

AN EXPLORATORY STUDY OF BIAS IN THE NOMINATIONS OF
BEST AND WORST WORKER IN THE FORM OF
CRITICAL INCIDENTS OF JOB PERFORMANCE

A THESIS

Presented to

The Faculty of the

Division of Graduate Studies

by

Steven L. Johnson

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Psychology

Georgia Institute of Technology

June 1978

AN EXPLORATORY STUDY OF BIAS IN THE NOMINATIONS OF
BEST AND WORST WORKER IN THE FORM OF
CRITICAL INCIDENTS OF JOB PERFORMANCE

Approved:

W. W. Ronan, Chairman

C. M. York

Sherman Dallas

Date Approved by Chairman 5/22/78

ACKNOWLEDGMENTS

The author wishes to thank his thesis committee, Drs. Ronan and York in Psychology and Dr. S. F. Dallas in Industrial Management, for providing guidance, assistance, and support throughout this study. I am especially grateful to Dr. Ronan, my chairman, teacher, colleague, and friend, for his effort and magnanimous giving of his time. Also acknowledged is the considerable assistance I received from faculty and students in the completion of this thesis.

This thesis is dedicated to the author's wife and parents, whose love and support were necessities during his graduate experience at Georgia Institute of Technology.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
LIST OF TABLES	iv
SUMMARY	v
Chapter	
I. INTRODUCTION	1
Selected Literature: Human Performance	
Appraisal	
Study Purpose	
Hypotheses	
II. METHODOLOGY	16
Critical Incident Technique	
Sample Population	
Statistical Analysis	
III. RESULTS	26
IV. CONCLUSION	39
Limitations of the Research	
Future Research	
APPENDICES	46
REFERENCES	66

LIST OF TABLES

Table	Page
1. Subordinate Composites for Different Race and Sex Categories of Supervisors	22
2. Overall Analysis Using Supervisor Race and Sex by Subordinate Race and Sex	27
3. Analysis Using Subordinate Variables Race and Sex	29
4. Analysis Using Subordinate Variable Race.	31
5. Analysis Using Subordinate Variable Sex	33
6. Analysis of White Female Supervisor Nominations Using Subordinate Variable Race	34
7. Analysis of Black Female Supervisor Nominations Using Subordinate Variable Race	36
8. Analysis of White Male Supervisor Nominations Using Subordinate Variable Race	37

SUMMARY

Unfair discrimination is a current problem in employment practices and is the subject of special concern to the Industrial and Organizational Psychologist. The literature reviewed in this study generally found the performance ratings of black people to be worse than their white peers. The literature review included performance appraisal and variables effecting performance appraisal which might have an influence on this general finding.

The purpose of this study was an exploration to determine whether the nominations of best and worst workers in the form of critical incidents of job performance are "biased". In this study, the operational definition of bias is a deviation of the observed frequency from the expected frequency of nominations. The expected frequency assumes the assignments of nominations to be a chance event, independent of race or sex of subordinate.

A group method of data collection was used to get the critical incidents of job performance. This group method and the Critical Incident Technique are well documented procedures (Flanagan, 1954, 1974).

The results showed there was a significant bias in nominating workers based on race of subordinates. There was no significant bias in nominating workers based on sex

of subordinates. The bias found in nominating best and worst workers seemed to have an exclusive source in the white female supervisors. This bias yielded more negative nominations and less positive nominations than expected for the black subordinates. More positive and less negative nominations than expected were reported for white subordinates.

CHAPTER I

INTRODUCTION

Unfair discrimination is a current problem in employment practices. The distinction between "necessary" and "unfair" discrimination was made more than a decade ago by a prominent Industrial and Organizational Psychologist. Necessary discrimination is the employer's right to be able to differentiate treatment of qualified and unqualified applicants. "Unfair discrimination exists when persons with equal probabilities of success of the job have unequal probabilities of being hired for the job" (Guion, 1966, p. 26).

The focus of this thesis is an exploration to determine whether the nomination of workers as one's best or worst worker was "biased". The nominations of workers was made in the form of a critical incident of job performance behavior (critical incidents will be defined in the Methods sections). The nomination of the best worker will be called the positive nomination. The nomination of the worst worker will be called the negative nomination. In this study, the operational definition of bias was a deviation of the observed frequency from the expected frequency of reported nominations. The expected frequency assumes the assignments of nominations to be a chance event, independent of race or

sex of subordinate.

In the personnel selection/evaluation literature particular attention has been drawn to testing and charges of test bias. Test bias occurs when too high or too low a criterion score is consistently predicted for numbers of a subgroup by a particular selection device. These subgroups are those for which a historical basis for discrimination has existed (i.e. race, sex, religion, etc.).

It has been pointed out by Einhorn and Bass (1971) that two elements are involved in the validation of a test: the predictor (the test) and the criterion (some evaluation of job performance). These two components are equally important and deserve critical analysis to avoid bias. However, little research has been reported with regard to the investigation of bias in criteria of performance evaluation.

A major problem concerning these investigations is that the criteria "...can be subjected to no wholly satisfactory empirical test of its adequacy. The criterion must, consequently, be logically justifiable as valid in its own right" (Brogden and Taylor, 1950, p. 160). Requirements of a criterion, according to these test specialists (1950), are "... the criterion should give an accurate and unbiased measure of the extent to which individuals in the validation population contribute to or detract from the efficiency of the organization" (pp. 160).

A bias in the criterion would be any variable, except measurement errors and sampling errors, which cause the obtained criterion scores to deviate from the "true" criterion scores. Brogden and Taylor's (1950) classification of biasing factors includes:

1. Criterion deficiency - the omission of necessary elements in the criterion.
2. Criterion contamination - the inclusion of extraneous elements in the criterion.
3. Criterion Scale Unit bias - the inequality of scale units in the criterion.
4. Criterion Distortion - improper weighting in combining elements in the criterion.

The most important aspect of criterion bias is its effect on the validity of the predictor. If the bias correlates with the predictor, the effect is the reduction or elimination of predictive validity when correlated with the "true" criterion. On the otherhand, "test free" bias, bias which does not correlate with the predictor, acts only as an error in measurement in regard to its effect on the validity coefficient. The relative magnitude of the validity and the partial regression coefficients remain unaffected in this instance.

There are, however, two potential side effects of "test-free" bias. One, it can render statistics less stable from sample to sample. Second, it may distort estimates

of the reliability of the criterion in an indeterminate manner. While the effect of predictor related bias is undeniably detrimental the effect of "test free" bias is not so acutely determined.

Inn, Hulin, and Tucker (1972) found evidence that individual differences in performance measures were dimensional. Therefore, they argue, "optimum prediction is achieved only if the dimensions of individual differences on the criterion variable are represented similarly on the predictor variable" (pp. 81). This same point has been stressed by Ronan and Prien (1971). This finding supports the contention that any bias should be considered undesirable.

Selected Literature: Human Performance Appraisal

A major question to be asked of human performance appraisers is whether they use or attempt to use relevant job performance information as opposed to irrelevant information. In a study by Wiener and Schneiderman (1974) it was found that interviewers rely on relevant job information. However, the availability of relevant job information did not eliminate the use of irrelevant information. The same result was found whether experienced or inexperienced interviewers were used. The use of relevant information is also supported by the finding that supervisor ratings are linearly predictable based on the supervisor's knowledge of the ratee's objective performance (Casio and Valenzi, 1978).

There is mixed support for ethnic group differences in the relationships of job performance criteria. Bass and Turner (1973) found a significant difference between blacks and whites in the relationship of subcriteria to an overall job performance criterion. Their results showed that objective subcriteria had a higher correlation with the overall job performance measure for blacks than for whites; whereas, the subjective subcriteria had a higher correlation with the overall performance measure for the white group than the black group. Salary increases for blacks correlated more highly with attendance records than they did for whites. Salary increases for whites seemed to be linked to supervisor ratings. Arnold (1968), however, found no significant difference between measures of an overall job effectiveness criterion for two ethnic groups working on the same job. The subcriteria accounted for variance in the overall measure similarly for both racial groups and there was no difference in factor structure of the subcriteria for the two ethnic groups. A restriction of variance in the ratings may have occurred due to the extremely low level of skill required for the task performance of wire solderers in this study.

Similar inconsistency can be found in this same topical area with regard to sex differences. Different variables were perceived as important for the evaluation of male applicants than for the evaluation of female applicants (Cecil et. al., 1973). Factor analysis of these variables

also produced distinct separate factors for males and females. A study done to look at the influence of sex of incumbent and sex of analyst on job analysis results showed no inflation or devaluation in any treatment condition (Avery et. al., 1977).

Either one or both of the studies in these inconsistent pairs of studies relied on ratings as a measure of performance. In this writer's opinion, ratings as a measure of job performance and specific discussion of sexual and racial effects in ratings seems appropriate, particularly since they are widely used as a performance measure.

O'Reilly (1973) found large discrepancies between a supervisor's perception of a subordinate's skills and knowledge and the subordinate's perception of his skills and knowledge. A factor analysis of an intercorrelation matrix of 20 variables rated from three vantage points - self rating, peer rating, and supervisor ratings - revealed three orthogonal factors labelled for each of the three vantage points (Klimoski and London, 1974). Generally, the difference in ratings across groups is not only a magnitude difference but also a dimensional difference.

It is frequently "...found that superiors do not agree with peers of the people being rated about just what the given job actually entails. That is, they are actually rating two different jobs when they rate performance on what is supposedly a unitary, well-defined job" (Ronan and Schwartz,

1974, pp. 75).

Relatively few studies have been done on the accuracy of ratings. Gordon (1970) compared the relative accuracy of ratings of correct behavior compared to the ratings of incorrect behavior. The results showed a difference in the accuracy of these ratings, the ratings of correct behavior had significantly greater accuracy than the ratings of incorrect behavior. The results of a peer nomination study (Kaufman and Johnson, 1974) found the frequency of positive peer nomination to be the best predictor of performance, but negative peer nomination accounted for very little of the variance in performance. This result supports the difference in accuracy of negative and positive ratings in the prior study.

Variability of performance and pattern of performance (ascending or descending) does not seem to affect overall performance ratings (Scott and Hamner, 1975). These two factors do appear to trigger stereotypic responses in ratings of ability and motivation. Workers with high variability in performance were rated as having greater ability than workers with low variability in performance. Workers with ascending patterns of performance were rated as being more motivated than workers with descending patterns. Also workers with low variability of performance were rated as being more motivated than were the workers with high variability of performance.

Personal consequences and immediacy have been shown to influence ratings (Sewell, 1973). Ratings were considerably more conservative when there were personal consequences to the rater. Ratings were also more conservative when made in close time proximity to the performance being rated.

Attitudinal differences in the rater can effect performance ratings (Klores, 1966). Raters who placed greater emphasis on "initiating structure" (a leadership attitude variable) gave significantly lower ratings than did raters who placed greater emphasis on "consideration" (a leadership attitude variable). Further, the same raters who placed more emphasis on "initiating structure" had greater variance in their ratings than did the more "considerate" group.

Two types of studies have caused serious questions to be raised concerning the validity of ratings. Attempts to show convergent validity between ratings of performance and objective measures of performance have generally had a dismal outcome. In a review of studies of this type it was found that "...ratings show little or no relative to objective measures of performance" (Ronan and Schwartz, 1974, pp. 76). Factor analytic studies have also caused similar concern. These studies also provide evidence of the lack of congruence between performance and ratings. The studies reviewed by Ronan and Schwartz (1974) showed ratings loaded on a single factor separate from the other indices of performance.

Race and sex differences in ratings are specifically germane to the present study. It is hoped that one will keep

in mind these factors which influence ratings in general and consider the added complexity of including race and sex differences. Further, the potential for interactions between these general influencing factors and race and/or sex differences is very complex and has serious implications for the adequacy of performance appraisals.

One principal racial difference is a rather clear tendency for supervisors to give higher ratings to one's own ethnic group (Campbell, 1972; DeJung and Kaplan, 1961; Quinn, 1969). This tendency seems to be supported by a general review of the rating literature for black and white differences. The studies reviewed, which made note of supervisory race, had a predominance of white supervisors. This could be the reason for the relatively lower ratings for black than white subordinates (Arnold, 1968; Gavin and Ewen, 1974; Greenhaus and Gavin, 1972; Huck and Bray, 1976; Quinn, 1969; Schmidt and Hill, 1977).

Campbell (1972) also found that black raters had higher "validities" when rating black incumbents than when rating white incumbents. However, Mexican-American and Caucasian raters had higher validities when rating members of ethnic groups other than their own. In his conclusion Campbell thought that "...there is little doubt that ethnic group of rater and ratee does make a difference" (pp. 16). Anastasi (1972) in concurrence stated, "The results certainly suggest that ratings are a questionable type of criterion measure for test validation when different ethnic groups

are involved" (pp. 85).

There is some evidence which shows that a similar sex influence occurs in ratings. Rosen and Jerdee (1974a, 1974b) in a series of studies showed that females were unfairly discriminated against in personnel decisions involving promotion, development, and supervision. Males were discriminated against in personnel decisions resulting from competing role demands of work and family circumstances. The evaluations of male applicants were higher and they were accepted for managerial positions more frequently than were females. Another study (Cline et. al., 1977) found that opposite sex combinations of rater-ratee produced devalued ratings in comparison to same sex rater-ratee pairs.

There has been some research, although very little, concerned with reducing racial effect in ratings. Schmidt and Johnson (1973) found that in an industrial setting in which 50% of the workers were black and 50% were white and where a human relations training program was used, no race effect in peer ratings was found. Schmidt and Hill (1977) also suggest the use of peer ratings to minimize race and sex effects. They further suggest the use of methods using actual observation of behavior evaluations. In support of this second suggestion (Norton et. al., 1977) it was found that the use of behaviorally based scales showed less leniency, less variance, and no rater sex-ratee sex interaction.

The use of behaviorally based scales have generally been shown to be superior to more subjective ratings (Borman and Dunnette, 1975; Campbell et. al., 1973). However, the use of such behaviorally based scales has not eliminated the race and sex effect. Hamner, et. al. (1974) and Bigoness (1976) found a race and sex influence despite the use of two very distinct performance levels and using a behaviorally based "objective" measure of performance. The results (Hamner et. al., 1974) found black males were rated as about average despite their level of performance.

The existence of prejudice in interracial relations has been a proverbial boiling pot for years. The areas and nature of prejudice are numerous, complex, and dynamic. McGuire (1973) found that police officials were biased in their reporting of accidents. The biasing influences were sex and occupation. It was also found that these officials were biased in issuing citations. In the case of issuing citations, the officials were biased according to race and sex.

Prejudice is frequently covert and almost invariably denied, whether vocally or as the result of the more blatant racial attitude measures. Porter (1974) found that a measure of intimacy allowance to blacks predicted opinion change from black or white speech sources more accurately than did a measure of prejudice.

The dynamic nature of racial attitudes was exhibited in a study by Hamm et. al. (1975). They found mere exposure to photographs of black people can yield an increase in

favorability towards blacks. In an extension of this study results showed interpersonal attraction of either whites or blacks to stimulus groups of whites or blacks can also be increased by exposure to photographs of that stimulus group.

One should also look at the work environment to assess the forms and sources of prejudice. Haefner (1977) found that white workers preferred to work with whites rather than blacks. Females indicated no bias and younger workers showed less racial preference than did older workers. The results showed an equivalent effect for blacks - black workers would prefer to work with blacks.

Ledvinka (1977) states that one of the interracial rules is that blacks should not express interracial conflict or hostility to whites. He then hypothesizes that a black job seeker would not offer reasons of rejection (the workers rejection of the company or the company's rejection of the worker) to a white interviewer. The results of the study supported this hypothesis. Black job seekers offered more rejection reasons and fewer nonrejection reasons for leaving a job to black interviewers. Black job seekers conversely offered fewer rejection reasons and more non-rejection reasons for leaving a job to white interviewers.

Controversial results in studies of interracial supervision were found. Richards and Jaffee (1972) found performance ratings of black supervisors were lower than the performance ratings of white supervisors. Parker (1976)

found the subordinates view of supervisor's behavior to be a function of three things: supervisor's race and role, race of subordinate, and numerical status of racial groups in the work groups. Contrary to the previous finding, black supervisors were seen as more effective by all subordinates than a white supervisor.

The topic of supervisory behavior toward subordinates is also important to this discussion. A study of supervisory coercion (Kipnis et. al. 1973) revealed the frequency of coercion used by supervisors with black subordinates was significantly greater than for whites. In this study there was no difference in length of employment of the two groups nor in the kinds of problems manifested by white or black workers.

This review has covered criterion bias, ratings, interracial relations, and racial prejudice. Apparent conclusions are: criterion bias is a serious problem, prejudicial race attitudes do exist, and performance ratings have been shown to contain race and sex effects from rater-ratee combinations of these variables. This problem with performance ratings is very serious in regard to equal opportunity in the employment world. It appears research efforts should be directed toward a thorough investigation of the problem, causes, and methods to eliminate the problem. One aspect of the problem is to determine whether or not supervisors "see" job performance differences as related to

sex and/or race of subordinates.

Study Purpose

This exploratory study is to assess the nomination of workers as best or worst for possible bias by race or sex. Race and sex of supervisor and subordinates, and corresponding interaction effects, represent the principal variables studied. It is acknowledged that this scope cannot determine the effect on the criterion; however, the question seems to merit investigation in its own right. Positive results would show that a criterion developed from behavioral incidents describing the nominations which were biased might not proportionately represent the actual work behaviors of black and white workers. It is hoped that at the very least this study can provide valuable insight for future investigations of this problem.

Hypotheses

H_1 - Test of H_0 - that the frequency of positive and negative nominations does not differ from that expected if the nominations are chance events, independent of race of subordinate.

H_2 - Test of H_0 - that the frequency of positive and negative nominations does not differ from that expected if the nominations are chance events, independent of sex of subordinate.

H_3 - Test of H_0 - that the frequency of positive and negative nominations does not differ from that expected if

the nominations are chance events, independent of an interaction based on race x sex of subordinate.

CHAPTER II

METHODOLOGY

Critical Incident Technique

The critical incident technique consists of a set of procedures for collecting direct observations of human behavior in the performance of some specified activity. The technique is a procedure for collecting observed incidents having special significance and meeting systematically defined criteria (Flanagan, 1954).

An incident is an observed human behavior which is itself a completed action related to some defined behavioral outcome. Essentially the intent is to collect objective, observable behaviors to serve as the basis for the development of performance criteria. The incident must occur and be observed in a situation where the action and its consequences are determined to be critical to performance, either negatively or positively.

The critical incident technique is aimed at obtaining certain important facts concerning behavior in a defined situation. The method is a flexible set of principles which allows for modification to adapt to the specific needs of a particular situation. The ultimate outcome is a description of observable behaviors that are "critical" to effective performance in a defined activity.

The basic principles of the technique are: only simple judgments are required of the observer; only qualified observers are used; the observations must meet predetermined standards for the purpose of the study. The objectivity of the reported observation depends upon the precision which the studied characteristic is defined by the experimenter and the competence of the observer in the capacity of observation.

There are five specific conditions (Flanagan, 1951) which must be met in collecting critical incidents. First, the incidents must be observations of actual performance. Second the objective and consequences of the activity must be known to the observer. Third, the standards (or rules) for the specific judgments must be clearly defined. Fourth, the observer must be qualified to judge the behavioral outcomes of the actions and the observer must have the opportunity to make actual observations of work behavior. Fifth, care must be exercised to ensure accuracy, i.e. require that only "relatively" recent observations be used and that subjective inferences of traits or opinions be eliminated.

The standards or rules are described in the instructions and are a very important element of the critical incident technique. These precise instructions focus observer attention on behavior crucial to the definition of the work activity. The standards specify that it is behavior that

is "particularly effective" in attaining the goals of the activity that are of interest. Further, the standards for the incidents must describe the situation in which the observations are made; the relevance of the incidents to the general purpose; the extent of the effect of the incidents on the general purpose of the work activity; the persons to make the observations.

There are various forms of collecting critical incidents. A primary distinction is generally made whether the reports are recalled or made at the time of the observation. Campion et. al. (1973) found that both methods were equally effective in providing reliable behavioral incidents.

In this study a group interview method was used to collect incidents. The group interview method has been used frequently since its introduction by Ronan (1953). The method seems to maximize research efficiency with no noticeable reduction in the reliability or accuracy of the incidents. The incidents are recorded by the observers on forms submitted to them (see Appendix A). One advantage of this method is that the incidents are recorded in the observer's own language.

The analysis of the data is primarily intended to increase the usefulness of the data while sacrificing as little comprehensiveness, specificity, and validity as possible. The main considerations involving this step are: the frame of reference, category formulation, and general

behaviors. The frame of reference delimits the ways in which a given set of incidents can be classified. The principle consideration here is how the data are to be used.

The category formulation is largely subjective rather than objective. The quality and usability of the categories formulated are mainly dependent upon the ability of the formulator. This category formulation is divided into five stages: the identification of incidents as "critical" behavior, the development of a classification system which facilitates the grouping and ordering of similar behaviors, placing the incidents into the categories developed, subgrouping the categorized incidents so a behavioral statement about the subgroup can be written, and to reduce the number of descriptive statements which represent the incidents.

The reliability of the categories is measured by the amount of inter-rater agreement between two or more independent raters. The result of this reclassification can indicate where the grouped behavior may be inconsistent or the category may be inappropriately defined. In this study an agreement of .95 for positive incidents and .91 for negative incidents between two independent raters was obtained.

The last problem is to determine the specificity-generality of behaviors to be used. The problem is to weigh between the use of more specific incidents or the use of fewer rather general headings. Usually, this decision is dependent upon how the data are to be used.

The critical incident technique has been shown to produce criteria which are reliable, relevant, and show substantial content validity (Andersson and Nilsson, 1964). These results have been corroborated and substantial concurrent validity using criteria developed from this method was also found (Ronan and Latham, 1974).

Flanagan (1974) believes that this technique has maximized objectivity while minimizing subjective components of inference and interpretation. He believes that the slight deficiency in objectivity is due to current limitations in psychological knowledge. Primarily this limitation lies in the development of a classification system.

Sample Population

The sample of supervisors for this research consisted of virtually every first line supervisor in a government agency. The only exceptions were people who for personal reasons could not be present or could not be scheduled.

Prior to conducting the research, criteria for eliminating subjects were established. One criteria was that each supervisor had to supervise both blacks and whites. Also, supervisors were eliminated if they did not report any incidents. One problem in this regard was a black female who reported a positive incident but not a negative incident. This subject was retained in the positive analyses and eliminated from the negative analyses. This explains the

difference in probability estimates (expected) between positive analysis groups and negative analysis groups which contain black female supervisors. Neither the inclusion of this subject nor exclusion of this subject altered the results for any analysis appreciably. One supervisor was completely eliminated for not reporting any critical incidents. Seven supervisors were eliminated due to violations of the requirement of having to supervise both black and white subordinates. All seven of those eliminated had no black subordinates. Of these supervisors two were white males and the rest were white females.

The sample of supervisors studied consisted of eleven black females, ten white males, and sixty white females. They represent over 90% of the supervisors in the agency. The race and sex breakdown of subordinates supervised by these supervisor categories is found in Table 1. (The number of subordinates in each subordinate category for individual supervisors is given in Appendix D.)

This sample of supervisors was somewhat unusual in that they had been trained in writing critical incidents quite extensively by this agency. Further, the supervisors were required to submit behavioral incidents about their subordinates along with the performance evaluations of the subordinate. This fact was reflected in the quality of incidents reported and the lack of questions directed to the interviewer. The instructions were standardized for

Table 1. Subordinate Composites for Different
Race and Sex Categories of Supervisors.

Subordinates	Supervisors		
	Black Female	White Male	White Female
Black Males	4	8	19
White Males	9	31	110
Black Females	88	24	315
White Females	146	81	905

Total number of subordinates supervised is 1,740.

all groups of supervisors (see Appendix B).

The interviews were conducted in groups of four to eight supervisors. The most frequent number of supervisors interviewed was six. Six was also the number of subjects per interview requested. The scheduling of supervisors was the sole responsibility of the government agency.

After the instructions were given, a folder containing the incident forms and a demographic data questionnaire (see Appendix A) about the composition of each supervisor's work group was administered. At the bottom of each incident form was a short demographic data questionnaire to describe the subordinate in the reported incident. The supervisors were instructed to give their best estimate of certain variables which had caused considerable concern due to uncertainty, i.e. age and education level. These variables which were causing concern were not variables relevant to the study; they were included to camouflage the primary intentions of the study. Supervisor race and sex were noted along with an identifying number of the folder given them while the supervisors were filling out the forms.

When the supervisors completed the forms they brought them to the interviewer. They were instructed to allow the interviewer to peruse their forms before they left. This served a dual purpose - one, the interviewer could check to ensure that behavioral incidents were actually reported; second, idiosyncratic terms could be explained and vague

incidents could be more clearly defined.

In general, the incidents obtained were of high quality, probably as a result of prior training and experience on the part of the supervisors. Only minor editing or clarification was necessary on the part of the experimenter.

As previously stated, inter-rater reliabilities of .95 and .91 were obtained for the categorization of positive and negative incidents, respectively. The categories formulated and the frequencies including racial breakdown of nominations for each behavioral statement were included (see Appendix C).

Statistical Analysis

A statistic computationally equivalent to Pearson's "Chi-Square" statistic for testing goodness of fit was used in the data analyses. The expected estimates were the number of nominations occurring in a given category that probability would predict to occur by chance. This estimate assumes the assignments of nominations to be a chance event, independent of race or sex of subordinate. The operational definition of bias is a deviation of the observed frequency from the expected frequency of reported nominations provided by the supervisors.

The expected frequency is the expected total number of subordinates nominated that would fall in a given subordinate category if each supervisor's nomination is made in such a way that each of his/her subordinates is equally likely

of being chosen. The expected frequency is the summation of probabilities derived from each individual supervisor. To get the probabilities for each supervisor the number of subordinates in each category of subordinates was divided by the total number of subordinates under that supervisor. To get the expected frequency of the nominations on black females who were supervised by white males, for example, one would summate the probabilities of nominating on this subordinate category across all white male supervisors.

The probability of a nomination being reported for a given subordinate category is not the same for all supervisors. The classical Pearson Chi-square statistic uses N independent observations where the probability of observations is distributed identically in all trials. Since the Chi-square is such a generally applied statistic, this use of the Chi-square, while a deviation from the classical use, is believed to approximate the Chi-square distribution reasonably well (Walker, 1978). The author could find no reference in the literature dealing with this deviation from the classical use of the Chi-square; however, there is evidence to support and justify this present use as an approximation of the Chi-square distribution (see Appendix E).

The Yates' correction was used in all the Chi-square analyses to correct for 1 degree of freedom in some cases and to correct for the small expected frequencies in other cases (Kolstoe, 1973).

CHAPTER III

RESULTS

The Chi-square test for goodness of fit was used in the data analyses. The analyses test the observed distribution against the Chi-square distribution which assumes the assignments of nominations to be a chance event, independent of race or sex of subordinate.

Table 2 portrays the overall Chi-square analysis of the positive and negative nominations made by the agency supervisor. The black male subordinates and white male subordinates have been pooled because of the small number of male subordinates in the study. The expected frequency (E) is the summation of probabilities from the individual supervisor in the three supervisor categories of nominating subordinates of different categories (this procedure is described in the Methodology section). The observed frequency (O) is the actual frequency of nominations reported by each supervisory group on the different subordinate categories.

Chi-squares were calculated for each of the three supervisor categories for both the positive and negative nominations. The Chi-square values for positive nominations are .6778, .0355, and 9.0747 for black female supervisors, white male supervisors, and white female supervisors, respectively. The results for black female supervisors and white

Table 2. Overall Analyses Using Supervisor
Race and Sex by Subordinate Race
and Sex

<u>Positive Nominations</u>				
Subordinates		Supervisors		
		Black female	White male	White female
Male	O	1	3	5
	E	.5789	2.7084	5.7376
Black female	O	2	2	4
	E	3.919	1.6667	14.0104
White female	O	8	5	51
	E	6.502	5.625	40.252
Chi-square values		.6778	.0355	9.0747*
<u>degrees of freedom</u>		2	2	2
*p-value is less than .025, but greater than .01				
<u>Negative Nominations</u>				
Subordinates		Supervisors		
		Black female	White female	White female
Male	O	1	2	10
	E	.5262	2.7084	5.7376
Black female	O	4	1	23
	E	3.5628	1.6667	14.0104
White female	O	5	7	27
	E	5.9109	5.625	40.252
Chi-square values		.031	.2393	11.6514**
<u>degrees of freedom</u>		2	2	2
** p-value is less than .005 but greater than .001				

male supervisors have p-values greater than .75, and greater than .975, respectively, with two degrees of freedom. The results for white female supervisors has a p-value less than .025 but greater than .01 with two degrees of freedom. The positive nominations of white female supervisors had fewer black females and more white females than expected.

The chi-square values for negative nominations are .031, .2393, and 11.6514 for black female supervisors, white male supervisors, and white female supervisors, respectively. The results for black female supervisors and white male supervisors have p-values greater than .975 and greater than .75, respectively, with two degrees of freedom. The results for white female supervisors has a p-value less than .005 but greater than .001 with two degrees of freedom. The negative nominations of the white female supervisors had more males, more black females, and less white females than expected.

Results of the analysis using the subordinate variables race and sex are given in Table 3. Supervisor categories are not separated and black male subordinates are not pooled with the white male subordinates. The expected frequency (E) is the summation of probabilities of all the supervisors without regard to supervisor category of nominating subordinates in the different subordinate categories. The observed frequency (O) is the actual frequency of nominations of subordinates in the four subordinate race x sex categories.

Table 3. Analysis Using Subordinate Variable Race and Sex

<u>Positive Nominations</u>		
Subordinates		
Black males	O	3
	E	1.4431
White males	O	6
	E	6.9828
Black females	O	8
	E	19.8779
White females	O	64
	E	52.6966

p-value is less than .025 but greater than .01

Chi-square = 9.5349

degrees of freedom = 3

<u>Negative Nominations</u>		
Subordinates		
Black males	O	5
	E	1.4253
White females	O	8
	E	6.8966
Black females	O	28
	E	19.6322
White females	O	39
	E	52.046

p-value is less than .005 but greater than .001

Chi-square = 12.863

degrees of freedom = 3

The chi-square is 9.5349 for the positive nominations. This result has a p-value less than .025 but greater than .01 with three degrees of freedom. The chi-square is 12.863 for the negative nomination. This result has a p-value less than .005 but greater than .001 with three degrees of freedom.

The major part of the chi-square value in the positive nomination is accounted for by the deviation of reported nominations from expected nominations for the black female subordinates and the white female subordinates. Black female subordinates received fewer positive nominations than expected. White female subordinates received more positive nominations than expected.

The major part of the chi-square value in the negative nomination is accounted for by the deviation of reported nominations from expected nominations for the black male subordinates, the black female subordinates, and the white female subordinates. Black male subordinates received more negative nominations than expected. Black female subordinates received more negative nominations than expected. White female subordinates received fewer negative nominations than expected.

In Table 4, the subordinate variable of race is used and the supervisor categories are not separated. The expected frequency (E) is the summation of the probability for all supervisors of nominating black or white subordinates. The observed frequency (O) is the actual frequency of

Table 4. Analysis Using Subordinate Variable Race

<u>Positive Nominations</u>		
Subordinates		
Blacks	O	11
	E	21.321
Whites	O	70
	E	59.6794

p-value is less than .025 but greater than .01
 chi-square = 6.1398 degrees of freedom = 1

<u>Negative Nominations</u>		
Subordinates		
Blacks	O	33
	E	21.0575
Whites	O	47
	E	58.9426

p-value is less than .005 but greater than .001
 chi-square = 8.4392 degrees of freedom = 1

nominations reported for black or white subordinates.

The chi-square is 6.1398 for the positive nominations. This result has a p-value less than .025 but greater than .01 with one degree of freedom. The chi-square is 8.4392 for the negative nominations. This result has a p-value less than .005 but greater than .001 with one degree of freedom.

Black subordinates had fewer positive nominations than expected. White subordinates had more positive nominations than expected. Black subordinates had more negative nominations than expected. White subordinates had fewer negative nominations than expected.

In Table 5, the subordinate variable of sex is used and the supervisor categories are not separated. The expected frequency (E) is the summation of the probability for all supervisors of nominating a male or female subordinate. The observed frequency (O) is the actual frequency of nominations reported for male or female subordinates.

The chi-square is .0008 for the positive nominations. This result has a p-value greater than .9 with one degree of freedom. The chi-square is 2.3413 for the negative nominations. This result has a p-value greater than .1 but less than .25 with one degree of freedom.

White female supervisors were the only supervisor group in Table 6. The subordinate variable was race. The expected frequency (E) is the probability for white female supervisors of nominating black or white subordinates. The

Table 5. Analysis Using Subordinate Variable Sex

<u>Positive Nomination</u>		
Subordinates		
Male	0	9
	E	8.4259
Female	0	72
	E	72.5745

p-value is greater than .9

Chi-square = .0008 degrees of freedom = 1

<u>Negative Nomination</u>		
Subordinates		
Male	0	13
	E	8.3219
Female	0	67
	E	71.6782

p-value is less than .25 but greater than .1

chi-square = 2.3413 degrees of freedom = 1

Table 6. Analysis of White Female Supervisor
Nominations Using Subordinate Variable
Race

<u>Positive Nominations</u>		
<u>Subordinates</u>		
Black	0	6
	E	14.8555
White	0	54
	E	45.1445

p-value is less than .025 but greater than .01

chi-square = 6.2461 degrees of freedom = 1

<u>Negative Nominations</u>		
<u>Subordinates</u>		
Black	0	27
	E	14.8555
White	0	33
	E	45.1445

p-value is less than .001

chi-square = 12.1312 degrees of freedom = 1

observed frequency (O) is the actual frequency of nominations reported for black and white subordinates by white female supervisors.

The chi-square is 6.241 for the positive nominations. This result has a p-value less than .025 but greater than .01 with one degree of freedom. The chi-square is 12.1312 for the negative nominations. This result has a p-value less than .001 with one degree of freedom.

Black subordinates had fewer positive nominations than expected. White subordinates had more positive nominations than expected. Black subordinates had more negative nominations than expected. White subordinates had fewer negative nominations than expected.

Black female supervisors were the only supervisor group in Table 7. The subordinate variable was race. The expected frequency (E) is the probability for black female supervisors of nominating black or white subordinates. The observed frequency (O) is the actual frequency of nominations reported for black and white subordinates by black female supervisors.

The chi-square is .9922 for the positive nominations. This result has a p-value less than .5 but greater than .25 with one degree of freedom. The chi-square is .3832 for the negative nominations. This result has a p-value less than .75 but greater than .5 with one degree of freedom.

White male supervisors were the only supervisory group in Table 8. The subordinate variable was race. The expected

Table 7. Analysis of Black Female Supervisor
Nominations Using Subordinate Variable
Race

<u>Positive Nominations</u>		
<u>Subordinates</u>		
Black	0	2
	E	4.0971
White	0	9
	E	6.9028

p-value is less than .5 but greater than .25

chi-square = .9922 degrees of freedom = 1

<u>Negative Nominations</u>		
<u>Subordinates</u>		
Black	0	4
	E	3.7247
White	0	6
	E	6.275

p-value is less than .75 but greater than .5

chi-square = .3832 degrees of freedom = 1

Table 8. Analysis of White Male Supervisor
Nominations Using Subordinate Variable
Race

<u>Positive Nominations</u>		
<u>Subordinates</u>		
Black	0	3
	E	2.2223
White	0	7
	E	7.7778

p-value is less than .9 but greater than .75

chi-square = .0446 degrees of freedom = 1

<u>Negative Nominations</u>		
<u>Subordinates</u>		
Black	0	2
	E	2.2223
White	0	8
	E	7.7778

p-value is less than .9 but greater than .75

chi-square = .0446 degrees of freedom = 1

frequency (E) is the probability for white male supervisors of nominating black or white subordinates. The observed frequency (O) is the actual frequency of nominations reported for black and white subordinates by white male supervisors.

The chi-square is .0446 for the positive nominations. This result has a p-value less than .9 but greater than .75. The chi-square is .0446 for the negative nominations. This result has a p-value less than .9 but greater than .75 with one degree of freedom.

CHAPTER IV

CONCLUSIONS

In this study, the operational definition of bias was a deviation of the observed frequency from the expected frequency of reported nominations. The expected frequency assumes the assignments of nominations to be a chance event, independent of race or sex of subordinate.

The significance or non-significance in the results was reported using the alpha level of .05 which was set prior to data collection. There was a significant bias in the nomination of subordinates based on race of the subordinate. No significant bias was observed in the nomination of subordinates based on sex of the subordinate. The bias found in the nomination of subordinates seems to have an exclusive source in the white female supervisors.

Caution must be exercised in the interpretation of these results. The probability of a nomination being reported for a given subordinate cell is not the same for all supervisors. This violation has an unknown effect on the alpha region. For this reason interpretation of the significance of results must be guarded.

The noted bias yielded more negative nominations and less positive nominations than expected for the black

subordinates. More positive and less negative nominations than expected were reported for the white subordinates.

The direction of this bias seems to concur with the literature regarding ratings. In the relevant studies reviewed, considerable evidence was found showing that blacks received relatively lower performance ratings than whites (Arnold, 1968; Gavin and Ewen, 1974; Greenhaus and Gavin, 1972; Huck and Bray, 1976; Quinn, 1969; Schmidt and Hill, 1977).

This study found no evidence of a bias in nomination of subordinates based on sex of the subordinate to support other research where a sex effect on ratings has been demonstrated (Cline et al., 1977). It must be noted that the supervisors studied had a disproportionate number of female subordinates in contrast to most studies where male subordinates were more numerous.

From the results of this study it appears possible that these white female supervisors could be biased in their performance appraisals of their subordinates. This possibility, however, cannot be investigated from this study. Clearly previous studies have shown a bias, specifically a racial bias, in ratings. This study has attempted to show a potential source of such a bias. Whether this bias in nominations does affect the assigned evaluations for performance appraisal or criterion remains to be shown. It should be recalled that "... the criterion should give an

accurate and unbiased measure of the extent to which individuals in the validation population contribute to or detract from the efficiency of the organization" (Brogden and Taylor, 1950, pp. 160).

All the employees under the supervisors studies were required to meet basic Civil Service standards for initial employment. Further, minimum standards must be met for the various promotional stages. These considerations plus the lack of bias in white male supervisor nominations would argue against the existence of any "real" difference in performance of the two ethnic groups. However, since performance data was not made available for this research the alternative hypothesis that there was a performance difference between the two ethnic groups cannot be conclusively ruled out. The author feels obligated to remind his reader that the "bias" discussed here is a statistical "bias" and cannot be interpreted as prejudice. This is to say that attitudinal prejudice may or may not have influenced the results.

The implications of this research for criterion construction using the critical incident technique are somewhat nebulous. The intent of the technique is to collect effective and ineffective incidents describing job behaviors. This study did collect such incidents and has shown that more effective behavioral incidents or positive nominations are attributed to white subordinates and more ineffective

behavioral incidents or negative nominations are reported for black subordinates by white female supervisors. However, it cannot be determined with these data whether or not such perceptions would be biasing in actual performance appraisals. A performance appraisal instrument developed from the incidents provided by the nominations may not reflect the work behavior of the ethnic groups in an unbiased manner. The work behaviors described in the incidents might be shown by any worker regardless of race or sex. However, the effect of this bias influence could distort the generation of critical incidents in general.

The analysis of the nominations could be used as a preliminary identification device, where actual performance data cannot be obtained or would be extremely costly to obtain, to identify organizations or groups where potential racially discriminatory practices of performance evaluation exist. This possibility is proposed tentatively with the understanding that the performance component cannot be separated from the perceptual component of the positive or negative nomination. Further, it must be realized that extensive validation would be required before this identification device could be used.

Limitations of the Research

One limitation of this research is that the incidents were collected in a limited geographical region, the South-eastern part of the country. This fact raises serious question

as to the generalizability of the results to other geographical areas.

The research was also limited in the number of male subordinates concerning whom incidents could be reported. It would be desirable to replicate the study with the numbers of male and female personnel more nearly equal. This same limitation applies to the reporting supervisors. Here, the number of white female supervisors was preponderant.

Future Research

The most important extension of this study would be the inclusion of job performance data, which would determine whether the effect found is due to the perceptual component or to the performance component of the nomination. The perceptual component could be composed of elements such as attitudinal factors, opportunity for observation factors, etc. The performance component could be composed of elements such as work history and experience factors, training factors, etc.

It would be desirable to design a study to determine if the racial bias demonstrated in this research carries over to actual subordinate performance appraisals. Further, it would be desirable to explore the possibility of a bias in the dimensions or categories reported based on race, sex, etc. Research of this nature could lead to a similar examination of performance appraisals since the link between reporting of incidents and the use of appraisal instruments subsequently

developed must be determined.

The literature review showed that supervisors generally give lower performance ratings to blacks. Individual studies have shown that the race combination of rater-ratee biases the performance ratings, specifically, that raters give same race subordinates higher ratings than opposite race subordinates. However, no studies were found which reported such differences when a behavioral checklist based upon critical incidents was used. It seems appropriate to investigate this aspect of bias in the use of a behavioral checklist. If such a bias exists the effect of "balancing" the critical incidents should be investigated.

Another related research study would be a comparison of appraisals based upon more objective measures, a productivity count for example, with appraisals obtained with a behavioral check list and performance ratings. Distributions of performance scores, by race and/or sex, could be compared to assess whether or not bias in any or all of the appraisals exists as related to some more objective performance measure.

An interesting research possibility for the present organization would be to establish a training program in order to lessen the bias among the white female supervisors. Presumably the bias has some sort of attitudinal base and might be susceptible to correction with appropriate training.

Generally, it might be pointed out that this entire area, bias in performance appraisal, has received only

limited attention. Most emphasis has been placed upon selection of employees even though biased appraisals can be as harmful as biased selection. The personnel function of performance evaluation is in need of studies to determine the existence of and methods of correcting such bias.

APPENDIX A

Think of the best worker you supervise. Will you please write an incident describing something done in the job that made you think this. Remember, this is a story describing something that was done.

Please indicate employee's: Approx. Age _____, Education level (approx.) _____, Race _____, Sex (circle one) male or female, G.S. Rating _____, Branch _____, Measured or Unmeasured (circle one), Position status (circle one) Seasonal WAE or Intermittent or permanent.

Think of the worst worker you supervise. Will you please write an incident describing something done in the job that made you think this. Remember, this is a story describing something that was done.

Please indicate employee's: Approx. Age _____, Education level (approx.) _____, Race _____, Sex (circle one) male or female, G.S. rating _____, Branch _____, Measured or Unmeasured (circle one), Position status (circle one) Seasonal WAE or Intermittent or Permanent.

Please indicate the number of workers you supervise in each of the following demographic categories. Please give your best estimate of this information.

Black males _____, White males _____,

Black females _____, White females _____.

Under 25 _____, 26-30 _____,

31-40 _____, 41-50 _____,

51 and over _____.

G.S. ratings: 1 _____, 2 _____, 3 _____,
4 _____, 5 _____, 6 and over _____.

Is your best worker also your most dependable? (circle one)

YES or NO

Is your worst worker also your least dependable? (circle one)

YES or NO

APPENDIX B

INSTRUCTIONS GIVEN SUPERVISORS

I would like to thank all of you for being here and say that we really appreciate your cooperation. If you have any questions while filling out these forms please feel free to bring them to me. When you are finished I would like to briefly check them before you leave.

We are asking you to give us 'incidents' concerning the work behavior of your subordinates. Incidents are 'stories' concerning your better and poorer workers. For example, in a recent study, college students were asked to think of the 'best professor' they had during the present school year. Many of them indicated that the professor was 'interesting'. People can be interesting in many ways so the students were directed to describe what the professor did that made him interesting. An example of one of these incidents is: 'Professor X brought in real life examples to illustrate his lectures, like when he was discussing harmonic distortion, he brought in his guitar and gave us an example of harmonic distortion.'

This is what we are asking you for, a specific behavioral incident, something that was actually done. So think of the best employee that you supervise and write an incident describing something done by this employee that made you think of this subordinate as your best. You will

also be given forms of dependability which we would like you to fill out the same way. So think of your most dependable worker and describe something the person did that made you think that this person was your most dependable subordinate.

I will give you four sheets. At the top of the first it says 'think of the best worker you supervise'. Write an incident about your best employee. The next sheet says 'think of the worst worker you supervise'. Write an incident about your worst employee. The other sheets are about your most dependable and least dependable workers. REMEMBER what we need is a story about something that was done.

At the bottom of each sheet of paper is a short form concerning demographic data which are of interest to our research. Please fill in this information about the subordinate being described. There is also a final form concerning demographic data of all your workers, please fill it out also. Be sure to include all your employees, permanent and temporary, that are presently working. All parties involved in this research shall remain anonymous.

CATEGORIES OF POSITIVE CRITICAL
INCIDENTS OF JOB PERFORMANCE

I. Unit oriented behavior

A. Will do other work not specifically assigned to them.
frequency = 11; frequency of incidents on blacks =

2

B. Willing to do any type of job (not picky about assignment). frequency n = 7; frequency of incidents on blacks = 0

C. Accepted or selected for special assignment (completed accurately and efficiently).
frequency = 12; frequency of incidents on blacks = 1

II. Work oriented behavior

A. Willing to work overtime (or simply works overtime).
frequency = 5; frequency of incidents on blacks = 1

B. Schedules leave in advance and around work schedule.
frequency = 6; frequency of incidents on blacks = 1

C. Never tardy
frequency = 4; frequency of incidents on blacks = 0

D. Never absent.
frequency = 2; frequency of incidents on blacks = 0

E. Makes personal sacrifices for work.
frequency = 7; frequency of incidents on blacks = 0

F. Volunteers for special assignments (or does).
frequency = 15; frequency of incidents on blacks = 3.

- III. Has learned, or attempted to learn, all job skills.
- A. Studies or does extra work to improve oneself (own training).
frequency = 11; frequency for blacks = 2
- B. Competence - exhibits greater ability to do job than most.
frequency = 11; frequency of blacks = 2
- IV. Production (Quantity and Quality).
- A. Quantity of work produced.
frequency = 19; frequency of blacks = 1
- B. Quality of work produced.
frequency = 20; frequency of blacks = 1
- V. Interpersonal relationships (with peers and superior).
- A. Helps other employees.
frequency = 21; frequency of blacks = 3
- B. Always keeps composure and retains authority.
frequency = 2; frequency of blacks = 0
- C. Accepted or took upon themselves a leadership role or extra responsibilities, effectively accomplished task.
frequency = 4; frequency of blacks = 0
- D. Acts as a communication link for the supervisor; acts as a confidant or to give feedback on work performance.
frequency = 4; frequency of blacks = 0.
- E. Diplomatic and tactful in relations with other people.
frequency = 1; frequency of blacks = 0.

IV. Concern for job improvements and work flow.

A. Improves upon assignment - does more than required;
tries to do things in a "better" way.

frequency = 11; frequency of blacks = 1

B. Does things to keep work unit producing and timely.

frequency = 3; frequency of blacks = 1

C. Willingness to try new methods and to expend the
effort to make them work.

frequency = 2; frequency of blacks = 0

frequency = total frequency including blacks.

frequency of blacks = includes only the blacks in the category.

CATEGORIES OF NEGATIVE CRITICAL
INCIDENTS OF JOB PERFORMANCE

I. Attendance at work.

- A. Excessive tardiness or late coming back from breaks.
frequency = 12; frequency of blacks = 10
- B. Excessive absenteeism or does not report in leave.
frequency = 13; frequency of blacks = 8
- C. Physician or other excuse (believed fraudulent).
frequency = 3; frequency of blacks = 0
- D. Abuses or uses leave as it accumulates.
frequency = 6; frequency of blacks = 3

II. Work Performance

- A. Low quantity of work (doesn't do fair share).
frequency = 18; frequency of blacks = 6
- B. Does bare minimum or looks for easiest work.
frequency = 10; frequency of blacks = 6
- C. Low quality of work.
frequency = 24; frequency of blacks = 10
- D. Proceeds without asking questions (when should have asked!)
frequency = 1; frequency of blacks = 0
- E. Excuses poor work performance.
frequency = 10; frequency of blacks = 2

F. Learning time excessive (repeats mistakes).

frequency = 5; frequency of blacks = 3

III. Relations with Supervisors

A. Insubordinate.

frequency = 3; frequency of blacks = 1

B. Defies supervisor.

frequency = 2; frequency of blacks = 1

C. Fails to cooperate (help in emergencies, etc.)

frequency = 4; frequency of blacks = 1

D. Fails to follow specific job instructions.

frequency = 3; frequency of blacks = 1

E. Makes unjustified complaints or grievances.

frequency = 6; frequency of blacks = 2

IV. Relations with other persons

A. Impolite to public

frequency = 1; frequency of blacks = 0

B. Irritates, argues, disturbs, etc. peers.

frequency = 7; frequency of blacks = 4

C. Will not cooperate with peers.

frequency = 1; frequency of blacks = 1

V. General Attitude toward work.

A. Will not follow work rules.

frequency = 20; frequency of blacks = 10

B. Abuses telephone, visiting, etc.

frequency = 7; frequency of blacks = 4

- C. Careless with important and confidential material.
frequency = 1; frequency of blacks = 1
- D. Wastes time on irrelevant activities.
frequency = 13; frequency of blacks = 8
- E. Will not accept responsibility (have to watch them).
frequency = 8; frequency of blacks = 2
- F. Sleeps on job.
frequency = 6; frequency of blacks = 3
- G. Complains, grumbles, criticizes.
frequency = 11; frequency of blacks = 6
- H. Fail to report for scheduled overtime.
frequency = 1; frequency of blacks = 1

APPENDIX D

I.D. No.	Supervisor Race & Sex	Subordinate Categories			
		Black Male	White Male	Black Female	White Female
1	WF	0	1	3	10
2	WF	0	1	1	12
3	WM	**	**	**	**
4	WF	0	2	6	19
5	WF	0	1	18	21
6	BF	1	1	8	13
7	WF	0	1	4	17
8	WM	**	**	**	**
9	WM	0	3	3	4
10	WF	0	2	16	20
11	WF	0	2	8	26
12	WF	0	0	3	17
13	WF	0	0	1	3
14	WF	0	0	3	15
15	BF	0	0	10	11
16	WF	0	3	2	17
17	BF	0	0	14	16
18	WM	1	4	10	17
19	WM	2	5	0	4
20	WF	0	1	5	26
21	WF	**	**	**	**
22	WF	0	1	5	19
23	WF	0	0	10	11
24	BF	1	2	3	3
25	WF	0	2	7	15
26	WF	0	0	3	7
27	WF	0	2	4	10
28	WF	0	2	5	13
29	WF	0	1	2	25
30	WF	0	2	2	10

Supervisor I.D. No.	Race & Sex	Subordinate Categories			
		Black Male	White Male	Black Female	White Female
31	WF	0	2	1	16
32	BF	0	1	19	11
33	WF	0	1	11	15
34	WF	2	12	30	66
35	WF	0	0	2	14
36	WF	0	2	5	4
37	BF	0	0	1	13
38	WF	0	1	1	12
39	WF	1	3	2	19
40	WM	1	4	3	4
41	BF*	0	1	4	21
42	WF	2	1	15	15
43	WF	1	2	7	22
44	WF	0	0	1	7
45	WF	0	2	2	9
46	WM	2	4	0	0
47	BF	0	0	6	16
48	WM	2	8	2	4
49	WM	0	1	2	7
50	WF	0	1	1	30
51	WF	0	1	3	18
52	WF	0	0	4	16
53	WF	0	1	4	7
54	WF	**	**	**	**
55	WF	0	1	4	14
56	WF	0	2	4	23
57	WF	1	3	6	12
58	WF	1	11	0	2
59	WF	1	2	3	23
60	WF	**	**	**	**
61	WM	0	0	1	7

Supervisor I.D. No.	Race & Sex	Subordinate Categories			
		Black Male	White Male	Black Female	White Female
62	WF	2	2	0	2
63	WF	1	9	10	6
64	WF	2	4	12	13
65	BF	1	3	2	19
66	WF	0	0	5	22
67	WM	0	2	2	22
68	WF	1	0	3	16
69	WM	0	0	1	12
70	WF	0	1	2	12
71	WF	0	0	5	10
72	WF	**	**	**	**
73	BF	1	1	7	13
74	WF	2	5	10	16
75	WF	**	**	**	**
76	WF	2	3	7	10
77	WF	0	0	3	21
78	BF	0	0	14	10
79	WF	0	0	1	2
80	WF	0	0	7	11
81	WF	0	0	3	12
82	WF	0	3	11	28
83	WF	0	2	5	12
84	WF	0	2	8	7
85	WF	0	2	3	21
86	WF	**	**	**	**
87	WF	0	1	1	16
88	WF	0	2	2	4
89	WF	0	1	3	7

* is the black female supervisor who did not report a negative nomination.

** are supervisors who were dropped.

WF - is a White female supervisor

BF - is a Black female supervisor

WM - is a White male supervisor

APPENDIX E

Dr. S. A. Mulaik has shown in a Monte Carlo simulation that the distribution generated in an experiment from a table similar to Table 3 in the text is approximately distributed as Chi-square with $k - 1$ degrees of freedom. It was found that the empirical probability that a value greater than 7.82 would be obtained by a Chi-square distribution was .05. The probability of such a result in the simulated distribution was .027. A value greater than 11.34 has a probability level of .01 in a Chi-square distribution. The same result has a probability of .005 in the simulated distribution. That a value exceeding 12.84 would be obtained has a probability of .005 in the Chi-square distribution and a probability of .003 in the simulated distribution. Therefore, in this instance assuming a Chi-square distribution with $k - 1$ degrees of freedom has provided a conservative test of significance.

REFERENCES

- Anastasi, A. An investigation of sources of bias in the prediction of job performance: A six year study. Princeton, New Jersey: Educational Testing Service, 1972.
- Andersson, B. E. & Nilsson, S. G. Studies in the reliability and validity of the critical incident technique. Journal of Applied Psychology, 1964, 48, 398-403.
- Arnold, B. C. Comparison of Caucasian and Negroe subgroups on criterion indices of overall job effectiveness. Unpublished doctoral dissertation, Colorado State University, 1968.
- Avery, R. D., Passino, E. M., & Lounsbury, J. W. Job analysis results as influenced by sex of incumbent and sex of analyst. Journal of Applied Psychology, 1977, 62, 411-416.
- Bass, A. R. & Turner, J. N. Ethnic group differences in relationships among criteria of job performance. Journal of Applied Psychology, 1973, 57, 101-109.
- Bigoness, W. J. Effect of applicant's sex, race, and performance on employers' performance ratings: Some attitudinal findings. Journal of Applied Psychology, 1976, 61, 80-84.
- Borman, W. C. & Dunnette, M. D. Behavior-based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 1975, 60, 561-565.
- Brogden, H. E. & Taylor, E. K. The theory and classification of criterion bias. Educational and Psychological Measurement, 1950, 10, 159-186.
- Campbell, J. P., Dunnette, M. D., Avery, R. D., & Hellervick, L. V. The development and evaluation of behaviorally based rating scales, Journal of Applied Psychology, 1973, 57, 15-22.
- Campbell, J. T. An investigation of sources of bias in the prediction of job performance: A six year study. Princeton, New Jersey: Educational Testing Service, 1972.

- Campion, J. E., Greener, J., & Wernli, S. Work observation versus recall in developing behavioral examples for rating scales. Journal of Applied Psychology, 1973, 58, 286-288.
- Cascio, W. F. & Valenzi, E. R. Relations among criteria of police performance. Journal of Applied Psychology, 1958, 63, 22-28.
- Cecil, E. A., Paul, R. J., & Olins, R. A. Perceived importance of selected variables used to evaluate male and female job applicants. Personnel Psychology, 1973, 26, 397-404.
- Cline, M. E., Holmes, D. S. & Werner, J. C. Evaluation of the work of men and women as a function of the sex of the judge and type of work. Journal of Applied Social Psychology, 1977, 7, 89-93.
- DeJung, J. E. & Kaplan, H. Some differential effects of race of rater and ratee on early peer ratings of combat aptitude. Journal of Applied Psychology, 1961, 46, 370-377.
- Einhorn, H. J. & Bass, A. R. Methodological considerations relevant to discrimination in employment testing. Psychological Bulletin, 1971, 75, 261-269.
- Flanagan, J. C. Defining the requirements of the executive's job. Personnel, 1951, 28-35.
- Flanagan, J. C. The critical incident technique. Psychological Bulletin, 1954, 51, 327-358.
- Flanagan, J. C. Measuring human performance, Pittsburgh, Pennsylvania; American Institute for Research, 1974.
- Gavin, J. F. & Ewen, R. B. Racial differences in job attitudes and Performance; Some theoretical considerations and empirical findings. Personnel Psychology, 1974, 27, 455-464.
- Gordon, M. E. The effect of the correctness of the behavior observed on the accuracy of ratings. Organizational Behavior and Human Performance, 1970, 5, 366-377.
- Greenhaus, J. H. & Gavin, J. F. The relationship between expectancies and job behavior for white and black employees. Personnel Psychology, 1972, 25, 449-455.

- Guion, R. M. Employment tests and discriminatory hiring. Industrial Relations, 1966, 5, 20-37.
- Haefner, J. E. Sources of discrimination among employees: A survey investigation, Journal of Applied Psychology, 1977, 62, 265-270.
- Hamm, N. H., Baum, M. R., & Nikels, K. W. Effects of race and exposure on judgments of interpersonal favorability. Journal of Experimental Social Psychology, 1975, 11, 14-24.
- Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. Journal of Applied Psychology, 1974, 59, 705-711.
- Huck, J. R. & Bray, D. W. Management assessment center evaluations and subsequent job performance of white and black females. Personnel Psychology, 1976, 29, 13-30.
- Inn, A., Hulkin, C. L., & Tucker, L. Three sources of criterion variance: Static dimensionality, dynamic dimensionality, and individual dimensionality, Organizational Behavior and Human Performance, 1972, 8, 58-83.
- Kaufman, G. G. & Johnson, J. C. Scaling peer ratings: An examination of the differential validities of positive and negative nominations, Journal of Applied Psychology, 1974, 59, 302-306.
- Kipnis, D., Silverman, A., & Copeland, C. Effects of emotional arousal on the use of supervised coercion with black and union employees. Journal of Applied Psychology, 1973, 57, 38-43.
- Klimoski, R. J. & London, M. Role of the rater in performance appraisal. Journal of Applied Psychology, 1974, 59, 445-451.
- Klores, M. S. Rater bias in forced-distribution performance ratings, Personnel Psychology, 1966, 19, 411-421.
- Kolstoe, R. H. Introduction to statistics for the behavioral sciences. Homewood, Illinois: The Dorsey Press, 1973.
- Ledvinka, J. Race of employment interviewer and reasons given by black job seekers for leaving their jobs. Journal of Applied Psychology, 1973, 58, 362-364.

- McGuire, F. L. The nature of bias in official accident and violation records. Journal of Applied Psychology, 1973, 57, 300-305.
- Norton, S. D., Gustafson, D. P., & Foster, C. E. Assessment for management potential: Scale design and development training effects, Academy of Management Journal, 1977, 20, 117-131.
- O'Reilly, A. P. The supervisor and his subordinate's self-actualization, Personnel Psychology, 1973, 26, 81-85.
- Parker, W. S. Black-white differences in leader behavior related to subordinates' reactions. Journal of Applied Psychology, 1976, 61, 140-147.
- Porter, D. T. An experimental investigation of the effects of racial prejudice and racial perception upon communication effectiveness. Speech Monographs, 1974, 41, 177-184.
- Quinn, J. L. Bias in performance appraisal. Personnel Administrator, 1969, 32, 40-43.
- Richards, S. A. & Jaffee, C. L. Blacks supervising whites: a Study of interracial difficulties in working together in a simulated organization. Journal of Applied Psychology, 1972, 56, 234-240.
- Ronan, W. W. Training for emergency procedures in multi-engine aircraft. Pittsburgh, Pennsylvania: American Institute for Research, 1953.
- Ronan, W. W. & Latham, G. P. The reliability and validity of the critical incident technique: A closer look. Studies in Personnel Psychology, 1974, 6, 53-64.
- Ronan, W. W. & Prien, E. P. Perspectives on the measurement of human performance, New York: Appleton-Century-Crofts, 1971.
- Ronan, W. W. & Schwartz, A. P. Ratings as performance criteria. International Review of Applied Psychology, 1974, 23, 71-82.
- Rosen, B. & Jerdee, T. H. Influence of sex role stereotypes on personnel decisions. Journal of Applied Psychology, 1974a, 59, 9-14.

- Rosen, B. & Jerdee, T. H. Effects of applicant's sex and difficulty of job on evaluations of candidates for managerial positions. Journal of Applied Psychology, 1974b, 59, 506-512.
- Schmidt, F. L. & Johnson, R. H. Effect of race on peer ratings in an industrial situation. Journal of Applied Psychology, 1973, 57, 237-241.
- Schmitt, N. & Hill, T. E. Sex and race composition of assessment center groups as a determinant of peer and assessor ratings. Journal of Applied Psychology, 1977, 62, 261-264.
- Scott, W. E. & Hamner, W. C. The influence of variations in performance profiles on the performance evaluation process: An Examination of the validity of the criterion. Organizational Behavior & Human Performance, 1975, 14, 360-370.
- Sewell, A. F. Person perception as a function of the personal consequences and immediacy of a decision. The Journal of Psychology, 1973, 85, 157-164.
- Walker, J. W. Personal communication, May 15, 1978.
- Wiener, Y. & Schneiderman, M. L. Use of job information as a criterion in employment decisions of interviewers. Journal of Applied Psychology, 1974, 59, 699-704.