



12-1980

Differences in Assessment Center Performance as a Function of the Race and Sex of Ratees and the Race of Assessors

Mark Joel Friedman
University of Tennessee

Follow this and additional works at: https://trace.tennessee.edu/utk_graddiss

Recommended Citation

Friedman, Mark Joel, "Differences in Assessment Center Performance as a Function of the Race and Sex of Ratees and the Race of Assessors. " PhD diss., University of Tennessee, 1980.
https://trace.tennessee.edu/utk_graddiss/6115

This Dissertation is brought to you for free and open access by the Graduate School at TRACE: Tennessee Research and Creative Exchange. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of TRACE: Tennessee Research and Creative Exchange. For more information, please contact trace@utk.edu.

To the Graduate Council:

I am submitting herewith a dissertation written by Mark Joel Friedman entitled "Differences in Assessment Center Performance as a Function of the Race and Sex of Ratees and the Race of Assessors." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

Ohmer Milton, Major Professor

We have read this dissertation and recommend its acceptance:

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

(Original signatures are on file with official student records.)

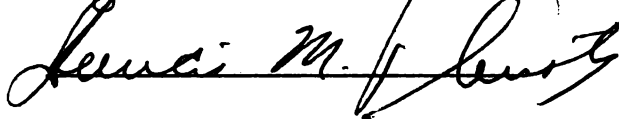
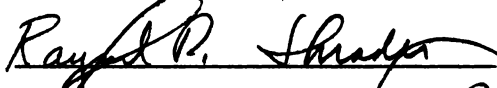
To the Graduate Council:

I am submitting herewith a dissertation written by Mark Joel Friedman entitled "Differences in Assessment Center Performance as a Function of the Race and Sex of Ratees and the Race of Assessors." I recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Psychology.

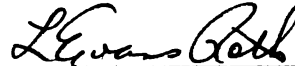


Ohmer Milton, Major Professor

We have read this dissertation
and recommend its acceptance:



Accepted for the Council:



Vice Chancellor
Graduate Studies and Research

DIFFERENCES IN ASSESSMENT CENTER PERFORMANCE AS A
FUNCTION OF THE RACE AND SEX OF RATEES AND THE
RACE OF ASSESSORS

A Dissertation
Presented for the
Doctor of Philosophy
Degree
The University of Tennessee, Knoxville

Mark Joel Friedman

December 1980

ACKNOWLEDGMENTS

It has probably been said many times before that writing a dissertation is demanding. For me, the experience has been rewarding, although at times painful. There are many people who helped to reduce the pain. They gave me their love, support, encouragement, and friendship.

I would like to extend my sincere appreciation to each of the members of my committee, Dr. Kenneth Newton, Dr. Raymond Shrader, and Dr. Francis Trusty. Each made his own unique and helpful contribution to this work. As a committee, they functioned in the true spirit of academe--as enablers, rather than disablers of the educational process.

I am especially grateful to Dr. Ohmer Milton. As a teacher, a mentor, a friend, and as the chairman of the committee, he helped to guide me over the rough spots. He will always have my deepest respect and admiration.

I would also like to thank Gwen Sweeden, a friend, who gave me encouragement and support throughout this project.

I owe an especially large amount of gratitude to my supervisors at the Tennessee Valley Authority--Arthur H. Allen, Marilyn F. Taylor, and Sue E. Wallace for their assistance and patience. A special note of thanks is owed to Evelyn Carr for typing this manuscript.

I would also like to thank Cabot L. Jaffee, of Assessment Designs, for his personal interest in me and this project, and for allowing me the use of the assessment materials that formed the basis of this study.

And last, but certainly not least, I would like to thank my parents, William and Dorothy Friedman, who gave me their love and understanding, and taught me the value of perserverance.

ABSTRACT

The purpose of this study was to investigate the possibility of any systematic differences in assessment center ratings as a function of the race and sex of the ratees or the race of the assessors. In effect, the overriding question was to determine if there were discrimination due to race or sex-related biases.

The subjects were 256 employees of a large southeastern utility who participated in an assessment center as the initial step for selection into a two-year management development and training program. Participants spent one day in the center, which consisted of three simulations of typical management activities. The exercises were designed to measure the following nine skill dimensions identified by a job analysis as being critical for job success: (1) Leadership, (2) Perception, (3) Adaptability, (4) Decisiveness (refers to the number of decisions made), (5) Decision-making (refers to the quality of decisions), (6) Organization and Planning, (7) Sensitivity, (8) Written Communications, and (9) Oral Communications.

The data were analyzed using a Three-way Analysis of Variance design, and results indicated no race or sex-linked discrimination. Females scored significantly higher than males on four of the dimensions, while blacks scored significantly lower than whites on all nine of the skill dimensions.

Because the technique appears to be free of sex or race-linked biases, the data suggest that the assessment center technique can be

useful in promoting equal employment opportunities. Several steps are recommended which organizations could take to further prepare blacks for managerial responsibility. Directions for future research are also discussed.

TABLE OF CONTENTS

CHAPTER	PAGE
I. INTRODUCTION	1
What is an Assessment Center?	2
Varied Uses of Assessment Centers.	4
Length and Size of a Center.	5
Evaluative Studies on Assessment Centers--	
Reliability and Validity	5
Studies on Sex and Race Bias	9
Research on Sex-Role Stereotyping.	10
Research on Racial Bias.	13
The Present Study.	21
II. METHOD	23
Subjects	23
Instrumentation.	25
Scoring and Dependent Variables.	27
Procedure.	29
III. RESULTS.	32
Analysis of Overall Assessment Center Scores	32
Analysis of Mean Performance Scores on Each	
Assessment Exercise.	33
Analysis of Mean Performance Scores Across	
Assessment Exercises	34
IV. DISCUSSION	40
Directions for Future Research	47

CHAPTER	PAGE
V. SUMMARY	50
REFERENCES	53
APPENDIXES	59
Appendix A.	60
Appendix B.	62
Appendix C.	67
Appendix D.	72
VITA.	73

LIST OF TABLES

TABLE	PAGE
1. Demographic Breakdown of <u>Ss</u>	23
2. Analysis of Variance Comparing Overall Assessment Score Across Participant Race, Participant Sex, and Racial Composition of Assessor Teams	32
3. Analysis of Variance Comparing Mean Performance Scores on Each Exercise Across Participant Race, Participant Sex, and Race of the Assessor	34
4. Analysis of Variance With Repeated Measures on Assess- ment Exercise Comparing Mean Performance Scores on Each Exercise Across Participant Race, Participant Sex, and Assessment Exercise	35
5. Results of Repeated Measures ANOVA on Each Assessment Dimension	37
6. Mean Performance Scores by Race on Each Skill Dimension	38
7. Mean Performance Scores by Sex on Each Dimension	38

CHAPTER I

INTRODUCTION

It is increasingly clear that today, perhaps as never before, there is a need for competent managers, administrators, and executives in both the public and private sectors. Because of the growth and complexities of our technology and economy, it is essential that we have a supply of well trained individuals who can manage material, as well as human resources (Moses, 1973). It is also clear that more and more managerial jobs have been going unfilled. This has been caused by rapid industrial growth, lowered birth rates during the 1930's, and until very recently, a disinterest on the part of college graduates to enter the business world. Thus, over the last 25 years there has been a growing concern about identifying and developing top managerial talent (Dunnette, 1971; Wollowick & McNamara, 1969).

Numerous methods of dealing with managerial shortages have been attempted--including elaborate selection systems, computerized manpower and skill inventories, management training and development programs, and the use of executive search firms, or "head hunters." Another approach of the last 25 years has been that of multiple assessment procedures. This approach, generally known as the Assessment Center Method, has grown in popularity because it is a formal and systematic technique characterized by careful planning, standardized procedures, and an impressive amount of positive evaluative research (see reviews by Dunnette, 1971; Finkle, 1976; Huck, 1977). In addition, traditional testing procedures utilizing paper-and-pencil

instruments, have come under heavy legal attack and increasing government scrutiny (see Huck & Bray, 1976; Uniform Guidelines, 1978) because of a lack of sound validity data and the possibility that these instruments may be biased unfairly against minorities (see Ash & Kroeker, 1975).

What Is An Assessment Center?

An assessment center is not a place--rather it is a technique--a technique for simulating and sampling job activities that have been identified through job analysis as being critical for successful job performance. In a sense, the assessment center method provides a means for organizations to allow candidates to "try out" for a particular position. As such, it has been used as a selection method for such diverse positions as top executives, managers, salespeople, police and firemen, highly skilled blue-collar workers, and stockbrokers (Dunnette & Borman, 1979). Over 4,000 organizations, including private industry, government and other non-profit agencies, have widely used the technique and over 300,000 people are estimated to have participated in assessment centers around the world (Development Dimensions International, 1977). Depending upon the purpose of the center and the organization involved, assessment activities may consist of paper-and-pencil personality tests, IQ and attitude inventories, personal interviews and job simulation exercises such as leaderless group discussions, in-basket exercises, and business simulation games (see also Arvey, 1979; Bray, 1976; Jaffee & Sefcik, 1980).

Finkle (1976, p. 862) points out that the key factor distinguishing assessment centers from other selection devices is "assessment in groups, assessment by groups, use of multiple measurement techniques

with a heavy emphasis on situational exercises, and a special appeal to management."

The proceedings of the Third International Congress on the Assessment Center Method, Standards and Ethical Considerations for Assessment Center Operations (Moses, 1975, p. 2-3), explicitly defines what does and does not constitute an assessment center. This document outlines seven minimum requirements which must be met before a selection procedure can be called an assessment center:

1. Multiple assessment techniques must be used. At least one of these techniques must be a simulation.

A simulation is an exercise or technique designed to elicit behaviors related to dimensions of performance on the job by requiring the participant to respond behaviorally to situational stimuli. The stimuli present in a simulation parallel or resemble stimuli in the work situation. Examples of simulations include group exercises, in-basket exercises, and fact-finding exercises.

2. Multiple assessors must be used. These assessors must receive training prior to participating in a center.
3. Judgments resulting in an outcome (i.e., recommendation for promotion, specific training or development) must be based on pooling information from assessors and techniques.
4. An overall evaluation of behavior must be made by the assessors at a separate time from observation of behavior.
5. Simulation exercises are used. These exercises are developed to tap a variety of predetermined behaviors and have been pretested prior to use to insure that the techniques provide reliable, objective, and relevant behavioral information to the organization in question.
6. The dimensions, attributes, characteristics, or qualities evaluated by the assessment center are determined by an analysis of relevant job behaviors.
7. The techniques used in the assessment center are designed to provide information which is used in evaluating the dimensions, attributes, or qualities previously determined.

Those activities that do not constitute an assessment center include:

1. Panel interviews or a series of sequential interviews as the sole technique.
2. Reliance on a specific technique (regardless of whether a simulation or not) as the sole basis for evaluation.
3. Using only a test battery composed of a number of paper and pencil measures, regardless of whether the judgments are made by a statistical or judgmental pooling of scores.
4. Single assessor assessment - . . . measurement by one individual using a variety of techniques such as pencil and paper tests, interviews, personality measures, or simulations.
5. The use of several simulations with more than one assessor where there is no pooling of data, i.e., each assessor prepares a report on performance in an exercise, and the individual reports (unintegrated) are used as the final product of the center.
6. A physical location labeled as an "assessment center" which does not conform to the requirements noted above.

Varied Uses of Assessment Centers

The data collected at a center may have several uses in addition to that of selection. For example, in a recent study of how organizations utilize assessment center results, Alexander (1979) indicates that the centers can be used to (1) make specific recommendations for individual development based on strengths and weaknesses identified in the center, (2) develop managerial replacement and succession plans, (3) identify high potential individuals in the organization, (4) determine overall strengths and weaknesses of individuals, as well as identifying areas within the organization that have specific skill development needs, (5) aid in career planning, and (6) serve as a training tool for assessors, who are generally higher level managers. Assessor training helps managers to learn how to observe

and record behavior; with these new skills, they are better equipped to coach and counsel subordinates, as well as to conduct more effective performance appraisals. One additional use of assessment centers (Blumenfield, 1971; Cohen & Bunker, 1975; Huck & Bray, 1976) is the identification of minority group members who have potential to assume managerial positions. Thus, assessment centers can aid an organization in the implementation of its affirmative action program (Moses & Boehm, 1975).

Length and Size of a Center

Assessment centers vary in length depending upon the number and types of exercises, the purpose of the center and the level of responsibility for which it is being used. Generally centers designed for training and development require more time than those designed primarily for selection and promotion. This is because in a center for development, participants receive immediate feedback and take part in training activities. These centers may require up to a week. Centers designed for selection of first-line supervisors, salespersons, or blue collar workers may last for a day or less. Centers designed to select top executives very often consist of two to two and a half days of exercises and interviews (Development Dimensions, 1977). The general procedure calls for groups of either six or twelve individuals to be assessed simultaneously, with one assessor for every two participants; however, some centers use an assessor-participant ratio of one-to-one.

Evaluative Studies on Assessment Centers--Reliability and Validity

In the 25 or so years since AT&T first initiated the use of assessment centers in industry (Bray, 1964; Bray, Campbell, & Grant,

1974) much research has been conducted on both the reliability and validity of the method. In his review of the literature, Huck (1977) points out that interrater reliabilities reported across studies are consistently high (see also, Greenwood & McNamara, 1967; Schmitt, 1977).

The majority of studies show similar results for predictive validity (see reviews by Dunnette, 1971; Finkle, 1976; Huck, 1973, 1977). Byham (1970) concludes that the research evidence accumulated across various organizations and studies lends considerable credibility to the overall validity of the method. He writes:

In a survey of the 20 companies that operated centers, I uncovered some 22 studies in all that showed assessment more effective than other approaches and only one that showed it exactly as effective as some other approaches. None showed it less effective. As I suggested before, these studies exhibit correlations between center prediction and achievement criteria such as advancement, salary grade, and performance ratings that range as high as .64 (p. 154).

Regardless of the format of the center, the criteria used, the methodology or the type of job being assessed, results have been positive and consistent (Klimoski & Strickland, 1977).

Studies have also shown that assessment centers are equally as valid for predicting the performance of minorities and females as they are for predicting performance of white males (Boche, 1977; Huck & Bray, 1976; Moses, 1973; Moses & Boehm, 1975). Moses & Boehm (1975) compared assessment center ratings of over 8,000 males and 4,500 females who had been assessed sometime between 1963 and 1971. The distribution of ratings for men and women were quite similar evidencing a rank order correlation of .75 ($p < .01$). It was found

that the four assessment center dimensions (overall rating, leadership, decision-making, and organization and planning) which correlated most highly with the criterion of level of management for men, also had the highest correlations for women. The overall assessment ratings for women correlated .37 with management progress; for men the corresponding correlation was .44. Moses and Boehm (1975) concluded that overall, nearly identical proportions of males and females do well in the assessment center. Thus, the method appears to be an excellent way to avoid adverse impact when making selection decisions for management positions.

Using two samples, Huck & Bray (1976) compared assessment center performance for white and black females. The primary sample consisted of 126 nonmanagement women (91 white and 35 black) who had been promoted to one of two supervisory positions. The secondary sample consisted of 479 women who were not promoted (238 black and 241 white) and who had attended the center during the same period, 1966-1971, as had the women in the primary sample. The supplementary sample was used to facilitate additional internal analyses. All participants (in both samples) were rated on several assessment center variables and given an overall assessment rating. All the women in the primary sample, (who had been promoted) were given supervisory ratings on several criterion variables, including an overall job performance scale and a potential for advancement scale.

Several analyses were conducted on these data, including a comparison of the factor structure for both samples compared by race. The analysis indicated similar factor structures for both samples, for both races. Another analysis (conducted on the primary sample)

involved the relationship between the overall assessment rating and the two criterion variables--overall job performance and potential for advancement. Validity coefficients between overall assessment ratings and overall job performance were .41 for the white group and .38 for the black group; both were significant. The correlations between the overall assessment rating and potential for advancement were somewhat higher, .59 for whites and .54 for blacks; again, both were significant. Because the correlations in both of these analyses were slightly lower for the black group, the two regression lines were compared. The lines did not differ, indicating that a common regression line could be applied to both groups.

The results of this study indicated that the assessment center was equally valid for both black and white women. It is also interesting to note that non-promoted black women received significantly lower overall assessment ratings than the non-promoted white women--the mean difference being 0.40 (2.40 vs. 2.80). A difference of the same magnitude was noted in the primary sample, black women scoring an average of 3.00 and whites scoring 3.40 on the overall assessment rating. This difference did not reach statistical significance due to the smaller sample size of the promoted group (Huck & Bray, 1976). The data also indicated that white assessees in both the primary and supplementary samples were rated significantly higher on several variables. Black women were also rated lower on one of the criterion measures--supervisory ratings on overall job performance. Because the black women did somewhat less well than the white group in their assessment center performance and they also received somewhat lower criterion ratings, the correlations for both groups

between assessment and criterion variables were not significantly different. Thus, the authors concluded that assessment centers are valid for different ethnic groupings.

Studies on Sex and Race Bias

The literature cited above would lead one to conclude that the assessment center method is indeed a valid technique for predicting future job performance of white males, as well as for minorities and females. However, a recent article by Klimoski & Strickland (1977) has called into question the evidence of criterion-related validity in these studies. They suggest (p. 354) that there has been a "curious homogeneity in the criteria used for this research." In the main, criteria have been measures of advancement or indices of advancement such as salary growth, rate of promotion, increase in managerial responsibility, demotions, ratings and rankings of overall performance, rating of potential for advancement and personal data records (Huck, 1977).

Wernimont & Campbell (1968) make the point that this class of criterion measures constitutes a sign of behavior, rather than a direct sample of behavior. Klimoski & Strickland (1977) also argue that these criteria are not a direct measure of performance effectiveness since there are other forces in the organization that may determine advancement, pay progression, rate of promotion, etc. Thus, they argue that the use of these criteria may lead to a special case of criterion contamination--since these criteria may have more to do with "managerial adaptation and survival" (p. 355) than with managerial effectiveness. Therefore, their argument continues, the obtained validities may be spurious. If this is indeed true, it may

be true that instances of racial and sexual discrimination are also being obscured. This in fact does seem to be the case.

It is well known (Guion, 1965; Landy & Trumbo, 1976; Schmitt & Hill, 1977) that supervisory ratings are prone to judgmental errors and biases. It is certainly possible that these same types of errors are occurring systematically in predictor measures. Moreover, there is a substantial and growing body of literature that demonstrates the existence of discrimination in such personnel decisions as selection and placement, training and development, and compensation (see e.g., Bigoness, 1976; Cash, Gillen, & Burns, 1977; Cohen, 1976; Cohen & Bunker, 1975; Dipboye, Arvey, & Terpstra, 1977; Rosen & Jerdee, 1973, 1974 abc). Discrimination need not be a conscious, active process. Rather, sex and racial biases tend to be culturally ingrained and may operate unconsciously (Cohen & Bunker, 1975; Rosen & Jerdee, 1973, 1974 abc; Schein, 1973, 1975).

Research on Sex-Role Stereotyping

In a series of studies examining the effects of sex-role stereotyping on various personnel decisions, Rosen and Jerdee (1973, 1974 abc) found sex-role discrimination against both males and females. Discrimination was more pervasive against females; males tended to receive greater organizational support. When it came to the choice of selecting or promoting a male or an equally qualified female, males were favored over females. Males were also favored over females for career development; young promotable males were selected significantly more often than young females. Interestingly, respondents in the survey (Rosen & Jerdee, 1974c) favored sending the older, loyal employee for development regardless of sex.

Sex-role discrimination occurred against males in that management expected males to give top priority to their jobs, regardless of family responsibilities. Females, on the other hand, were treated more leniently when it came to taking time off for family matters; this occurred because females are expected to sacrifice their careers to family obligations.

Cohen & Bunker (1975) and Cohen (1976) also found evidence that selection decisions are affected by sex-role stereotyping. These two studies indicated that there is a significant interaction effect between applicant sex and the traditional sex-role connotation of the job. Cohen and Bunker (1975) found that significantly more females were recommended for a traditionally female job (editorial assistant), while more males were recommended for a personnel technician job which was perceived as a traditionally male occupation. Cohen (1976) points out that when recruiters were asked to rate applicants on their qualifications, no significant differences between the sexes were found. In both studies, recruiters were given a packet of information on each of the candidates for each of the jobs. The information they were given for both male and female applicants was the same for the same job description, except for the applicant's name. This manipulation tends to make a strong case for the interaction effect reported. On the basis of their data, Cohen & Bunker (1975) concluded that both sexes appear to be victims of sex-related biases in the recruitment interview.

The results of these studies lend further support to the impact of sex-role stereotyping on career opportunities for both women and men. The effect though has a greater and more pervasive impact on

women. For example, Cohen & Bunker (1975) reported that women are disproportionately underrepresented in upper level management and professional positions. There are 61 percent white collar males employed in these positions as compared to 32 percent white collar females.

Schein's (1973, 1975) work on perceptions of management characteristics helps to further clarify how sex-related biases limit career opportunities for women. She administered an adjective check list to a sample of male managers (Schein, 1973) and female managers (Schein, 1975) which was designed to solicit characteristic descriptions of males in general, females in general, and middle managers in general. Intra-class correlation coefficients were then computed between the descriptions of men and the descriptions of managers, and between the descriptions of women and the descriptions of managers. Although both correlations were significant, the degree of correspondence between males and managers was significantly greater than the correspondence between women and managers (Schein, 1975). These results suggest that men, more than women, are perceived to have the qualities and temperaments ascribed to successful middle managers. Interestingly, both male and female managers shared these perceptions. Schein (1975, p. 343) concludes that:

To the extent that this association between sex-role stereotypes and requisite management characteristics fosters a view of women as being less qualified than men for managerial positions, the results imply that female managers are as likely as male managers to make selection, promotion, and placement decisions in favor of men.

Indeed, this finding corroborates those of Rosen & Jerdee (1973, 1974c) which show both male and female managers often share negative

attitudes toward women in management positions. Apparently both sexes are similarly influenced by sex-role stereotypes (see also studies by Dipboye, Arvey, & Terpstra, 1977; Goldberg, 1968; Pheterson, Kiesler, & Goldberg, 1971; Shaw, 1972, and Terborg & Ilegen, 1975).

Research on Racial Bias

Of course, discrimination is not necessarily limited to sex-role stereotyping. As Wexley & Nemeroff (1974) have pointed out, skin color is an extremely salient personal characteristic which can affect a person's attitudes and feelings about another. Over the last thirty years various studies have focused on the systematic effects of race on ratings of performance. Results often are subtle, the statistical or practical differences not being very strong. Nevertheless, there is some consistency across studies showing that racial effects do exist. Therefore this is an important area for study, particularly given the increasing legislation for fair employment, the increase in integrated work groups and the ever increasing need for more effective employees as well as more effective managers.

Two early studies (Cox & Krumboltz, 1958; DeJung and Kaplan, 1962) which focused on the effects of race on peer ratings, found that raters tended to give higher ratings to those of their own race. DeJung & Kaplan (1962) found this effect more pronounced for black than white raters. In a later study, again designed to evaluate the effect of race on peer ratings, Schmidt & Johnson (1973) found a tendency for same race raters to rate those of their own race higher. The results did not reach statistical significance, and in this case the effect was greater for white than for black raters. On the basis

of these results, Schmidt & Johnson (1973) concluded that racial bias is not an inevitable outcome when using peer ratings. The study was designed so that approximately 50 percent of the peer group consisted of blacks. In addition, raters participated in a training program which emphasized interracial fairness and understanding. The authors concluded that the use of a racially balanced rating team and the use of a human relations training program emphasizing interracial awareness may have accounted for the nonoccurrence of racial biases in the ratings.

In another study of the effects of race and sex on performance ratings, Hamner, Kim, Baird, & Bigoness (1974) found that raters tended to rate members of their own racial group higher than members of the other group on a measure of overall task performance. Although the differences were only significant at the .05 level, the data are particularly noteworthy since actual performance was standardized by the use of a work sample procedure. In addition, the work samples viewed by the raters were filmed, thereby decreasing any extraneous factors which could have led to differing ratings among raters. Another finding was that while raters were clearly able to distinguish between high performing whites and low performing whites, they tended to rate high performing blacks only slightly better than low performing blacks--i.e., they tended to rate both high and low performing blacks as average workers. There was a tendency to favor low performing blacks over low performing whites, while favoring high performing whites over high performing blacks. The authors conclude that "the fact that blacks received significantly lower ratings than whites from white raters when performance levels were identical indicates a potentially serious problem of racial bias" (p. 709).

In a replication of the above study, Bigoness (1976), using the same films and the same methodology, found no significant differences as a function of the race of the performer. In an analysis of the simple effects of the interaction between the ratee's race and performance, however, it was found that low-performing blacks were rated significantly higher than low performing whites (7.15 vs. 6.33). As in the previous study, Bigoness (1976) concludes that performance ratings were biased as a function of the performer's race. Bigoness (1976) suggests that these effects might be mitigated by using a more clearly objective measure of performance and by training raters about the potential impact of sex and race biases on performance ratings.

To test the notion that a behaviorally anchored measuring instrument would be resistant to racial biases, Brugnoli, Campion, & Basen (1979) designed a study using a work sample for the job of maintenance mechanic. They further hypothesized that the work sample should be maximally representative of the critical job behaviors. In order to test these hypotheses, 56 white male maintenance mechanics, thoroughly familiar with the tasks and equipment were used as raters. All Ss were volunteers.

Four videotapes were prepared--each representative of an experimental condition. These consisted of a job relevant task and a job irrelevant task performed by both a black mechanic and a white mechanic. The work samples were devised as the result of a thorough job analysis performed for a previous study by Campion (1972). Each of the videotaped performances was prepared identically in order to avoid any extraneous sources of variance in the ratings. Additionally, videotapes showed only the hands and arms of the performers in

order to avoid any possible contamination due to the attractiveness (Dipboye, Fromkin, & Wiback, 1975) of the applicant. The 56 Ss were divided randomly into two groups--one group using only a global rating scale, the other using a behaviorally specific recording form followed by the global rating scale.

The results indicated that black applicants were rated lower only in the task irrelevant condition when the global rating scale was used. Blacks and whites were not rated significantly differently in the task relevant condition when the behaviorally specific scale was used nor when the global scale was used. The authors concluded that a well developed work sample procedure which is representative of performance that is critical to success and failure on the job will generally be resistant to the effects of racial bias. In addition, the instrument used to measure performance should be behaviorally specific, rather than relying on global evaluations of performance.

In terms of the methodology of this study, several comments should be made. First, this study involved a non-managerial job in which critical behaviors may have been easier to specify than those in managerial positions. Secondly, only white raters were used. Thus, it was not possible to compare the effects of the interaction between rater's race and ratee's race on performance scores. This study does, however, extend the findings of Hamner et al. (1974), since their investigation used only a global rating scale. Had Hamner et al. (1974) used a more behaviorally specific measuring instrument, perhaps they would not have found any effects due to race. This would certainly be a fruitful area for further research.

Brugnoli, Campion, & Basen (1979) suggested that a possible remedy to the problem of racial bias in evaluations is to have raters simply describe what they observe rather than evaluate it. This certainly seems like a reasonable course of action since the conclusions reached by Brugnoli et al., corroborate those of several authors, including Dunnette & Borman (1979), Hamner et al. (1974), Rosen & Jerdee (1974 abc), and Schein (1975). These authors contend that when interviewers or raters have little task relevant information on which to base a judgment, and must make global evaluations about a person's potential performance, they are forced to supplement what little information they have. Thus they tend to rely on stereotypes and preconceived notions in order to reach their final conclusions (see also Wiener & Schneiderman, 1974). Remember too, that the Rosen & Jerdee (1974 abc) and Schein (1973, 1975) studies involved managerial positions. Therefore, it would seem that the Brugnoli et al. (1979) conclusions would generalize to management-type jobs.

The Brugnoli et al. (1979) study lends further empirical support to the argument advanced by Wernimont & Campbell (1968). They contend that using a behavioral consistency approach would reduce several measurement problems, such as response sets, faking, and discrimination (both for race and sex). They believe that psychologists should return to the study of behavior. After all, the best way to predict future performance is by having a sample of past performance. Both predictor and criterion measures should be as behaviorally specific as possible. In order to accomplish this, a thorough job analysis must be conducted before the development of any evaluation instrument. The analysis must identify not only those

behaviors that are most representative of performance, but also those which are critical to successful, as well as unsuccessful performance on the job.

Several other studies present evidence of the influence of ethnic biases on evaluations. For example, in a recent literature review comparing black versus white leadership behavior, Bartol, Evans, & Stith (1978) point out that whites are evaluated on different criteria than are blacks. Black leaders are more often judged on interpersonal skills than on task-related or job content factors. The Huck & Bray (1976) study is cited as an example. In this investigation it was found that assessment center ratings on administrative skills and effective intelligence were more predictive of job performance for white women than for black women. On the other hand, sensitivity (to the social environment, one's strengths and weaknesses, and company liabilities) seemed to be a better predictor of job success for the black women.

In a study by Richards & Jaffee (1972), subordinates with more liberal attitudes tended to give their black supervisors higher ratings, particularly on human relations skills, than subordinates with less liberal attitudes. Black supervisors were rated significantly lower than white supervisors and subordinates supervised by whites behaved differently from those supervised by blacks. An additional finding was that differences in subordinate behavior appeared to decrease the effectiveness of black supervisors.

A second conclusion reached by Bartol et al. (1978) was that ethnic characteristics of both the rater and the ratee may affect judgments about performance. They point to several studies, including Cox & Krumboltz (1958), DeJung & Kaplan (1962), and Hamner et al.

(1974), as well as several others, as evidence of this phenomenon. In sum, these studies show that raters tend to rate individuals in their own ethnic group higher than those in other ethnic groups. The evidence is complex, and Bartol et al. (1978) point out that not all studies have found the same results (see e.g., Schmidt & Johnson, 1973; and Vinson & Mitchell, 1975).

Schmitt and Hill (1977) investigated the possibility that the ethnic composition of a group in an assessment center may affect overall performance scores. While the results were of marginal statistical and practical significance, the data suggest that the ethnic and sex composition of an assessee group can have an effect on the assessee's performance, as well as on the ratings received.

In another study using a work sample procedure, Schmitt & Lappin (1980) found that the ratings of black raters were significantly less variable when rating white ratees, than when rating blacks. Ratings of white raters were significantly less variable when they were rating blacks than when they were rating whites. Additionally, raters indicated that they felt more confident in their judgments when they were rating individuals in their own ethnic grouping. This was a partial confirmation of Schmitt & Lappin's (1980) hypothesis that people feel more comfortable when evaluating others who they perceive to be similar to themselves. Raters will use more of the scale when evaluating those that they perceive as being similar to themselves and thus, ratings will show more variation. Interestingly, although the hypothesis was confirmed for the ethnic subgroups, the same effects did not obtain for sex subgroupings. Another interesting finding was that while black raters tended to

rate blacks higher than whites, white raters also rated black ratees slightly higher than white ratees. This study thus corroborates the Bartol et al. (1978) contention that while ethnic characteristics do affect performance ratings, the data are neither straight-forward, nor consistent.

In line with the Richards & Jaffee (1972) and Schmitt & Lappin (1980) studies reported above, Wexley & Nemreroff (1974) investigated the influence of applicant race and biographical similarity of the applicant to the interviewer on outcomes of a selection interview. Subjects consisted of 120 (96 males, 24 females) white undergraduates who volunteered to participate as interviewers. Subjects completed the Multifactor Racial Attitude Inventory (MRAI), as a measure of their attitude toward blacks. Based on the MRAI scores, Ss were divided into two groups of either high or low prejudice.

Regardless of the race of the applicant or the prejudice of the interviewer, perceived similarity of background (between interviewer and applicant) proved to be the major determinant of interviewer evaluations. The race of the applicant had little effect on the evaluations. While it was also found that low prejudiced Ss gave significantly higher ratings to applicants than did those high in prejudice, this was attributable to the fact that high prejudiced Ss gave lower ratings to those they perceived as biographically dissimilar to themselves. Thus, while the result of this study does not in and of itself show a negative impact due to ethnic bias, in conjunction with the Schmitt & Lappin (1980) study, it shows that evaluations are affected by perceptions of dissimilarity between an evaluator and a ratee. In an assessment center, where assessors are not

familiar with the background of a candidate, ethnic or sex differences could contribute to a bias in the performance ratings.

The Present Study

In sum, this review suggests that sex and race biases do operate to influence selection decisions, evaluation decisions, and other personnel decisions. Worse yet, studies such as Bigoness (1976), Hamner et al. (1974), and Schmitt & Lippin (1980), show that these sex and race-linked biases operate in work sample procedures, where the actual behavior of males and females, and blacks and whites, is identical and has been carefully controlled. This leads one to question whether these biases operate in an assessment center--which after all, is a work sample. Perhaps the fact that assessors assign overall performance scores in teams, or perhaps because assessors record actual behavior, assessment center participants may not suffer from discrimination attributable to sex-related or ethnic-related biases. Few, if any, studies have investigated systematic bias on assessment ratings due to the race and sex of assessees or the race and sex of assessors. The purpose of this study is to investigate the impact of assessor race (sex is not being examined since only three females served as assessors) and the sex and race of assessment center participants on assessment ratings. More specifically, the following questions will be investigated:

1. Are there systematic differences in the overall assessment scores of participants between all white assessor teams versus racially mixed assessor teams across participant race, sex, and the various combinations thereof?
2. Overall, are there systematic differences in participant

scores due to the assessor's race or the participant's race or sex?

3. Are there systematic differences in performance ratings across assessment exercises as a function of the participant's race or sex?

The answers to the above questions can be helpful in determining if assessment centers are indeed a viable means for reducing "adverse impact" in selection, and for meeting affirmative action goals.

CHAPTER II

METHOD

Subjects

A total of 265 employees of a large southeastern utility company participated in the assessment center. Participants were candidates for a management development training program, the main purpose of which was to prepare them for positions in middle and upper management. A second purpose of the program was to increase opportunities for minorities and women to reach these positions. The program included such activities as the initial assessment, formulation of an individual development plan, seminars, job rotation, continuing education, and special assignments (such as placements with outside organizations). Of the 265 candidates, 50 were selected for the development program.

Table 1 shows the demographic breakdown of the Ss.

Table 1. Demographic Breakdown of Ss

<u>Group</u>	<u>N</u>	<u>%</u>
Black Male	68	25.66
Black Female	30	11.32
White Male	68	25.66
White Female	90	33.96
Hispanic & Orientals	9	3.40

Because of the small number of Orientals and Hispanics in the sample, they were dropped from the analysis. Therefore, the total remaining sample size was equal to 256.

The mean age of participants was 35.34 (S.D.=7.83) years with a range from 22.58 to 58.33 years old. The mean tenure of participants was 8.46 years (S.D.=6.71), with a range of nine months to 33 years. The mean tenure in present position was 2.63 years (S.D.=2.21) with a range from nine months to 17.75 years.

The majority of candidates (approximately 225) were nominated for the program by their supervisors. Supervisors completed a form recommending a candidate on the basis of nine dimensions. A description of the nomination form and the dimensions appears in Appendix A. The remaining 40 candidates were self-nominated. These candidates submitted the same nomination forms, completing the forms themselves, and then applied for consideration into the program through the Equal Employment Opportunity (EEO) Office. Only individuals currently employed on the lower management pay schedule, the technical/professional pay schedule, or the administrative pay schedule were eligible to be nominated. Participants were selected from each of the six major organizational units and from each of the three major regional areas.

A total of 23 managers served as assessors. They were selected from varying levels of the Management Schedule, representing mainly middle and upper management. They represented a cross sample of the major organizational units. Assessors ranged in age from 26.50 to 56.08 years, the mean age being 41.19 (S.D.=8.60). The mean tenure for assessors was 13.65 years. Assessors' mean tenure in their

present position was 2.83 years (S.D.=3.91), with a range of 0.17 to 8.58 years. The assessor group included four black males, one black female, 16 white males, and two white females.

Instrumentation

Three exercises were developed to simulate the typical job activities of middle managers. Each of the simulations was based on a thorough job analysis of the target level jobs. The job analysis had been conducted just prior to the development of the simulations for the specific purpose of ensuring the content validity of the exercises (see e.g., Campion, 1972; O'Leary, 1973; Wernimont & Campbell, 1968). The simulations were then developed by an outside consulting firm specializing in assessment center design.

A brief description of each of the simulations follows (a more complete description of the exercises, with examples, can be found in Appendix B):

1. In-Basket Exercise and Interview

This is an individual exercise. Each candidate is required to assume the role of a hypothetical person in an organization and is given a set of memorandums and briefings as might be found in a manager's in-basket. A hypothetical organizational situation is given to each candidate, and she/he is asked to take the appropriate action on each item by writing letters, memos, and notes to him/herself, to subordinates, or superiors. After completing the in-basket each candidate is interviewed by an assessor regarding her/his approach to the task, her/his rationale for taking the actions indicated, and the opinions she/he developed regarding subordinates, peers, supervisors, and the organization. This exercise is

timed, candidates having 75 minutes to complete the in-basket and having 20 minutes for the interview.

2. Leaderless Group Discussion

Six individuals participate in a panel discussion. Each of them is given general background information about some hypothetical situation. Additionally, each is given a different position to defend. The objective of the discussion is for each panel member to convince the others that his/her position should be accepted. The panel is required to ultimately arrive at a group decision about which of the positions is most meritorious. Participants have 20 minutes to review the background information and 45 minutes to discuss the problem and arrive at a group consensus.

3. Task Direction Problem

In this exercise, each candidate is given a hypothetical business problem and is required to assimilate and process a considerable amount of data. In this case, the candidate is required to schedule both people and machines in order to meet production deadlines. The participant is given a set of materials which includes the problem and all relevant information required to solve the problem. In addition, candidates are assigned an assistant to help complete the task. The assistant is actually a specially trained role-player who acts as a distractor. The objective of this exercise is to observe how the candidate manages both the task and the assistant.

The exercise proceeds in four phases. First, the candidate is given a package of information and directions and has five minutes to look it over. Next, the assistant is introduced and the participant

is told that she/he has 45 minutes to complete the task. The third phase consists of a feedback/counseling session. Candidates are required to give feedback to the assistant about her/his performance during the task. They have 15 minutes to accomplish this. In the last portion the participant is asked to write a short summary of the initial task and what was accomplished, and to write a short summary of the counseling session. They have 15 minutes to complete both summaries.

Scoring and Dependent Variables

As a result of the job analysis, nine dimensions were identified to be important to successful managerial performance within the organization. They are: (1) Leadership, (2) Perception, (3) Adaptability, (4) Decisiveness (refers to the number of decisions made), (5) Decision-making and Judgment (refers to the quality of decisions made), (6) Organization and Planning, (7) Sensitivity (refers to the ability to deal effectively and sensitively with others), (8) Written Communication, and, (9) Oral Communications. A more complete definition of the dimensions and the Exercise Report Forms appear in Appendix C.

Because the instruments are designed to elicit different modes of behavioral responses (e.g., oral, interactive, individual problem-solving, and written responses) the three simulations do not all measure the nine dimensions. The In-Basket measures eight of the nine, not measuring adaptability. The Group Discussion (LGD) also measures eight dimensions, excluding written communication. The Task Direction Exercise does measure all nine dimensions.

Scores on each of the dimensions for each of the exercises are

based on a seven-point scale: (1) poor, (2) well-below average, (3) below average, (4) average, (5) above average, (6) well above average, and (7) excellent. Scores are assigned on the basis of the observed participant behaviors. An overall dimensional score is calculated across the three exercises. This score is based on each of the assessor's judgment of a candidate's overall performance on the dimension, rather than just the arithmetical mean of the scores for each exercise on that particular dimension. The overall score is derived after the three assessors (each assessor rates the participant on one exercise) compare and discuss the candidate's performance on each of the three simulations. The overall score is also based on a weighting factor roughly representing the ability of the simulation to measure that particular dimension. For example, oral communications is weighted "one" on the in-basket, "two" on the LGD, and "three" on the task direction. This means that oral communication is most observable on the Task Direction exercise, next most observable on the LGD, and least observable on the In-Basket. A skill matrix weighting form appears in Appendix C which helps to clarify this.

The dependent variables in this study are derived directly from the scores described above. In analyzing question one, the overall score, which is derived by a three-person assessor team, is the dependent variable. In question two, the individual scores on each of the exercises are used as the dependent variable since this score is supposed to be the individual judgment of each assessor. Unfortunately, in some cases, assessors changed their initial rating of a candidate after discussion of the candidate's performance with the two other assessors on their team. Since there was no way to control

this, or to isolate instances when it occurred, it represents a possible confounding of the data.

In looking at the third research question, the dependent variable is the mean of the nine dimensional scores on each exercise. A further analysis involves comparison of the scores on each of the dimensions across the three exercises.

Procedure

After arriving at the motel where the sessions were to be held, participants received their schedules for the day. Twelve candidates were assessed each day except when one or two candidates were not able to attend; these individuals were re-scheduled for another day. A brief orientation meeting was held to inform participants about the purpose of the center, what exercises would be used, what dimensions would be assessed, to introduce the assessment center staff, and to answer any questions candidates might have. Six assessors, divided into two teams of three, served each day of the center's operation. Originally, the design called for a different group of six assessors to serve each week of the center's operation. Because of other work commitments not all assessors could serve for a full week. As a result, replacements were made when necessary from the pool of 23 assessors.

Assessment operations began immediately following the orientation meeting. Both assessors and participants changed rooms after each exercise. This was to ensure that each assessor and participant pair was correctly matched. (A sample schedule for both assessors and participants appears in Appendix D.) Care was taken to ensure that no participant was assessed by his/her supervisor or by a

person in his/her chain of command. After participants completed the exercises a debriefing session was held. Here, candidates had the opportunity to engage in a dialog with the program administrator around their initial expectations about the center, and their experiences of the day. Additionally, candidates could ask any questions they had about the management development program, the center, the use and confidentiality of the data, and how this experience might affect their careers. The entire procedure took approximately six hours. The center began each day promptly at 9 a.m. and candidates were finished about 3 p.m.

After the participants left assessors began the rating procedures. First, behaviors recorded during the center were categorized into the nine dimensions and then a rating of one to seven was assigned. Since each assessor was responsible for assessing two candidates on each of the three exercises, assessors had six Exercise Report Forms to complete (see Appendix C). After completing the Exercise Report Forms, assessors met in teams of three to determine each participant's overall assessment score. The final step in the process involved summarizing the data on the Skill Matrix Weighting Form (see Appendix C) which all assessors did as a part of the team meeting.

Because of the large number of participants that were assessed, the center was run every working day over a five-week period (from June 11 until July 19, 1979). Approximately 60 participants were assessed during each week of operation.

Two additional points should be mentioned. Two weeks before the center began, the line managers who were to serve as assessors

underwent a three-day training program to learn how to observe and rate behavior. Additionally, an attempt was made to balance the assessor group as much as possible with blacks and whites, and males and females in order to avoid any charges of discrimination.

CHAPTER III

RESULTS

Analysis of Overall Assessment Center Scores

The design for the first analysis consisted of a 2x2x2 factorial analysis of variance performed on the overall assessment scores. The independent variables were participant race (black versus white), participant sex (male versus female), and racial composition of the assessor team (racially mixed versus all white). The analysis was conducted using the General Linear Model procedure of the 1979 version of the Statistical Analysis System (Helwig & Council, 1979). This procedure was chosen because it can accommodate unequal cell sizes.

Examination of Table 2 indicates that each of the main effects was significant.

Table 2. Analysis of Variance Comparing Overall Assessment Score Across Participant Race, Participant Sex, and Racial Composition of Assessor Teams

Effect	df	MS	F
Participant Race (A)	1	17.266	24.43 ^a
Participant Sex (B)	1	2.782	3.94 ^c
A x B	1	0.375	0.53 ^b
Racial Composition of Team (C)	1	5.508	7.79 ^b
A x C	1	0.011	0.02
B x C	1	1.618	2.29
A x B x C	1	0.435	0.62
Error	248	0.707	

NOTE. R^2 for model = 0.154

a. $p < .0001$

b. $p < .006$

c. $p < .05$

Blacks scored lower than whites (3.502 versus 4.114), males scored lower than females (3.732 versus 4.047), and the racially mixed assessor teams scored candidates slightly lower than did the all white assessor teams (3.809 versus 4.056). Although statistically significant the main effects are of marginal practical significance since the largest mean difference (for blacks versus whites) is only 0.612 on a seven point scale. None of the interaction effects even approached significance.

Analysis of Mean Performance Scores on Each Assessment Exercise

In this analysis, the dependent variable was the mean performance score on each of the three assessment exercises across the nine dimensions. Each participant had three mean scores, one for the in-basket, one for the leaderless group discussion, and one for the task direction exercise. The independent variables were participant race, participant sex, and assessor race.

Table 3 indicates that only the main effect for participant race reached significance. As would be expected from the previous analysis, blacks were rated lower than whites (3.535 versus 4.133). Again, the mean difference is small, being 0.598. As would also be expected, females scored higher than males - (4.054 versus 3.772) - but not significantly so.

Although none of the other effects reached significance, two findings are noteworthy. First, the ratings associated with black assessors are almost identical to those associated with white assessors (3.90 versus 3.91). Secondly, when the interaction between Assessor Race x Participant Race is examined, it is found that black assessors rated blacks slightly higher than white assessors rated

blacks (3.56 versus 3.53), while black assessors rated white participants slightly lower than white assessors rated white participants (4.07 versus 4.16). The interaction was not significant, and the mean differences are small enough to be trivial, but the differences are in the direction found in several other studies, including Cox & Krumboltz (1958), DeJung & Kaplan (1962), and more recently, Hamner, et al. (1974).

Table 3. Analysis of Variance Comparing Mean Performance Scores on Each Exercise Across Participant Race, Participant Sex, and Race of the Assessor

Effect	df	MS	F
Participant Race (A)	1	29.089	25.32 ^a
Participant Sex (B)	1	12.664	2.32
A x B	1	1.008	0.88
Race of Assessor (C)	1	0.072	0.06
A x C	1	0.306	0.27
B x C	1	0.136	0.12
A x B x C	1	0.022	0.02
Error	760	1.149	

NOTE. $R^2 = 0.075$

a. $p < .0001$

Analysis of Mean Performance Scores Across Assessment Exercises

For this analysis, the dependent variable again was the mean performance score on each of the three assessment exercises. The design used was a 2x2x3 analysis of variance with repeated measures on the third factor. The independent variables were participant race (black versus white), participant sex (male versus female), and

assessment instrument (in-basket, leaderless group discussion, task-direction exercise). In addition to the analysis being done across the nine assessment dimensions, the same analysis was conducted for each dimension. This analysis was conducted in order to see on which, if any, of the nine dimensions participant performance differed.

Results of the overall analysis are shown in Table 4. Examination of the table indicates three significant effects, the main effects for Participant Race and Assessment Exercise, and the interaction of Assessment Exercise x Participant Sex.

Table 4. Analysis of Variance With Repeated Measures on Assessment Exercise Comparing Mean Performance Scores on Each Exercise Across Participant Race, Participant Sex, and Assessment Exercise

Effect	df	MS	F
Participant Race (A)	1	47.264	24.367 ^a
Participant Sex (B)	1	4.879	2.516
A x B	1	1.791	0.923
Error	252	1.940	
Assessment Exercise (C)	2	12.239	17.768 ^a
A x C	2	0.266	0.386 ^b
B x C	2	4.138	6.008 ^b
A x B x C	2	0.365	0.530
Error	504	0.689	

a. $p < .0001$

b. $p < .003$

Black participants scored 3.535 as compared to a mean score of 4.133 for white participants. Mean scores across all participants for the assessment exercises were 4.057, 4.014, and 3.642, for the

in-basket, leaderless group discussion, and the task direction exercise, respectively. A Duncan's Multiple Range Test (Kirk, 1968) performed across the means shows that participants scored significantly lower ($p < .05$, error $df = 504$) on the task direction exercise compared to either the in-basket or leaderless group discussion exercises. The significant B x C interaction shows that females scored significantly higher than males on the in-basket (4.345 versus 3.803) and task-direction exercises (3.808 versus 3.495), but scored about the same as males (4.009 versus 4.018) on the leaderless group discussion. Males made their strongest showing on the leaderless group discussion (3.803, 4.018, 3.495 for the in-basket, leaderless group, and task-direction, respectively).

Table 5 shows the results of the F-tests across all nine skill dimensions, as well as the F-values for each skill dimension.

Black participants were rated significantly lower than whites on all nine of the assessment skill dimensions. Table 6 shows the mean performance scores for black versus white participants.

Females scored significantly higher than males on four of the nine skill dimensions. These include sensitivity, organization and planning, and oral communications and written communications. Mean performance scores for each of the dimensions are shown in Table 7.

Although females scored significantly higher than males on four dimensions, examination of Table 7 shows that the mean performance scores for females were higher on every skill dimension except for adaptability, where males scored only slightly higher. These findings are consistent with those of Bigoness (1976) and Hamner et al. (1974) who also found females rated higher than males on performance.

Table 5. Results of Repeated Measures ANOVA on Each Assessment Dimension

Effect	df	Over-all F	Lead F	Sensit F	Percep F	Dec-Mk F	Decisive F	O & P F	Adapt. F	Oral F	Written F
Participant Race (A)	1	24.367 (0.0001)	15.150 (0.0001)	4.555 (0.034)	29.410 (0.0001)	27.135 (0.0001)	7.230 (0.008)	21.108 (0.0001)	6.664 (0.010)	7.012 (0.009)	21.296 (0.0001)
Participant Sex (B)	1	2.516	0.0	4.779 (0.030)	1.888	1.185	0.100	3.951 (0.048)	0.903	4.924 (0.027)	12.332 (0.001)
A x B	1	0.923	0.492	1.269	0.548	0.387	0.002	1.389	0.134	2.101	0.997
Error MS	252	1.940	4.108	2.844	3.256	3.318	3.861	4.185	2.525	2.011	1.677
Assessment Exercise (C)	2	17.768 (0.0001)	2.611 (0.074)	8.174 (0.0001)	63.757 (0.0001)	41.148 (0.0001)	16.692 (0.0001)	33.905 (0.0001)	7.071 (0.008)	1.720	2.019
A x C	2	0.386	3.058 (0.048)	1.063	2.467 (0.086)	0.529	4.423 (0.012)	0.697	1.762	0.255	8.214 (0.005)
B x C	2	6.008 (.003)	4.302 (0.014)	0.330	2.190	4.019 (0.019)	4.764 (0.009)	5.954 (0.003)	1.042	0.337	1.394
A x B x C	2	0.530	0.218	0.057	0.326	0.328	1.164	0.416	0.0	0.486	0.495
Error MS	504	0.689	2.112	1.685	1.648	1.559	1.829	1.667	1.356	0.806	0.913

NOTE. Numbers in parentheses are the probabilities associated with the F-tests.

Table 6. Mean Performance Scores by Race on Each Skill Dimension

Dimension	Black	White
Overall	3.535	4.133
Leadership	3.279	3.911
Sensitivity	4.130	4.506
Perception	3.170	4.002
Decision-making	2.908	3.705
Decisiveness	3.895	4.323
Org & Plan	3.092	3.937
Adaptability	3.306	3.655
Oral Comm.	4.228	4.608
Written Comm.	3.694	4.399

NOTE. N = 294 blacks, 474 whites

Table 7. Mean Performance Scores by Sex on Each Dimension

Dimension	Male	Female
Overall	3.772	4.054
Leadership	3.610	3.736
Sensitivity	4.211	4.533
Perception	3.515	3.875
Decision-making	3.250	3.569
Decisiveness	4.088	4.239
Org & Plan	3.400	3.856
Adaptability	3.537	3.504
Oral Comm.	4.331	4.611
Written Comm.	3.871	4.421

NOTE. N = 408 males, 360 females

Table 5 indicates that participants' mean performance was significantly different on six of the nine skill dimensions across assessment exercises. This result would be expected since the instruments were designed to measure different aspects of the nine skills. Examination of Table 5 also indicates that several of the two-way interactions are significant. However, these differences do not appear to be theoretically interesting or meaningful.

CHAPTER IV

DISCUSSION

The purpose of this study was to determine whether there were any systematic differences in assessment center ratings related to the race and sex of the ratees or the race of the assessors. In effect, the study was designed to investigate the possibility of discrimination in assessment centers due to race or sex-linked biases. It is clear from the data, that at least in this center, no such discrimination was evident. There are several possible explanations for these findings.

The assessors both work with and supervise minorities and females. This routine on-the-job association may have helped to reduce sex or race-related stereotypes that may have biased assessment center results. While it is still true that white males continue to dominate upper and middle management positions in this organization, the organization is concerned about affirmative action. Although all managers may not share the same convictions about affirmative action, there is a growing awareness about these issues among managers. As Schmidt & Johnson (1973) have suggested, awareness of human relations issues may help to reduce the impact of sex or race-related biases on ratings.

Another explanation, again in line with the findings of Schmidt & Johnson (1973), is that whenever possible both raters and assessees participated in racially and sexually mixed groupings. Group composition for both assessors and participants was purposely mixed in

order to eliminate, or prevent as much as possible, discrimination related to sex or race biases.

Only one recent study by Schmitt & Hill (1977), has explored the effects of sex and race composition of assessee groupings on assessment center outcomes. While the results of their investigation showed only marginal statistical and practical significance, their findings suggested that the ratings of black females may have been adversely affected by the race and sex of other group members. Given the trend towards increasingly heterogeneous work groups, even subtle effects may be of importance. To this end, the present study compared the ratings of racially mixed assessor teams versus all white assessor teams. Unfortunately it was not possible to form a team of all black assessors, because of the limited number of black assessors in the assessor pool (5 blacks versus 18 whites). Perhaps future investigations can be designed which include all possible combinations of sex and race in the assessor teams.

Racially mixed teams rated participants significantly lower than did the all white teams (see Table 2, page 32). However, mean differences were small (being 0.25) indicating marginal, if any, practical significance. Thus these data are difficult to interpret--particularly in light of the fact that individual ratings for black assessors were almost identical to those of the white assessors (see Table 3, page 34). A possible explanation for this outcome may be because assessors had to change teams frequently--one time working with an all white team, the next time working with a racially mixed team. This occurred because assessors often had other work commitments and could not serve the full week as planned. When an assessor

was absent, a replacement was made from the original assessor pool on an as available basis. Thus it may be that the experiences an assessor carried over from one team to another caused these results.

Another explanation may be that in some cases, assessors changed their initial ratings of candidates as a result of the team discussion. Although changes could have been made in either direction, the fact that changes were made does obscure the relationship of initial ratings to team ratings. Additional research on the impact of the team meetings on ratings, and on the sex and race composition of both assessor and participant groups, would provide some further insights.

Several authors, including Bigoness (1976), Brugnoli et al. (1979), Hamner et al. (1974), and Schmidt & Johnson (1973), have suggested that rater training may reduce the effects of stereotyping on ratings. In the present study assessors received three days of training on how to observe and record behavior. Although the training did not focus specifically on human relations awareness, it did focus on observing and rating behavior objectively. Assessor training appears to be one of the critical elements for ensuring that assessment centers are discrimination-free.

Another explanation is that assessment ratings were based on actual observations of behavior. In addition, assessors had to defend the ratings to other members of their teams. The defense always related to the recorded observations of the participant's behavior. Thus, even though the rating scale itself was global in nature, the ratings were based on observations of performance. Bigoness (1976), Brugnoli et al. (1979), and Rosen & Jerdee (1973, 1974 abc) all suggest that stereotyping and biases will be reduced

when the evaluation system is focused behaviorally. This is also consistent with the work of Campion (1972) and Wernimont & Campbell (1968).

The data shown in Table 7 (page 38) are also of some interest. Examination of Table 7 shows that the mean performance ratings for females were higher than the mean performance ratings for males on every dimension except Adaptability. Additionally, females scored significantly higher than males on four of the nine dimensions (see Table 5, page 37). Both Bigoness (1976) and Hamner et al. (1974) found that females were rated higher than males. Hamner et al. (1974) explained this finding by suggesting that when females are seen as performing equally well as males in a traditionally male job, then the women are perceived to be better performers. Bigoness (1976), on the other hand, suggested that sex-biases may be more of a problem when women are considered for professional positions, as opposed to non-professional jobs. In both these studies, the work sample used was that of stacking cans on a shelf.

In a replication of the above studies, Schmitt & Lappin (1980) had Ss stack books on a library shelf. They found no differences in ratings due to the sex of the subject. Thus the empirical data does not appear to be straightforward. Clearly, a study comparing sex effects across occupations that are stereotypically male or female and across professional versus non-professional occupations would be informative.

Another pervasive finding throughout this investigation was that white participants scored significantly higher than blacks. Examination of Table 5 (page 37) shows that whites were rated significantly

higher than blacks on every assessment dimension. Although the mean differences are small, if there were no differences between the two groups, one group could be expected to exceed the ratings of the other 50% of the time. A sign test (Siegel, 1956) calculated to establish the probability of the ratings given white participants consistently exceeding those of the blacks was significant at the .001 level (one-tailed test). Given the fact that the assessment center was designed to be a sample of the types of activities demanded on the job, these findings are somewhat distressing. If it can be assumed that the center is indeed a valid measure of job performance in this organization, then the only conclusion which can be reached is that the black participants in this study do not have skills comparable to those of the white participants.

Several explanations can be offered. Lerner (1980) recently pointed out that scores on national literacy tests such as the SAT have been on a steady decline. She construes this to mean that there has been a continuous longterm decline in academic preparedness and competency. She goes on to say that the skills which are necessary in organizational life are more and more the same as those needed for competent academic performance. Thus with the decline of academic performance there has been a parallel decline in our national productivity. Lerner (1980) contends that the greatest impact of this decline has been on the poor in general, and the black poor in particular.

In the present investigation, black participants working in the southernmost location of this organization received the lowest ratings. Interestingly, the white participants in that location

received lower ratings than whites in other locations (although their ratings were higher than those for black participants). These assesses live and work in an area of the United States considered a "deep-south" state. Approximately 20-25 years ago when these Ss attended grade school, the school systems were still segregated and the blacks attended "Black Schools." It is also well known that schools in the rural south were the weakest of all educational systems in the United States. This was true for both the black and white schools. As Lerner (1980) pointed out, a less than adequate background in the fundamental academic skills can influence later job performance. This may account for the poor performance of the whites from the southernmost location and the blacks in the present study.

Another possible explanation is that the blacks in this investigation may have had only limited opportunities for developmental experiences during their careers. Although legislation prohibiting discrimination is almost 20 years old, Nason (1976) contends that the effects of past discrimination are cumulative. Thus the impact of an inferior elementary education and the exclusion from experiences which teach individuals how to cope in an organizational environment leave those individuals less prepared to deal successfully in today's corporate environment. Add to this the effects of racial prejudice and discrimination which still operate in organizations (although it is no longer fashionable, or legal to discriminate overtly) and these findings are not all that surprising.

Actually, there is much historical evidence to show that on the average, blacks do poorer on tests, and perform worse academically than whites (Guion, 1965). Arvey (1979) points out that on the

average blacks tend to score between one and one and a half standard deviations lower than whites on tests of general cognitive ability (see also Dreger & Miller, 1960, 1968; and Shuey, 1966). Many reasons have been advanced to explain these differences ranging from cultural diversity (Shuey, 1966), to lowered self-esteem and heightened test anxiety (Samuda, 1975), or factors inherent in the test itself, such as culturally biased items (Arvey, 1979). The issues are complex and the problem still persists. Block & Dworkin (1976) present a good review of some of these issues.

Nason (1976) has suggested a program with three levels of action which must be taken not only to help organizations comply with society's changing values and laws, but to also make better use of our human resources. He suggests that organizations first do an internal analysis to eliminate and correct any barriers that currently exist which block the upward mobility of blacks. These barriers may include irrelevant job qualifications, biased application of criteria for selection and promotion, or institutionalized policies and procedures which unfairly limit the advancement of minority groups.

The second level of action involves direct financial and organizational support. This support can take several forms including: compensatory management development programs for blacks, job rotational assignments, assessment of individual developmental needs, and the opening up of more managerial positions to allow minorities to gain the experience they need. Organizations could sponsor black student scholarships, or provide high schools and colleges with the technical and financial assistance needed in order to develop programs that will create stronger and more competitive educational

backgrounds for blacks. Upward mobility programs could also be developed to assist presently employed blacks in managing and attaining their career goals.

The third level of action involves organizations initiating and supporting efforts to influence government (at all levels) to change funding priorities. Funds could then be channeled toward programs that would ameliorate the conditions that continue to deprive blacks of an equal standing in our society. Such efforts could be directed through community organizations which seek to upgrade the living conditions of blacks--this would be particularly critical in our inner cities. Thus there are several initiatives which organizations can make in order to increase the supply of blacks prepared to assume managerial responsibilities.

In summary, the findings of the present investigation have important implications for the continued use of assessment centers. They suggest that assessment centers are free of biases related to sex or race discrimination and therefore may be used to promote equal employment opportunities. The three factors which seem essential to achieve this end are (1) a careful job analysis upon which the exercises are based; (2) assessor training on how to observe and record behavior; and (3) the use of sexually and racially integrated assessor and participant groupings.

In addition, the findings indicate that organizations must take steps to assist blacks in further developing the skills necessary to assume managerial responsibilities.

Directions for Future Research

Happily, the findings of this study suggest that assessment

centers are a viable method for reducing adverse impact in selection decisions. As such, they can be a useful tool in helping organizations to meet affirmative action goals. Nevertheless, it must be remembered that this was an N=1 study, i.e., it was conducted in only one organization. Therefore the results may not generalize to other assessment centers conducted in other organizations. An obvious direction for future research would be the use of a multi-organizational design examining the ratings of several assessment centers. Certainly our confidence in the generalizability of the technique would be enhanced should future data corroborate the present findings.

One of the shortcomings of this study was the unavailability of more black and female assessors. Schmitt & Hill (1977) suggested that the race and sex composition of assessee groups could affect assessment ratings. More research needs to be directed toward understanding what impact sex and race have in both participant and assessor groupings. Perhaps future studies can shed more light on this issue.

A second shortcoming was that in some cases, assessors changed their initial ratings of a candidate as a result of the team discussion. Thus the relationship between initial ratings and team meetings was obscured. Very little research has focused on the function and value of the team meetings. Future efforts might be directed at such questions as:

- (1) Is the team meeting necessary?
- (2) What is the relationship of team scores to individual scores?

(3) What is the optimal number of team members?

(4) Does the use of racially and sexually integrated teams help to control or eliminate race and/or sex-linked biases?

And finally, scant attention has been paid to situational variables that surround assessment center testing. For example, Samuda (1975) contended that blacks score lower on tests than whites because of heightened test anxieties and lowered feeling of self-worth. Virtually no research has been directed toward exploring how these factors effect center performance for either whites or blacks. Future studies in this area would certainly be informative.

CHAPTER V

SUMMARY

This investigation was undertaken to determine the possibility of discrimination in assessment center ratings due to race or sex-linked biases, as a function of the race and sex of the center participants or the race of the assessors. More specifically, the following questions were examined:

1. Are there systematic differences in the overall assessment scores of participants between all white assessor teams of three versus racially mixed assessor teams of three across participant race, sex, and the various combinations thereof?
2. Overall, are there systematic differences in participant scores due to the assessor's race or the participant's race or sex?
3. Are there systematic differences in performance ratings across assessment exercises as a function of the participant's race or sex?

The 256 Ss in this study are employed in a large southeastern utility company and were candidates for a management development and training program. There were 68 black males, 30 black females, 68 white males, and 90 white females. The assessor group consisted of 23 managers, 18 were white and 5 were black.

Ss participated in a one-day assessment center in which they were rated on three simulation exercises. These included an

In-Basket and Interview, a Leaderless Group Discussion, and a Task-Direction Exercise. The exercises were designed to measure the following nine skill dimensions identified by a job analysis to be critical for job success: (1) Leadership, (2) Perception, (3) Adaptability, (4) Decisiveness (refers to the number of decisions made), (5) Decision-making (refers to the quality of decisions), (6) Organization and Planning, (7) Sensitivity, (8) Written Communications, and (9) Oral Communications.

The data were analyzed using a Three-way Analysis of Variance design. Although the results indicated that there was no discrimination due to sex or race-related biases, they also showed that females scored significantly higher than males on four of the nine dimensions and blacks scored significantly lower than whites on all of the nine dimensions.

The findings were discussed in terms of the implications for the continued use of assessment centers in selection decisions. It was suggested that because the technique appears to be free of biases related to sex or race discrimination that it may be used to promote equal employment opportunities. Three factors appear to be essential to achieve this: (1) A careful job analysis must be conducted upon which to base the exercises; (2) assessors must be trained on how to objectively observe and record behavior; and (3) the use of sexually and racially integrated assessor and participant groups appears to reduce the possibility of adverse impact.

The differences between the performance of black and white participants were discussed. It was suggested that organizations

should take steps to assist blacks in further developing the skills necessary to assume managerial responsibilities.

Several recommendations for future research directions were made:

1. Because this study was conducted in only one organization it would be desirable to replicate the investigation using a multi-organizational design examining the ratings of several assessment centers.

2. More research needs to be directed toward understanding how sex and race affect participant and assessor groups.

3. Of what value are the team meetings? Are they necessary? How do integrated teams help to control or eliminate race and/or sex-linked biases?

4. What affect do situational variables such as test anxiety or self-esteem have on assessment center performance?

REFERENCES

REFERENCES

- Alexander, L. D. An exploratory study of the utilization of assessment center results. Academy of Management Journal, 1979, 22(1), 152-157.
- Arvey, R. D. Fairness in selecting employees. Reading, Mass.: Addison-Wesley, 1979.
- Ash, P., & Kroecker, L. P. Personnel selection, classification, and placement. Annual Review of Psychology, 1975, 26, 481-507.
- Bartol, K. M., Evans, C. L., & Stith, M. T. Black versus white leaders: A comparative review of the literature. Academy of Management Review, April 1978, 3, 293-304.
- Bigoness, W. J. Effects of applicant's sex, race and performance on employer's performance ratings: Some additional findings. Journal of Applied Psychology, 1976, 61(1), 80-84.
- Block, N. J., & Dworkin, G. (Eds.) The IQ controversy: Critical readings. New York: Pantheon Books, 1976.
- Blumenfeld, W. S. Early identification of managerial potential by means of assessment centers. Atlanta Economic Review, December 1971, 35-38.
- Boche, A. Management concerns about assessment centers. In J. L. Moses, & W. C. Byham (Eds.), Applying the assessment center method. New York: Permagon Press, 1977, Chapter 13, 243-260.
- Bray, D. W. The management progress study. American Psychologist, 1964, 19, 419-420.
- Bray, D. W. The assessment center method. In R. L. Craig (Ed.), Training and development handbook: A guide to human resource development. (2nd ed.) New York: McGraw-Hill, 1976, Chapter 16, 1-15.
- Bray, D. W., Campbell, R. J., & Grant, D. L. Formative years in business. New York: Wiley, 1974.
- Brugnoli, G. A., Campion, J. E., & Basen, J. A. Racial bias in the use of work samples for personnel selection. Journal of Applied Psychology, 1979, 64(2), 119-123.
- Byham, W. C. Assessment centers for spotting future managers. Harvard Business Review, 1970, 48(4), 150-160 & appendix.
- Campion, J. E. Work sampling for personnel selection. Journal of Applied Psychology, 1972, 56, 40-44.

- Cash, T. F., Gillen, B., & Burns, D. S. Sexism and "beautyism" in personnel consultant decision-making. Journal of Applied Psychology, 1977, 62(3), 301-310.
- Cohen, S. L. The basis of sex-bias in the job recruitment situation. Human Resource Management, Fall 1976, 15-16, 8-10.
- Cohen, S. L., & Bunker, K. A. Subtle effects of sex-role stereotypes on recruiters' hiring decisions. Journal of Applied Psychology, 1975, 60(5), 566-572.
- Cox, J. A., & Krumboltz, J. D. Racial bias in peer ratings of basic airman. Sociometry, 1958, 21, 292-299.
- DeJung, J. W., & Kaplan, H. Some differential effects of race of rater and ratee on early peer ratings of combat aptitude. Journal of Applied Psychology, 1962, 46, 370-374.
- Development Dimensions International. Development Dimensions 1977-1978 Catalog. Pittsburgh: Author, 1977.
- Dipboye, R. L., Arvey, R. D., & Terpstra, D. E. Sex and physical attractiveness of raters and applicants as determinants of resume evaluations. Journal of Applied Psychology, 1977, 62(3), 288-294.
- Dipboye, R. L., Fromkin, H. L., & Wiback, K. Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. Journal of Applied Psychology, 1975, 60, 39-43.
- Dreger, R. M., & Miller, K. S. Comparative studies of Negroes and whites in the United States. Psychological Bulletin, 1960, 57, 361-402.
- Dreger, R. M., & Miller, K. S. Comparative studies of Negroes and whites in the United States: 1959-1965. Psychological Bulletin Monograph Supplement, 1968, 70(3), Part 2, 1-58.
- Dunnette, M. D. Multiple assessment procedures in identifying and developing managerial talent. In P. McReynolds (Ed.), Advances in psychological assessment. Vol. II. Palo Alto: Science & Behavior Books, 1971.
- Dunnette, M. D., & Borman, W. C. Personnel selection and classification. Annual Review of Psychology, 1979, 30, 477-525.
- Equal Employment Opportunity Commission, U.S. Civil Service Commission, U.S. Department of Justice, & U.S. Department of Labor, 1978. Uniform guidelines on employee selection procedures. Federal Register, August 25, 1978, 43(No. 166), 38290-38315.
- Finkle, R. B. Managerial assessment centers. In M.D. Dunnette (Ed.), Handbook of industrial and organizational psychology. Chicago: Rand-McNally, 1976, 861-888.

- Goldberg, P. A. Are women prejudiced against women? Transaction, 1968, 5, 28-30.
- Greenwood, J. M., & McNamara, W. J. Interrater reliability in situational tests. Journal of Applied Psychology, 1967, 51, 101-106.
- Guion, R. M. Personnel testing. New York: McGraw-Hill, 1965.
- Hamner, W., Kim, J. S., Baird, L., & Bigoness, W. Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. Journal of Applied Psychology, 1974, 59, 705-711.
- Helwig, J. T., & Council, K. A. (Eds.) The SAS user's guide, 1979 edition. Raleigh, N.C.: SAS Institute Inc., 1979.
- Huck, J. R. Assessment centers: A review of the external and internal validities. Personnel Psychology, 1973, 26(2), 191-212.
- Huck, J. R. The research base. In J. L. Moses, & W. C. Byham (Eds.), Applying the assessment center method. New York: Permagon Press, 1977, Chapter 14, 261-291.
- Huck, J. R., & Bray, D. W. Management assessment center evaluations and subsequent job performance of white and black females. Personnel Psychology, 1976, 29(1), 13-30.
- Jaffee, C. L., & Sefcik, J. T., Jr. What is an assessment center? Personnel Administrator, 1980, 25(2), 40-43.
- Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Brooks Cole, 1968.
- Klimoski, R. J., & Strickland, W. J. Assessment centers-valid or merely prescient. Personnel Psychology, 1977, 30, 353-361.
- Landy, F. J., & Trumbo, D. A. Psychology of work behavior. Homewood, Ill.: The Dorsey Press, 1976.
- Lerner, B. The war on testing: Detroit Edison in perspective. Personnel Psychology, 1980, 33, 11-16.
- Moses, J. L. The development of an assessment center for the early identification of supervisory potential. Personnel Psychology, 1973, 26, 569-580.
- Moses, J. L. (Chm.) Standards and ethical considerations for assessment center operations. Task Force on Development of Assessment Center Standards. Quebec, Canada: Third International Congress on the Assessment Center Method, May 1975.
- Moses, J. L., & Boehm, V. R. Relationship of assessment center performance to management progress of women. Journal of Applied Psychology, 1975, 60(4), 527-529.

- Nason, R. W. The dilemma of black mobility in management: Discrimination still a problem. In D. L. Ford, Jr. (Ed.), Readings in minority-group relations. La Jolla, Calif.: University Associates, 1976, 297-314.
- O'Leary, L. R. Fair employment, sound psychometric practice and reality: A dilemma and partial solution. American Psychologist, 1973, 28(2), 147-150.
- Pheterson, G. I., Kiesler, S. B., & Goldberg, P. A. Evaluation of the performance of women as a function of their sex, achievement, and personal history. Journal of Personality and Social Psychology, 1971, 19, 114-118.
- 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.
- Rosen, B., & Jerdee, T. H. The influence of sex-role stereotypes on evaluations of male and female supervisory behavior. Journal of Applied Psychology, 1973, 57(1), 44-48.
- Rosen, B., & Jerdee, T. H. Effects of applicant's sex and difficulty of job evaluations of candidates for managerial positions. Journal of Applied Psychology, 1974a, 59, 511-512.
- Rosen, B., & Jerdee, T. H. Influence of sex-role stereotype on personnel decisions. Journal of Applied Psychology, 1974b, 59, 9-14.
- Rosen, B., & Jerdee, T. H. Sex stereotyping in the executive suite. Harvard Business Review, 1974c, 52, 45-58.
- Samuda, R. J. Psychological testing of American minorities: Issues and consequences. New York: Dodd, Mead & Company, 1975.
- Schein, V. E. The relationship between sex-role stereotypes and requisite management characteristics. Journal of Applied Psychology, 1973, 57, 95-100.
- Schein, V. E. Relationship between sex-role stereotypes and requisite management characteristics among female managers. Journal of Applied Psychology, 1975, 60(3), 340-344.
- 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. Interrater agreement in dimensionality and combination of assessment center judgements. Journal of Applied Psychology, 1977, 62(2), 171-176.
- 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.

- Schmitt, N., & Lappin, M. Race and sex as determinants of the mean and variance of performance ratings. Journal of Applied Psychology, 1980, 65(4), 428-435.
- Shaw, E. A. Differential impact of negative stereotyping in employee selection. Personnel Psychology, 1972, 25, 333-338.
- Shuey, A. M. The testing of Negro intelligence. (2nd ed.) New York: Social Science Press, 1966.
- Siegel, S. Nonparametric statistics for the behavioral sciences. New York: McGraw-Hill, 1956.
- Terborg, J. R., & Ilgen, D. R. A theoretical approach to sex discrimination in traditionally masculine occupations. Organizational Behavior and Human Performance, 1975, 13, 352-376.
- Vinson, E., & Mitchell, T. R. Differences in motivational predictors and criterion measures for black and white employees. Proceedings of the 35th annual meeting of the Academy of Management, New Orleans, August 1975.
- Wernimont, P. F., & Campbell, J. P. Signs, samples, and criteria. Journal of Applied Psychology, 1968, 52, 372-376.
- Wexley, K. N., & Nemeroff, W. F. Effects of racial prejudice, race of applicant and biographical similarity on interviewer evaluations of job applicants. Journal of Social and Behavioral Science, 1974, 20(3), 66-78.
- Wiener, Y., & Schneiderman, M. L. Use of job information as a criterion in employment decisions of interviewers. Journal of Applied Psychology, 1974, 59(6), 699-704.
- Winer, B. J. Statistical principles in experimental design. (2nd ed.) New York: McGraw-Hill, 1971.
- Wollowick, H. B., & McNamara, W. J. Relationship of the components of an assessment center to management success. Journal of Applied Psychology, 1969, 53, 348-352.

APPENDIXES

APPENDIX A

NOMINATION FORM--MANAGEMENT TRAINING AND DEVELOPMENT PROGRAM

NAME: _____ JOB TITLE: _____

SCHEDULE AND GRADE: _____ OFFICE/DIVISION: _____

BRANCH: _____

FORMAL EDUCATION: _____

OTHER TRAINING/EDUCATION: _____

BRIEF WORK HISTORY: _____

CRITERIA: Describe the individual's potential/performance relative to each criterion using behavioral examples. (Descriptions of each criterion, and examples relative to each follow.)

1. Interpersonal competence - The ability to get along with other people on the job. Includes such behavior as putting others at ease in stressful situations; minimizing differences between self and others so that conflicts are managed.

2. Flexibility, broad perspective - The ability to understand and adjust to change. Includes such behavior as listening to points of view of others.

3. Initiative - The tendency to identify what needs to be done and do it without having to be told. Includes such behavior as seeking out new assignments while not letting present assignments suffer.

4. Creativity, innovativeness - The ability to look beyond the obvious. Includes such behavior as developing new approaches to problems.

5. Leadership - The ability to get others to perform while minimizing resistance and resentment. Includes such behaviors as delegating responsibility, encouraging teamwork, and supporting subordinates.

6. Problem-solving - The ability to reason things out. Includes such behavior as systematically thinking through complex problems or issues to a logical solution.

7. Decision-making - The ability to choose among various alternatives and act on the choice. Includes such behavior as selecting one of several alternatives which is subsequently borne out to be sound.

Summary Comments:

APPENDIX B¹

EXAMPLES OF ASSESSMENT EXERCISES

I. The In-basket Exercise

An example of the instructions that a candidate received are shown below. In addition to these instructions a candidate would receive a copy of an hypothetical organizational chart and a stack of memos and letters as might be found on the manager's desk.

Instructions

For the purpose of this exercise, you are to consider yourself Lee Baldwin, a Service Manager for the Consolidated Light and Power Company of the State of New York. Consolidated Light and Power Company is responsible for the production and distribution of electrical energy for the entire State of New York, excluding New York City, Long Island, and Westchester County. You have just been transferred to the Midwestern Division Headquarters where you have been appointed County Service Manager for Chemung County (one of four counties handled by this division).

Today is Sunday, May 30th, and you have just arrived at your new office. You were appointed on very short notice because your predecessor, Mr. Creech, died suddenly on Wednesday, May 26th. It is early in the morning and you are alone in your office without access to files because they are locked. You cannot use the telephone. You have come in to take care of any matters which might require your attention before Monday, June 7. You must leave your office in exactly one (1) hour to catch a train. Since this is the Memorial Day weekend there are a limited number of trains running. You will not be in the office until June 7th because the division where you have been working requires that you finish an important report needed for an Executive Committee meeting on Monday, June 7th. You will be spending your holiday and the following days completing this massive project. You will not have time to work on anything else while you are away.

On your desk you have found the following materials which have been gathered by your secretary, Joan Gore. In the next hour you are to deal with the materials in any manner you see fit. Prepare any letters, draft any memos, take any actions which you deem appropriate. Any decision or action you take should be indicated in writing.

¹Materials appearing in Appendixes B, C, and D were prepared by Assessment Designs, Inc., and are used with permission.

II. The Leaderless Group Discussion - Bioconversion Energy Problem.

Shown below are the general instructions a candidate received for the Leaderless Group Discussion Exercise. In addition, he/she also received background information on bioconversion and information regarding which location he/she would be representing.

Instructions

You are currently working on a special government project for the Energy Resource Management Agency (ERMA). The project is part of a long-range plan to utilize the oceans as a food and energy resource. Scientists are increasingly looking to the sea for the answers to the current energy and food shortages. You and your other committee members have been appointed to assist in this project because of your work in the area of energy and because of the geographical area in which you live.

Your committee has been studying solar-energy bioconversion. Bioconversion refers to the process in which biomass or fast growing plants trap the energy from solar photons and store it in ducts. Clean fuels, such as methane gas and methyl alcohol, can be added to gasoline or used on their own in internal combustion engines. These chemicals can also be used to make synthetic gasoline-type products. In addition, the plant biomass materials could be burned to produce steam or generate electricity.

You and the other committee members each represent your own geographical area. All of the represented areas have been chosen as possible sites for a kelp farm. You have compiled data on the relative advantages and disadvantages of locating and developing a commercial marine kelp farm in your state for the purpose of bioconversion to clean fuels and other products. These six sites were shown to be possible locations for kelp farms from an engineering standpoint. They are the only sites that were judged to be suitable for kelp farm development for topographical and climatic reasons. The task assigned to your committee is to decide which three of these areas are best suited for the development of a marine kelp farm and to list them as first, second, and third choices by considering all of the pertinent available data supplied by the committee members.

Your choices will be presented to ERMA which will make the final selection and will seek additional funding from Congress for the project development. It is therefore important to your committee that the best site be recommended so as to maximize the chances that the Congress will approve additional funds for the project. At the same time, since you each will continue to represent the site that you have collected data on, it would be of great advantage to your career if your site was chosen for the project.

You will have 20 minutes to look over the data on your site and prepare arguments for your position. At the end of that time, the

group will discuss the problem together for 45 minutes and come to a decision on which sites should be considered as first, second, and third choice.

Do not be restricted by governmental or departmental regulations or policies in making your recommendations.

III. Task Direction Problem - Casil Chemical Exercise

Below is an example of the instructions a candidate received. In addition to these instructions each candidate received a packet of information with all the data necessary to solve the problem.

Instructions

For the purpose of this exercise, you are to assume that you are Terry Sorinson, Supervisor of the Printing Department for the Casil Chemical Company, a manufacturer of pharmaceutical and agricultural chemical products.

The Printing Department is responsible for a large variety of printing work. This includes colored printing of various product packages, printing of the labels and shipping containers, specialized instructions and pamphlets for each chemical product produced by Casil Chemical, and advertising pamphlets which are distributed to your current and prospective retailers.

Today is Friday, August 15, and you have just received word that Casil Chemical products are currently selling at a rate which is exceeding by 7 percent the projected sales for this time of year. Although this is without a doubt good news, it also creates a number of problems. Current stocks of Casil Chemical agricultural products in retail stores and warehouses are declining at a time when products are needed to be stockpiled for the upcoming fall season demand (September - November). This unexpected 7 percent increase in sales is likely to continue through the year. This increase, on top of the normal sales during the end of the year, is likely to completely deplete current inventories and result in shortages at the retail level and a resulting loss of profit potential.

The management of Casil Chemical has decided to immediately add another shift of production employees and has informed all departments of the necessity of overtime work hours and an extra effort being required in order to ensure that adequate supplies are available through the rush. For the Printing Department, this will require large amounts of overtime for at least the next week. Since today is Friday, you must decide today who will work overtime, at what equipment, and how many hours. Furthermore, since you must allow sufficient time to inform your employees, you must make these decisions within the next 40 minutes.

Since this is a large task to complete in a short period of time you will have an assistant, Lynn Larson, to work with you.

The following information is to be used to help you make your decisions. At the end of the 40-minute period, you are to have prepared the schedule for the following week. This schedule is to include both the 40-hour workweek and all overtime hours. On this

schedule you are also to indicate the total number of hours each employee is to work and on what equipment. If an employee is to work on more than one type of machine or is to do more than one printing order, you are to specify what machine or what printing order is to be done first.

Lynn Larson has just been assigned to your department as a trainee. Note Lynn's performance as you work together in this exercise, because when the exercise is completed you will have fifteen minutes in which to meet with Lynn to discuss any work behavior that you consider important.

The last task in this exercise will be to write a short summary of what was accomplished in the main task and a brief summary of the performance evaluation session with Lynn. Both reports should be concise since only fifteen minutes is allotted for this part of the task.

If you have no questions you may begin. Remember, at the end of the 40 minutes you must have prepared the schedule for next week.

APPENDIX C

EXERCISE REPORT FORM, SKILL DEFINITIONS AND MATRIX WEIGHTING FORM

Exercise Report Form

ASSESSOR: _____ DATE: _____

CANDIDATE: _____ EXERCISE: _____

Using the rating key provided below, rate the participant on each of the following skills based on what you have seen him/her do only in this exercise.

Rating Key

- | | |
|-----------------------------|-----------------------------|
| 7 - Outstanding | 3 - Below satisfactory |
| 6 - Well above satisfactory | 2 - Well below satisfactory |
| 5 - Above satisfactory | 1 - Poor |
| 4 - Satisfactory | 0 - Not observed |

For any 100 participants you might observe, the following distribution of ratings is likely to occur: 5 percent of the participants are likely to be rated a "7"; 10 percent, a "6"; 20 percent, a "5"; 30 percent, a "4"; 20 percent, a "3"; 10 percent, a "2", and only 5 percent, a "1". Remember, these percentages are by no means binding, and you may consider several participants to perform in an outstanding manner on most skills; yet, when considering the entire group of participants, the full range of skills levels should be observable. The Skill/Exercise Matrix on the following page should help you in this evaluation task.

Exercise Summary (for comments on unique or extenuating circumstances only)

Leadership: Rating ____; Ability to take charge - to direct and coordinate the activities of others; to maintain control of situations and others; to achieve results through delegation and follow-up.

Sensitivity: Rating ____; Ability to be sensitive to the needs and feelings of others; to develop rapport and trust; to accept interpersonal differences; to deal effectively with others regardless of level or status.

Perception: Rating ____; Ability to identify, assimilate and comprehend the critical elements of a situation; to extract and interpret implications of courses of action; to attend to details of a problem (includes both data and people related issues).

Decision-Making: Rating ____; Ability to use logical and sound judgment in choosing a particular course of action (this refers to the quality as opposed to the quantity of decisions).

Decisiveness: Rating ____; Ability to take action when called upon to do so, (quantity of decisions); and to defend decisions when challenged.

Organizing and Planning: Rating ____; Ability to systematically structure tasks, plans and objectives; to establish priorities and set goals, to classify and categorize information.

Adaptability: Rating ____; Ability to alter normal posture with presentation of additional information; to appropriately change courses of action dictated by changes in the situation; to have the ability to behave in more than one way in a given situation; to adapt to stressful situations.

Oral Communication: Rating ____; Ability to effectively and clearly present and express information orally, in both formal and informal situations.

Written Communication: Rating ____; Ability to present and express information effectively and clearly through written means.

SKILL MATRIX WEIGHTING FORM

PARTICIPANT: _____ ASSESSOR: _____

SKILL	IN-BASKET I, II or III	GROUP DISCUSSION	TASK DIRECTION	OVERALL
Leadership	XX	XX	XXX	
Sensitivity	X	XX	XX	
Perception	XXX	XX	XX	
Decision-Making	XX	XX	XX	
Decisiveness	X	XX	XX	
Organizing & Planning	XXX	XX	XX	
Adaptability	N/A	XX	XX	
Oral Communication	X	XX	XXX	
Written Communication	XX	N/A	XX	

Weighting:

XXX - Very Strongly Measured

XX - Strongly Measured

S - Measured

N/A - Not Applicable

APPENDIX D

PARTICIPANT'S AND ASSESSOR'S SCHEDULES

Shown below are samples of the daily schedules for assessment center participants and assessors.

I. PARTICIPANT SCHEDULE

Participant #1

<u>Time</u>	<u>Activity</u>	<u>Room</u>
9:00- 9:15	Orientation	1
9:20- 9:40	LGD Review	1
9:40-10:25	LGD	1
10:30-11:45	In-Basket Review	2
12:05-12:25	In-Basket Interview	2
12:25-12:55	Lunch	2
12:55- 1:40	Task Direction/Scheduling	2
1:40- 1:55	Task Direction/Employee Counseling	2
1:55- 2:10	Task Direction/Written Report	2

II. ASSESSOR'S SCHEDULE

Assessor #6

<u>Time</u>	<u>Activity</u>	<u>Participant</u>	<u>Room</u>
9:20-10:05	Task Direction/Scheduling	7	7
10:05-10:20	Task Direction/Employee Counseling	7	7
11:00-11:45	LGD	9, 12	1
11:50-12:35	Task Direction/Scheduling	10	7
12:35-12:50	Task Direction/Employee Counseling	10	7
1:05-1:25	Assessor's In-Basket Review	8	6
1:25-1:45	In-Basket Interview	8	5
1:50-2:10	Assessor's In-Basket Review	11	8
2:10-2:30	In-Basket Interview	11	8

VITA

Mark Joel Friedman was born in New York City on October 13, 1946. He was raised and educated in the Bronx, graduating from Evander Childs High School in 1964. He then entered the City College of the City University of New York, where he graduated with a Bachelor of Arts degree in Psychology in 1969.

In the same year he entered The University of Detroit to pursue a Master of Arts degree in Industrial Psychology, which he completed in August 1974. While a student at Detroit, he decided to continue his education and so entered The University of Tennessee, Knoxville, in pursuit of the Ph.D. degree in Psychology. During the summer of 1973, Mr. Friedman served a three month internship with Elanco Products Company - A Division of Eli Lilly and Company in Indianapolis, Indiana. He worked in the Department of Personnel, doing management training and consulting.

Having completed his classwork at The University of Tennessee in the summer of 1975, Mr. Friedman assumed the post of Assistant Professor of Behavioral Science and Business Administration at Bloomfield College, in Bloomfield, New Jersey. He subsequently accepted a position with RCA Global Communications, in New York City, in January 1977, being soon promoted to Manager of Organizational Development and Training. However, still wanting to complete the final requirements for the Ph.D. degree, he returned to Knoxville in October 1978, accepting a position with the Tennessee Valley Authority as an Internal Development Consultant.

Mr. Friedman received the Doctor of Philosophy degree in Psychology in December 1980.