991; Yoder, 1994) that found a negative tokenism effect for women support Kanter’s results. Whereas Sackett, DuBois, and Noe (1991) found a strong tokenism effect for sex, they also found consistently lower ratings for African-Americans regardless of percentage within total group. As such, it is reasonable to assume that in some situations (i.e., various geographic locations) this effect is more pronounced. I hypothesize that the percentage of women in the measured

MINORITY BIAS IN SUPERVISOR RATINGS

sample will predict supervisor ratings of women when controlling for objective measures of job performance and demographic variables. Similarly, I hypothesize that the percentage of African-Americans in the measured sample will predict supervisor ratings when controlling for objective measures of job performance and demographic variables.

Hypothesis 2A: Percentage of African-Americans in the sample will predict unique variance in supervisor ratings of African-Americans over and above objective sales performance, controlling for demographic variables.

Hypothesis 2B: Percentage of women in the sample will predict unique variance in supervisor ratings of women over and above objective sales performance, controlling for demographic variables.

Previous research has provided conflicting results concerning the effects of the race of raters on ratings of performance, with some studies (Hamner et al., 1974; Landy & Farr, 1980) supporting this effect and others (Schmidt & Johnson, 1973) finding no such evidence. With the use of measures of objective and subjective job performance from the same employees, the current study should shed some additional light into determining any potential race of rater effect on ratings of subordinate job performance. I posit that race congruency (same race rater and ratee) will predict supervisor ratings for African-Americans when controlling for objective measures of job performance and demographic variables. Similarly, I posit that sex congruence (same sex rater and ratee) will predict supervisor ratings for women when controlling for objective measures of job performance and demographic variables.

MINORITY BIAS IN SUPERVISOR RATINGS

Research Question 1A: Race congruency will predict unique variance in supervisor ratings of African-Americans over and above objective sales performance, controlling for demographic variables.

Research Question 1B: Sex congruency will predict unique variance in supervisor ratings of women over and above objective sales performance, controlling for demographic variables.

Method

Participants

An archival database of performance data from a mid-western U.S. consulting firm provided all data used in this study. Organizations reporting both objective and subjective measures of job performance provided participant data. There was an initial sample of 4061 employees across a wide range of sales jobs and organizations. The final sample included 1894 employees after deletion of those employees that did not have either objective or subjective performance data available. This sample included 63.4% men and 36.6% women with a demographic racial distribution as follows: 78.0% Caucasian, 9.9% African American, 5.3% Hispanic, 2.0% Asian, and 4.9% other. Average age of employees sampled was 34.8 years old. The managers providing ratings for these employees were comprised of 84.9% males.

Objective Job Performance

Objective job performance was defined as average sales output over a twelve-month period, and was standardized with each organization.

Subjective Job Performance

MINORITY BIAS IN SUPERVISOR RATINGS

Subjective job performance was defined as global ratings of job performance via results of annual supervisor performance ratings. Global ratings are subjective indications of overall job performance made by employees' immediate supervisors and were standardized within each organization.

Predictor Variables

Sex. Employee sex was dichotomized as 0 (Female) and 1 (Male).

Race. Four dichotomous variables were created for separate racial comparisons. This included Caucasian (0) & African-American (1), Caucasian (0) & Hispanic (1), Caucasian (0) & Asian (1), and African-American (0) & Hispanic (1). Additionally, a variable indicating status as any racial minority was created. A separate variable was created to indicate employee status as a racial minority and/or female; being a minority of any kind.

Tokenism. Tokenism was defined as the percentage of a given minority group (sex- or race-based) in a respective workgroup.

Rater-Ratee Congruence. Rater-ratee congruence was defined as the similarity of rater and ratee on a given variable. Rater-ratee congruency exists when both rater and ratee are of the same sex or race. All other rater-ratee combinations (e.g., Caucasian rater-Asian ratee, African-American rater-Hispanic ratee, male rater-female ratee) would constitute rater-ratee incongruence.

Control Variables

Job tenure was measured as the number of months employed within an organization. Additionally, employee age was contained in employee records.

Data Analysis

MINORITY BIAS IN SUPERVISOR RATINGS

Descriptive statistics describing the sample and groups examined were calculated, and presented with variable intercorrelations. Hierarchical regression analyses were conducted to determine if race, gender, tokenism, and/or rater-ratee congruence add any incremental validity in predicting subjective ratings of job performance while controlling for the effects of objective sales performance, age, and job tenure.

Race Effect. In the first analysis, the outcome variable was subjective performance ratings, with sex, job tenure, objective sales performance, and age as predictors in the first step of the regression. In the second step, race of ratee (Caucasian vs. A.A.; Caucasian vs. Hispanic; Caucasian vs. Asian) was entered as a predictor to test Hypothesis 1A.

Sex Effect. In the second analysis, the outcome variable was subjective performance ratings, with race, job tenure, objective sales performance, and age as the predictors. In the second step, sex of ratee was entered as a predictor to test Hypothesis 1B.

Race Tokenism. In the third analysis, the outcome variable was subjective performance ratings, with sex, job tenure, objective sales performance, and age as the predictors. In the second step, race tokenism was entered as a predictor to test Hypothesis 2A.

Sex Tokenism. In the fourth analysis, the outcome variable was subjective performance ratings, with race, job tenure, objective sales performance, and age as the predictors. In the second step, sex tokenism was entered as a predictor to test Hypothesis 2B.

MINORITY BIAS IN SUPERVISOR RATINGS

Race Rater-Ratee Congruence. In the fifth analysis, the outcome variable was subjective performance ratings, with sex, job tenure, objective sales performance, and age as the predictors. In the second step, rater-ratee race congruency was entered as a predictor to test Research Question 1A.

Sex Rater-Ratee Congruence. In the sixth analysis, the outcome variable was subjective performance ratings, with race, job tenure, objective sales performance, and age as the predictors. In the second step, rater-ratee sex congruency was entered as a predictor to test Research Question 1B.

Results

Analyses focused on employees' objective and subjective measures of job performance as well as employees' membership in respective minority groups. Only those employees with both objective and subjective measures of job performance were included in final analyses. Given that several organizations were sampled and not all the variables of interest were available from every organization, only those employees' with relevant data for a given hypothesis or research question were included in a given analysis. Additionally, as different organizations used various performance rating scales and sales measurements, both variables were standardized within each organization before inclusion for analysis.

As there was insufficient data on the race of managers in the data set, Research Question 1 was untestable. The remaining hypotheses and Research Question 2 were examined using two-step linear regression analyses, to determine if any incremental validity was found with the addition of our variables of interest (ratee sex, ratee race, rater-ratee congruency, and tokenism), while controlling for other variables such as ratee

MINORITY BIAS IN SUPERVISOR RATINGS

age and job tenure. Means, standard deviations, and inter-correlations for all examined variables are shown in Table 1.

Prior to running analyses, a one-way ANOVA was completed to test for differences in ratings of performance between raters. Performance ratings differed significantly between raters, $F(196, 649) = 2.25, p < .001, \eta^2 = .39$.

Race Effects on Ratings of Performance

Results of a multiple regression analysis (displayed in Table 2) indicated that being African-American compared to Caucasian did not predict significant variance in supervisor ratings of job performance while controlling for ratee age, job tenure, ratee sex, and objective sales performance, $R^2 = .062, \Delta R^2 = .000, F(1, 956) = 12.60, p = .686, 90\% \text{ CI } [-.146, .241]$. Results of a multiple regression analysis (displayed in Table 3) indicated that being Hispanic compared to Caucasian did not predict significant variance in supervisor ratings of job performance while controlling for ratee age, job tenure, ratee sex, and objective sales performance, $R^2 = .062, \Delta R^2 = .000, F(1, 934) = 12.24, p = .913, 90\% \text{ CI } [-.211, .241]$. Results of a multiple regression analysis (displayed in Table 4) indicated that being Asian compared to Caucasian did not predict significant variance in supervisor ratings of job performance while controlling for ratee age, job tenure, ratee sex, and objective sales performance, $R^2 = .064, \Delta R^2 = .001, F(1, 900) = 12.29, p = .308, 90\% \text{ CI } [-.130, .554]$. Results of a multiple regression analysis (displayed in Table 5) indicated that being African-American compared to Hispanic did not predict significant variance in supervisor ratings of job performance while controlling for ratee age, job tenure, ratee sex, and objective sales performance, $R^2 = .033, \Delta R^2 = .000, F(1, 131) = 0.90, p = .887, 90\% \text{ CI } [-.315, .265]$ (Hypothesis 1A was not supported).

MINORITY BIAS IN SUPERVISOR RATINGS

Additional analyses were conducted to determine if racial minority status predicted significant variance in ratings of job performance. Results of a multiple regression analysis (displayed in Table 6) indicated that being a racial minority (non-Caucasian) did not predict significant variance in supervisor ratings of job performance while controlling for ratee age, job tenure, ratee sex, and objective sales performance, $R^2 = .052$, $\Delta R^2 = .001$, $F(1, 1072) = 11.72$, $p = .285$, 90% CI [-.046, .217].

Sex Effects on Ratings of Performance

Results of a separate multiple regression analysis (displayed in Table 7) indicated that Ratee Sex did not predict additional significant variance in supervisor ratings of job performance while controlling for ratee age, job tenure, ratee race, and objective sales performance, $R^2 = .051$, $\Delta R^2 = .000$, $F(1, 1073) = 14.36$, $p = .937$, 90% CI [-.102, .113] (Hypothesis 1B was not supported). A regression analysis was conducted to determine if inclusion in any minority group (sex- or race-based) predicted significant variance in ratings of performance. Results of a multiple regression analysis (displayed in Table 8) indicated that being a minority (non-Caucasian and/or female) did not predict significant variance in supervisor ratings of job performance while controlling for ratee age, job tenure, ratee sex, and objective sales performance, $R^2 = .051$, $\Delta R^2 = .000$, $F(1, 1073) = 14.41$, $p = .664$, 90% CI [-.076, .131].

Tokenism Effects on Ratings of Performance

Next, to test Hypothesis 2A and Hypothesis 2B, Race Tokenism and Sex Tokenism were entered into the second step of the analyses as predictors to determine if additional significant variance in subjective performance ratings was explained by these variables. Results of a multiple regression analysis (Displayed in Table 9) indicated that

MINORITY BIAS IN SUPERVISOR RATINGS

there was no tokenism effect for African-Americans, as the addition of the Race Tokenism variable did not contribute additional significant variance in explaining subjective performance ratings, $R^2 = .047$, $\Delta R^2 = .007$, $F(1, 74) = 0.74$, $p = .471$, 90% CI [-.064, .025]. (Hypothesis 2A was not supported). Follow-up analyses were conducted to examine group differences in objective and subjective job performance between Caucasians and African-Americans. Results of an independent samples t-test showed that Caucasians ($M = 0.01$, $SD = 1.02$) received significantly higher supervisor ratings of job performance than African-Americans ($M = -0.17$, $SD = 1.02$), $t(1386) = -2.10$, $p = .04$ (two-tailed). Also, Caucasians ($M = 0.07$, $SD = 1.03$) performed significantly better than African-Americans ($M = -0.12$, $SD = 0.90$) based on objective sales performance, $t(1381) = -2.13$, $p = .03$ (two-tailed). Additionally, results of a multiple regression analysis (Displayed in Table 10) indicated that there was no tokenism effect for Hispanics, as the addition of the Race Tokenism variable did not contribute additional significant variance in explaining subjective performance ratings, $R^2 = .231$, $\Delta R^2 = .049$, $F(1, 51) = 3.06$, $p = .079$, 90% CI [.007, .199]. Results of an independent samples t-test indicated that there was no significant difference in subjective ratings of performance between Caucasians ($M = 0.01$, $SD = 1.02$) and Hispanics ($M = -0.07$, $SD = 0.97$) in the overall sample, $t(1313) = -0.71$, $p = .48$, (two-tailed). Similarly, results of an independent samples t-test indicated that there was no significant difference in objective sales performance between Caucasians ($M = 0.07$, $SD = 1.03$) and Hispanics ($M = -0.03$, $SD = 0.89$) in the overall sample, $t(1309) = -0.82$, $p = .42$ (two-tailed).

When entering Sex Tokenism into the second step of the regression analysis (displayed in Table 11), results indicated that sex tokenism did predict a significant

MINORITY BIAS IN SUPERVISOR RATINGS

amount of variance in subjective ratings of job performance when employing the 20% threshold indicated by previous studies (Sackett, DuBois, & Noe, 1991), $R^2 = .031$, $\Delta R^2 = .025$, $F(1, 369) = 2.95$, $p = .002$, 90% CI [-.818, -.179]. Results of an independent samples t-test showed that when less than 20% of the workgroup was comprised of women, men ($M = 0.03$, $SD = 1.03$) received significantly higher supervisor ratings of job performance than women ($M = -0.22$, $SD = 0.94$), $t(583) = -2.15$, $p = .02$ (one-tailed). Additionally, women ($M = -0.30$, $SD = 0.46$) performed significantly worse than men ($M = 0.16$, $SD = 1.13$) based on objective sales performance, $t(583) = -6.57$, $p < .001$ (two-tailed). However, when not employing the 20% threshold and using Sex Tokenism as a continuous variable, no significant variance in subjective ratings of job performance were accounted for by Tokenism, $R^2 = .013$, $\Delta R^2 = .006$, $F(1, 369) = 1.17$, $p = .126$, 90% CI [-.003, .027] (Hypothesis 2B was not supported).

For workgroups containing greater than 20% of women, there was no significant difference between the performance ratings of men ($M = -0.02$, $SD = 1.00$) and women ($M = 0.01$, $SD = 1.04$), $t(1164) = 0.48$, $p = .63$ (two-tailed). Similarly, there was no significant difference between the objective sales performance of men ($M = -0.02$, $SD = 0.97$) and women ($M = 0.06$, $SD = 1.01$), $t(1164) = 1.28$, $p = .20$ (two-tailed). Although no significant differences were found in the analyses of workgroups containing greater than 20% of women, the relationships were in the expected direction, with women exhibiting more success in both the objective and subjective metrics when they are members of a more evenly distributed workgroup.

To better understand at which proportion women began to perform significantly worse than men, additional analyses were conducted to test group differences at different

MINORITY BIAS IN SUPERVISOR RATINGS

levels (proportions of women in respective workgroups). Results of independent samples t-tests indicated that women ($M = -0.30$, $SD = 0.46$) began to perform significantly worse than men ($M = 0.16$, $SD = 1.13$), based on objective sales performance, at approximately 30% (when women comprised less than 30% of the workgroup), $t(583) = -3.85$, $p < .001$. Below 30%, women ($M = -0.22$, $SD = 0.94$) were being rated accordingly, compared to men ($M = 0.03$, $SD = 1.03$), $t(583) = -2.49$, $p = .032$.

Rater-Ratee Congruency Effects on Ratings of Performance

Due to insufficient data on Manager Race, Research Question 1 was not tested. Results of a two-step regression analysis (displayed in Table 12) indicated that Sex Congruency did not predict additional significant variance in subjective ratings of job performance, $R^2 = .354$, $\Delta R^2 = .003$, $F(1, 525) = 71.96$, $p = .113$, 90% CI [-.229, .004] (Research Question 2B was not supported).

Discussion

The purpose of this study was to examine the effects of minority status, being female or a member of a racial minority, on subjective ratings of job performance. This study examined potential biases in supervisor ratings of subordinate job performance in an attempt to parse rater bias from objective performance (sales performance). Previous studies (e.g., Hamner et al., 1974; Landy & Farr, 1980; Sackett, DuBois, & Noe, 1991; Schmidt & Johnson, 1973; Waldman & Avolio, 1991) have been unable to produce conclusive evidence that rater bias occurs during the subordinate performance evaluation process. The primary issue with previous studies was that they were unable to determine if any observed group differences in performance ratings are the results of rater bias or group differences in performance. This study did not find a significant effect for Sex

MINORITY BIAS IN SUPERVISOR RATINGS

Tokenism but did find a significant difference between the objective performance of men and women in instances when women comprised a small proportion of their respective workgroups. I found a significant difference in the performance of men and women beginning when women comprised as little as 30% of the workforce, with the difference becoming more pronounced as this proportion decreased. Given that no Sex Tokenism or direct Sex effects were found, it can be assumed that raters are not being biased in their ratings and instead that women are being rated commensurate to their performance, but performing significantly lower than they otherwise would in workgroups with greater than 30% of women. This suggests that there is a real difference in performance between men and women in certain instances and that observed differences are not determined by sex or race specifically, but instead are determined by the situations in which they are placed; this is particularly true for women.

This study is an extension of the research conducted by Sackett, DuBois, and Noe, (1991) and examines some of the same questions. Sackett, DuBois, and Noe recognized their limitations, namely their inability to ascertain whether observed differences in performance ratings between majority and minority employees were due to differences in performance or ratings bias. Due to their lack of objective performance data, such conclusions were not possible. This study examined the same hypotheses as Sackett, DuBois, and Noe with the use of performance ratings and objective performance data to better ascertain the cause of lower performance ratings received by minorities. It should be noted that the current study used percentage of minority group across job groups within organizations as opposed to individual workgroups as used by Sackett, DuBois, and Noe (1991).

MINORITY BIAS IN SUPERVISOR RATINGS

Incongruent with the findings of Sackett, DuBois, and Noe (1991) this study did not find a significant effect for tokenism of women. However, the current study did observe similar performance decrements in women when they comprised less than 20% of the workgroup, which is similar to the findings of Sackett, DuBois, and Noe. However, the current study also found performance decrements at thresholds up to 30% as well. Given that no significant tokenism effect was found, it can be assumed that the observed differences in performance ratings between men and women are not attributable to rater bias, but are instead related to actual differences in performance in these workgroups with small proportions of women.

Potential Causes of Observed Performance Decrements

Perhaps environmental work factors such as lack of support from the majority group (i.e., withholding important information, exclusion in important group meetings, or harassment), lack of support from their supervisors, or their own perceptions (i.e., feelings of exclusion or stress from inclusion in a minority group) may be causal explanations for these observed performance differences.

In a study by Sacco and Schmidt (2005) the effects of diversity misfit were examined and found to lead to higher turnover rates within groups with lower representation within an organization. They suggested that being demographically different from the rest of one's workgroup resulted in attitudes or beliefs that lead to turnover. As such, it is likely that being demographically different (i.e., female or racial minority) may result in additional attitudes or beliefs detrimental to performance. Additionally, it is likely that the most qualified minority employees would quit a job in which they experience these detrimental attitudes and beliefs related to their minority

MINORITY BIAS IN SUPERVISOR RATINGS

status, which would result in the remaining minority employees performing at an overall lower level. As such, these more qualified minority employees leaving the workgroup for more optimal situations could exaggerate any observed differences in performance for those minorities employed in workgroups where they comprise a disproportionately small percentage of the workgroup.

Theoretical Implications

These findings allow for potential explanations of the inconsistent findings of much of the previous literature concerning sex differences. This study provides evidence that observed differences in performance ratings between men and women is likely due to differences in performance, rather than rater bias. It is likely that much of the previous research lacking support for a sex effect on performance ratings contained samples with higher proportions of women. Similarly, those studies showing support for a significant effect of sex on performance ratings had likely utilized samples containing lower proportions of women.

An important consideration in previous studies, as well as the current study is the nested effects of the raters providing ratings for subordinates, as well as the organizations within each study. Given the likely sources of error within raters, such as halo, leniency, etc., many raters are likely to exhibit low discrimination in their ratings of their subordinates, rating them similarly instead. With approximately 39% of the variance in subordinates' ratings of job performance being attributable to the raters providing the rating in the current study, it is important to address these many potential sources of additional error. Additionally, there are many other factors which contribute to an employees' performance ratings beyond the rater and their own objective performance.

MINORITY BIAS IN SUPERVISOR RATINGS

Additional sources of variance may include customer bias, region-based bias, peer bias. For example, any bias towards an employee on the part of a customer will typically go unreported in any performance ratings and would likely have a negative influence on the observed objective performance of an employee.

Additionally, this customer bias may be more pronounced in regions in which the employee is a member of a particular minority group. As such, it is difficult to parse out actual employee performance from the many potential sources of error. It is likely that this difficulty in separating the many sources of variance in job performance are the cause of the relatively low correlation between objective and subjective measures of job performance. Given these many sources of variance of job performance, it is important to account for as many of these sources as possible in any studies of performance. Previous studies that have found significant correlations between sex or race and performance ratings likely have not accounted for enough sources of variance within job performance to ascertain a clear understanding.

The current study found significant correlations between race (Caucasian-A.A.) and subjective performance ratings, $r = -.06$, but no significant effect when controlling for objective performance and job tenure. Also, a significant correlation between subjective performance and sex congruency ($r = -.11$) was found, yet this effect was suppressed when controlling for tenure and objective performance. This suggests that job tenure and objective performance account for a greater proportion of the variance in performance ratings than do ratee sex or race.

It is possible that no tokenism effect was found for African-Americans as raters may be more likely to fear perceptions of discrimination towards racial minorities

MINORITY BIAS IN SUPERVISOR RATINGS

compared to women. There is also the possibility that women are biased towards in work settings where they comprise a small proportion of the workgroup and are thus perceived as unfit for this male dominated position within the given organization, resulting in lower performance, thus, lower ratings of performance. Supervisors of workgroups with less than 30% of women, may have a biased schema of the “model subordinate” being male in these male dominated settings.

Practical Implications

This study highlights the importance for proper rater training and some of the potential windfalls when biases continue uncorrected. Supervisors should be made aware of the potential biases that may occur as well as their personal schemas for the “model subordinate.” Additionally, it is increasingly important to monitor workgroup perceptions of minority groups as the proportion of these minority groups fall below certain thresholds ($\approx 30\%$ for women) as performance detriments below these levels may be due to environmental factors such as workgroup interactions.

Limitations

A limitation of this study is that no workgroups within the sample contained greater than 18% of African-Americans. This limitation prevented any conclusions being reached concerning objective performance differences between African-Americans and Caucasians when workgroups contained greater than 20% of African-Americans. Another limitation of this study was the lack of race data for raters, as Research Question 1A could not be tested. Another limitation of this study was the exclusive use of sales jobs in the sample. Results may differ for job types with varying degrees of availability of objective performance readily available for supervisors to refer to when providing

MINORITY BIAS IN SUPERVISOR RATINGS

performance ratings. Additionally, sales performance only accounts for a portion of total job performance for these positions and supervisor ratings are likely influenced by additional factors. As such, accounting for more sources of variance in employees' performance would provide more conclusive results. Another limitation of this study included the relatively large proportion of variance (approximately 39%) in supervisory ratings accounted for at the supervisor level. Analyzing the current data with a multilevel statistical approach would likely result in more conclusive findings, as well as allow for finding potential differences between various groups of raters (i.e., male vs. female raters, Caucasian vs. Minority raters, etc.).

Future Research

Future research should examine whether performance detriments disappear in African-Americans as they do for women above a certain threshold. As such, future studies should replicate the analyses of this study with samples including workgroups containing larger proportions of African Americans. Although no congruency effect was found in the current study, future studies should examine whether a tokenism effect is still evident in those workgroups of lower diversity for minorities with supervisors whom are members of the same minority group. Specifically, examining organizations possessing low diversity but employing a diverse selection of supervisors would provide additional insight into the specific situations in which minorities are biased. Another direction for future research is to examine job types other than sales jobs, which present opportunities to objectively measure job performance. Such a study would help determine which types of objective measures of job performance, and thus which job types, are more susceptible to rater bias. Another important future consideration is the

MINORITY BIAS IN SUPERVISOR RATINGS

examination of additional sources of variance in job performance to better understand potential biases in ratings of performance.

Conclusion

The current study examined the effects of sex and race on subjective ratings of job performance. No significant effects of sex, race, tokenism, or congruency were found; however, performance decrements for women were found when the proportion of women in a workgroup fell below 30%. This suggests that women are being rated fairly but their performance is influenced by factors not present when they comprise greater than 30% of a workgroup. Overall, rater sex and race accounted for only 0-1% of the variance in subjective ratings of job performance. Additionally, employee token status accounted for only 1-5% of the variance in subjective ratings of job performance. This study provides support for the idea that observed differences in ratings of job performance between men and women and Caucasians and minorities is not due to rater bias, or that a trivial amount of variance in performance ratings is attributable to sex or race of rater. Instead, these observed differences in performance ratings are likely due to other factors such as job tenure, actual performance, or environmental factors.

References

- Arvey, R. D. & Murphy, K. R. (1998). Performance evaluation in work settings. *Annual Review of Psychology, 49*, 141-168.
- BeVier, C. A. (1999). Racial differences in cognitive abilities: A meta-analysis. Ph.D. dissertation, Clemson University, United States -- South Carolina. Retrieved July 18, 2011, from Dissertations & Theses: A&I. (Publication No. AAT 9929716).
- Bommer, W. H., Johnson, J. L., Rich, G. A., Podsakoff, P. M., & Mackenzie, S. B. (1995). On the interchangeability of objective and subjective measures of employee performance: A meta-analysis. *Personnel Psychology, 48*, 587-605.
- Chockalingam, V., 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.
- Den Hartog, D.N., Boselie, P., & Paauwe, J. (2004). Performance management: A model and research agenda. *Applied Psychology: An International Review, 53*, 557-569.
- DeNisi A. S. & Peters, L. H. (1996). Organization of information in memory and the performance appraisal process: Evidence from the field. *Journal of Applied Psychology, 81*, 6, 717-737.
- Feldman, J. M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. *Journal of Applied Psychology, 66*, 2, 127-148.

MINORITY BIAS IN SUPERVISOR RATINGS

- Ferris, G. R., Judge, T. A., Rowland, K. M., & Fitzgibbons, D. E. (1994). Subordinate influence and the performance evaluation process: test of a model. *Organizational Behavior and Human Decision Making*, 58, 101-135.
- Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. (1974). Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. *Journal of Applied Psychology*, 59, 705-711
- Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. *Journal of Applied Psychology*, 73, 1, 68-73.
- Heneman, R. L. (1986). The relationship between supervisory ratings and results-oriented measures of performance: A meta-analysis. *Personnel Psychology*, 39, 811-826.
- Isen, A. M. & Baron, R. A. (1991). Positive affect as a factor in organizational behavior. *Research in Organizational Behavior*, 13, 1, 1-53.
- Jensen, A. R. (1980). Bias in mental testing. New York, NY: The Free Press.
- Kanter, R. M. (1977). Men and women of the corporation. New York, NY: Basic Books.
- Kinicki, A. J., Hom, P. W., Trost, M. R., & Wade, K. J. (1995). Effects of category prototypes on performance-rating accuracy. *Journal of Applied Psychology*, 80, 3, 354-370.
- Kraiger, K., & Ford, J. V. (1985). A meta-analysis of rater race effects in performance ratings. *Journal of Applied Psychology*, 70, 1, 56-65.
- Kraiger, K., Ford, J. K., & Schechtman, S. L. (1986). Study of race effects in objective indices and subjective evaluations of performance: A meta-analysis of performance criteria. *Psychological Bulletin*, 99, 3, 330-337.

MINORITY BIAS IN SUPERVISOR RATINGS

Landy, F. J. & Farr. J. L. (1980). Performance Rating. *Psychological Bulletin*, 87, 1, 72-107.

Latham, G. P., & Wexley, K. N. (1981). *Increasing productivity through performance appraisal*. Reading, MA: Addison-Wesley.

Lee, C. (1985). Increasing performance appraisal effectiveness: Matching task types, appraisal process, and rater training. *The Academy of Management Review*, 10, 2, 322-331.

Lent, R. H., Aurbach, H. A., & Levin, L. S. (1971). Predictors, criteria, and significant results. *Personnel Psychology*, 24, 519-533.

Martinko, M. J. & Gardner, W. L. (1987). The leader/member attribution process. *The Academy of Management Review*, 12, 2, 235-249.

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. *Journal of Applied Psychology*, 69, 4, 581-588.

Pulakos, E. D. (2004). Performance management.

Pulakos, E. D., Caucasian, L. A., Oppler, S. H., & Borman, W. C. (1989). Examination of race and sex effects on performance ratings. *Journal of Applied Psychology*, 74, 5, 770-780.

Rich, G. A. et al. (1999). Apples and apples or apples and oranges? A meta-analysis of objective and subjective measures of salesperson performance. *Journal of Personal Selling & Sales Management*, 4, 41-52.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, 54, 297-330.

MINORITY BIAS IN SUPERVISOR RATINGS

- Sacco, J. M. & Schmitt, N. (2005). A dynamic multilevel model of demographic diversity and misfit effects. *Journal of Applied Psychology, 90* (2), 203-231.
- 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 Caucasian-African-American differences in performance ratings. *Journal of Applied Psychology, 76*, 2, 263-267.
- Salgado, J.F., Moscoso, S., & Lado, M. (2003). Test-retest reliability of ratings of job performance dimensions in managers. *International Journal of Selection and Assessment, 11*, 98-101.
- Sanchez, J. L. & De La Torre, P. (1996). A second look at the relationship between rating and behavior accuracy in performance appraisal. *Journal of Applied Psychology, 81*, 1, 3-10.
- Sinclair, R. C. (1988). Mood, categorization breadth, and performance appraisal: The effect of order of information acquisition and affective state of halo, accuracy, information retrieval, and evaluations. *Organizational Behavior and Human Decision Processes, 42*, 22-46.
- Schmidt, F. L., Johnson, R. H. (1973). Effect of race on peer ratings in an industrial situation. *Journal of Applied Psychology, 57*, 237-241.
- Schmitt, N., & Lippin, M. (1980). Race and sex as determinants of the mean and variance of performance ratings. *Journal of Applied Psychology, 65*, 4, 428-435.
- Tangen, S. (2003). An overview of frequently used performance measures. *Work Study: A Journal of Productivity Science, 52*, 347-354.

MINORITY BIAS IN SUPERVISOR RATINGS

- Tziner, A., & Kopelman, R.E. (2002). Is there a preferred performance rating format? A non-psychometric perspective. *Applied Psychology: An International Review*, 51, 479-503.
- Viswesvaran, C., & Ones, D.S. (2000). Perspectives on models of job performance. *International Journal of Selection and Assessment*, 8, 216-226.
- Waldman, D. A., Avolio, B. J. (1991). Race effects in performance evaluations: Controlling for ability, education, and experience. *Journal of Applied Psychology*, 76(6), 897-901.
- Wexley, K. N. & Klimoski, R. (1984). Performance appraisal: An update. In K. M. Rowland & G. R. Ferris (Eds.), *research in personnel and human resources management* (Vol. 2, pp 35-79). Greenwich, CT: JAI Press.
- Woehr, D. J., & Huffcutt, A. I. (1994). Rater training for performance appraisal: A quantitative review. *Journal of Occupational and Organizational Psychology*, 67, 189-205.
- Wong, K.F., & Kwong, J.Y. (2007). Effects of rater goals on rating patterns: Evidence from an experimental field study. *Journal of Applied Psychology*, 92, 577-585.
- Yoder, J. D. (1994). Looking beyond the numbers: The effects of gender status, job prestige, and occupational gender-typing on tokenism processes. *Social Psychology Quarterly*, 57, 150-159.

Table 1.

Summary of Intercorrelations, Means, and Standard Deviations for Control Variables, Subjective Performance Ratings, and Objective Sales Performance.

Variable	M	SD	1	2	3	4	5	6	7	8	9
1. Sex	0.63	0.48	--								
2. Age	34.81	10.66	-.17**	--							
3. Tenure	46.67	83.23	-.23**	.50**	--						
4. Obj. Perf.	0.03	1.00	.03	.03	.15**	--					
5. Subj. Perf.	0.00	1.01	.01	-.10**	-.04	.14**	--				
6. % Women	29.93	11.90	-.21**	.52**	-.01	-.02	.00	--			
7. % A.A.	8.47	4.37	-.31**	.35**	.45**	.03	-.01	.48**	--		
8. % Hispanic	6.25	3.18	-.24**	.56**	.39**	.01	.00	.51**	.51**	--	
9. Rater/Ratee Congruency	0.57	0.49	.56**	.11*	.09*	.06	-.11**	-.31**	--	--	--
10. Race: White-A.A.	0.11	0.32	-.04	.10**	.12**	-.06*	-.06*	.03	.23**	.15**	-.02
11. Race: White-Hispanic	0.06	0.24	-.02	.07*	-.03	-.02	-.02	.10**	.00	.11**	.04
12. Race: White-Asian	0.02	0.16	.02	-.03	-.04	-.05	.03	.04	-.02	.10**	-.01
13. Race: A.A.-Hispanic	0.35	0.48	.03	-.04	-.19*	.05	.05	.15*	-.33**	-.02	.10

** $p < .01$ (two-tailed).
* $p < .05$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 2.

Multiple Regression Analyses Testing the Effects of Ratee Race (Caucasian-African-American) on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.05	-0.06	[-0.15, 0.04]
Ratee Sex	0.02	0.02	[-0.09, 0.13]
Ratee Age	-0.17*	-0.18*	[-0.23, -0.12]
Job Tenure	0.25*	0.26*	[0.19, 0.32]
Obj. Sales Performance	-0.06	-0.06	[-0.12, 0.01]
Race: Caucasian-A.A.		0.05	[-0.15, 0.24]
R^2	.06		.06
F	15.72*		12.60*
ΔR^2			.00
ΔF			0.16

Note. $N = 962$. CI = confidence interval.

* $p < .001$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 3.

Multiple Regression Analyses Testing the Effects of Ratee Race (Caucasian-Hispanic) on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.05	-0.05	[-0.15, 0.04]
Ratee Sex	0.01	0.01	[-0.10, 0.13]
Ratee Age	-0.18**	-0.18*	[-0.23, -0.12]
Job Tenure	0.26**	0.26**	[0.20, 0.33]
Obj. Sales Performance	-0.09*	-0.09*	[-0.16, -0.03]
Race: Caucasian-Hispanic		0.02	[-0.21, 0.24]
R^2	.06	.06	
F	15.31*	12.24*	
ΔR^2		.00	
ΔF		0.01	

Note. $N = 939$. CI = confidence interval.

* $p < .05$ (two-tailed).

** $p < .01$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 4.

Multiple Regression Analyses Testing the Effects of Ratee Race (Caucasian-Asian) on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.03	-0.04	[-0.13, 0.06]
Ratee Sex	-0.01	-0.01	[-0.13, 0.11]
Ratee Age	-0.18*	-0.18*	[-0.24, -0.13]
Job Tenure	0.26*	0.26*	[0.19, 0.33]
Obj. Sales Performance	-0.07	-0.07	[-0.14, -0.00]
Race: Caucasian-Asian		0.21	[-0.13, 0.55]
R^2	.06		.06
F	15.10*		12.29*
ΔR^2			.00
ΔF			1.04

Note. $N = 906$. CI = confidence interval.

* $p < .001$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 5.

Multiple Regression Analyses Testing the Effects of Ratee Race (African-American-Hispanic) on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.13	-0.12	[-0.15, 0.04]
Ratee Sex	0.09	0.09	[-0.09, 0.13]
Ratee Age	-0.05	-0.05	[-0.23, -0.12]
Job Tenure	0.23	0.23	[0.19, 0.32]
Obj. Sales Performance	-0.23	-0.23	[-0.12, 0.01]
Race: A.A-Hispanic.		-0.03	[-0.15, 0.24]
R^2	.03	.03	
F	1.13	0.90	
ΔR^2		.00	
ΔF		0.02	

Note. $N = 137$. CI = confidence interval.

* $p < .001$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 6.

Multiple Regression Analyses Testing the Effects of Ratee Race (Any Racial Minority) on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.03	-0.04	[-0.13, 0.05]
Ratee Sex	0.01	0.01	[-0.10, 0.11]
Ratee Age	-0.16*	-0.17*	[-0.22, -0.11]
Job Tenure	0.23*	0.24*	[0.17, 0.30]
Obj. Sales Performance	-0.05	-0.05	[-0.11, 0.01]
Race: Any Racial Minority		0.09	[-0.05, 0.22]
R^2	.05	.05	
F	14.36*	11.72*	
ΔR^2		.00	
ΔF		1.15	

Note. $N = 1078$. CI = confidence interval.

* $p < .001$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 7.

Multiple Regression Analyses Testing the Effects of Ratee Sex on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.03	-0.03	[0.21, 0.67]
Ratee Age	-0.16*	-0.16*	[-0.02, -0.01]
Job Tenure	0.23*	0.23*	[0.16, 0.31]
Obj. Sales Performance	-0.05	-0.05	[-0.12, 0.02]
Ratee Sex		0.01	[-0.13, 0.12]
R^2	.05		.05
F	19.16*		14.36*
ΔR^2			.00
ΔF			0.01

Note. $N = 1078$. CI = confidence interval.

* $p < .001$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 8.

Multiple Regression Analyses Testing the Effects of Minority Status (Race or Sex Minority) on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.03	-0.04	[-0.11, 0.03]
Ratee Age	-0.16*	-0.17*	[-0.22, -0.11]
Job Tenure	0.23*	0.23*	[0.17, 0.30]
Obj. Sales Performance	-0.05	-0.05	[-0.11, 0.01]
Minority Status		0.03	[-0.08, 0.13]
R^2	.05	.05	
F	19.16*	14.41*	
ΔR^2		.00	
ΔF		0.19	

Note. $N = 1078$. CI = confidence interval.

* $p < .001$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 9.

Multiple Regression Analyses Testing the Effects of Race Tokenism-AA on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.13	0.14	[-0.56, 0.83]
Ratee Sex	0.20	0.08	[-0.39, 0.56]
Ratee Age	-0.06	-0.05	[-0.28, 0.18]
Job Tenure	0.20	0.21	[-0.07, 0.49]
Obj. Sales Performance	0.09	0.08	[-0.21, 0.37]
Race Tokenism-AA		-0.02	[-0.06, 0.03]
R^2	.04	.05	
F	0.80	0.74	
ΔR^2		.01	
ΔF		0.53	

Note. $N = 80$. CI = confidence interval. Race tokenism-AA was defined as the proportion of respective workgroups comprised of African-Americans.

* $p < .001$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 10.

Multiple Regression Analyses Testing the Effects of Race Tokenism-Hispanic on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.09	-1.00	[-1.92, -0.08]
Ratee Sex	-0.02	0.13	[-0.32, 0.58]
Ratee Age	-0.11	-0.15	[-0.40, 0.10]
Job Tenure	0.32	0.29	[-0.06, 0.65]
Obj. Sales Performance	-0.59**	-0.61**	[-0.90, -0.32]
Race Tokenism-Hispanic		0.10	[0.01, 0.20]
R^2	.18	.23	
F	2.90*	3.06*	
ΔR^2		.05	
ΔF		3.22	

Note. $N = 57$. CI = confidence interval. Race tokenism-Hispanic was defined as the proportion of respective workgroups comprised of Hispanics/Latinos.

* $p < .05$ (two-tailed).

** $p < .01$ (two-tailed)

MINORITY BIAS IN SUPERVISOR RATINGS

Table 11.

Multiple Regression Analyses Testing the Effects of Sex Tokenism on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	-0.06	-0.41	[-0.80, -0.02]
Ratee Age	0.04	-0.01	[-0.12, 0.11]
Job Tenure	-0.05	-0.03	[-0.15, 0.09]
Obj. Sales Performance	-0.06	-0.07	[-0.18, 0.04]
Sex Tokenism		0.01	[-0.00, 0.02]
R^2	.01	.01	
F	0.77	1.17	
ΔR^2		.01	
ΔF		2.35	

Note. $N = 374$. CI = confidence interval. Sex tokenism was defined as the proportion of respective workgroups comprised of a specific minority group (women in this analysis).

MINORITY BIAS IN SUPERVISOR RATINGS

Table 12.

Multiple Regression Analyses Testing the Effects of Sex Congruency on Subjective Supervisor Ratings of Ratee Job Performance.

Variable	Model 1 <i>B</i>	Model 2	
		<i>B</i>	90% CI
Constant	0.06	0.05	[-0.07, 0.17]
Ratee Age	-0.11	-0.11	[-0.24, 0.01]
Job Tenure	-0.16*	0.16*	[-0.28, -0.04]
Obj. Sales Performance	-0.47**	-0.47**	[-0.58, -0.35]
Sex Congruency		0.06	[-0.19, 0.31]
R^2	.36	.37	
F	35.09**	26.24**	
ΔR^2		.00	
ΔF		0.18	

Note. $N = 188$. CI = confidence interval. Sex congruency was defined as the congruency of rater and ratee with respect to sex (e.g., rater and ratee of the same sex would constitute congruency).

* $p < .05$ (two-tailed).

** $p < .01$ (two-tailed).

MINORITY BIAS IN SUPERVISOR RATINGS

Figure 13.

Predicted Performance Ratings for Women in Workgroups Comprised of Less than 20% Women Compared to Workgroups Comprised of Greater than 20% Women.

