

INFORMATION TO USERS

This reproduction was made from a copy of a document sent to us for microfilming. While the most advanced technology has been used to photograph and reproduce this document, the quality of the reproduction is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help clarify markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure complete continuity.
2. When an image on the film is obliterated with a round black mark, it is an indication of either blurred copy because of movement during exposure, duplicate copy, or copyrighted materials that should not have been filmed. For blurred pages, a good image of the page can be found in the adjacent frame. If copyrighted materials were deleted, a target note will appear listing the pages in the adjacent frame.
3. When a map, drawing or chart, etc., is part of the material being photographed, a definite method of "sectioning" the material has been followed. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.
4. For illustrations that cannot be satisfactorily reproduced by xerographic means, photographic prints can be purchased at additional cost and inserted into your xerographic copy. These prints are available upon request from the Dissertations Customer Services Department.
5. Some pages in any document may have indistinct print. In all cases the best available copy has been filmed.

**University
Microfilms
International**

300 N. Zeeb Road
Ann Arbor, MI 48106

8229971

Russell, Craig John

AN ANALYSIS OF A DECISION PROCESS IN AN ASSESSMENT CENTER

The University of Iowa

PH.D. 1982

**University
Microfilms
International** 300 N. Zeeb Road, Ann Arbor, MI 48106

Copyright 1982

by

Russell, Craig John

All Rights Reserved

PLEASE NOTE:

In all cases this material has been filmed in the best possible way from the available copy. Problems encountered with this document have been identified here with a check mark .

1. Glossy photographs or pages _____
2. Colored illustrations, paper or print _____
3. Photographs with dark background _____
4. Illustrations are poor copy _____
5. Pages with black marks, not original copy _____
6. Print shows through as there is text on both sides of page _____
7. Indistinct, broken or small print on several pages
8. Print exceeds margin requirements _____
9. Tightly bound copy with print lost in spine _____
10. Computer printout pages with indistinct print _____
11. Page(s) _____ lacking when material received, and not available from school or author.
12. Page(s) _____ seem to be missing in numbering only as text follows.
13. Two pages numbered _____. Text follows.
14. Curling and wrinkled pages _____
15. Other _____

University
Microfilms
International

AN ANALYSIS OF A DECISION PROCESS
IN AN ASSESSMENT CENTER

by

Craig John Russell

A thesis submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Business Administration
in the Graduate College of
The University of Iowa

July, 1982

Thesis supervisor: Associate Professor Duane Thompson

Graduate College
The University of Iowa
Iowa City, Iowa

CERTIFICATE OF APPROVAL

PH. D. THESIS

This is to certify that the Ph.D. thesis of

Craig John Russell

has been approved by the Examining Committee
for the thesis requirement for the Doctor of
Philosophy degree in Business Administration
at the July, 1982 graduation.

Thesis committee: *Duane E. Thompson*
Thesis supervisor

T. Anne Cleaves
Member

Michael K. Mount
Member

John P. Lewis
Member

C. R. Klassen
Member

Copyright by
CRAIG JOHN RUSSELL
1982
All Rights Reserved

TABLE OF CONTENTS

CHAPTER	Page
I. INTRODUCTION	1
Purpose	8
Definition of Key Terms	8
Limitations	9
II. REVIEW OF RELATED RESEARCH	11
Criterion Validity Research	11
Literature Related to Assessor Cognitive Processes	12
Perception of Assessment Dimensions	13
Group Influences	15
Future Directions	15
Integration of the Dimensions	16
Possibility of a Nonlinear Model	22
Future Directions	23
Use of the Overall Rating	24
Future Directions	26
Sex Effects	27
Future Directions	29
Summary	30
III. METHODOLOGY	32
The Assessment Center	32
Dimensions	33
Techniques	36
Assesseees and Assessment Staff	38
Rating Procedures	39
Data Collection	42
Method	43
Research Question 1	43
How well can the OAR be predicted by the eighteen dimensions?	44
Is there a simple structure underlying the eighteen dimensions corresponding to the a priori categories?	45

	Page
Do the assessors arrive at the OAR in the way they say they do?	45
Research Question 2	46
Effects of sex of the assessee	47
Effects of male to female ratio of assessor group	47
IV. RESULTS	48
Research Question 1	48
Double Cross Validation	48
How well can the OAR be predicted by the eighteen dimensions?	49
Is there a simple structure underlying the eighteen dimensions?	51
Do assessors arrive at the OAR in the way they say they do?	54
Research Question 2	57
Regressions Involving Assessee Sex	58
Assessor Group Sexual Composition	60
V. SUMMARY AND DISCUSSION	62
Discussion of Research Question 1	63
Predictive Power	63
Underlying Structure of the Eighteen Dimensions	65
Use of the Subscale Scores	67
Discussion of Research Question 2	69
Assessee Sex Effects	69
Assessor Group Sex Effects	73
VI. LIMITATIONS AND IMPLICATIONS	77
Limitations	77
Implications	79
Future Research	79
Implications for Practice	81
APPENDIX A. ASSESSMENT CENTER DEFINED	83
APPENDIX B. STAFF-ASSESSOR JOB DESCRIPTION	86
Job Summary	86
Job Duties and Responsibilities	86
Scope and Nature of Supervision	90
APPENDIX C. STAFF DIRECTOR JOB DESCRIPTION	91

	Page
Job Summary	91
Job Duties and Responsibilities	91
APPENDIX D. TABLES	95
REFERENCES	114

LIST OF TABLES

Table	Page
1. Sample Representation of Total Number Assessed by Overall Rating and Occurance of Promotion* . . .	96
2. Dimension Sources	97
3. Correlation Matrix of Consensus Ratings	98
4. Multiple Regression Analyses: Dependent Variable = Overall Assessment Rating	99
5. Principal Component Analysis varimax rotated Factor Pattern matrix: Three Factor Rotation*	100
6. Principal Components Analysis Oblimin Rotated Factor Pattern Matrix: Three Factor Rotation (delta=0)*	101
7. Principal Components Analysis Oblimin Rotated Factor Pattern Matrix: Three Factor Rotation (delta=-1)*	102
8. Subscale Loadings from Common Factor Analyses	103
9. Principal Components Analysis Varimax Rotated Factor Pattern Matrix: Four Factor Rotation*	104
10. Principal Components Analysis Oblimin Rotated Factor Pattern Matrix: Four Factor Rotation (delta=0)*	105
11. Principal Components Analysis Oblimin Rotated Factor Pattern Matrix: Four Factor Rotation (delta=-1)*	106
12. Regression of Overall Assessment Rating Onto Category Subscale Scores, 2-Way Interactions, Dummy Coded Assessee Sex Variable, and Sex by Subscale Score Interactions	107

13.	Regression of Overall Assessment Rating Onto Factor Subscale Scores, 2-Way Interactions, Dummy Coded Assessee Sex Variable, and Sex by Subscale Score Interactions	108
14.	Frequencies: Sex x OAR	109
15.	Predicted OAR Means for Males and Females by Variables Included in Regression Equation and Sex of Subsample the Regression Equation Was Derived From	110
16.	Regression of Overall Assessment Rating Onto Category Subscale Scores, 2-Way Interactions, Assessor Group Sex Composition (GSEX), and GSEX by Subscale Score Interactions	111
17.	Regression of Overall Assessment Rating Onto Factor Subscale Scores, 2-Way Interactions, Assessor Group Sex Composition (GSEX), and GSEX by Subscale Score Interactions	112
18.	Predicted Means and Standard Deviations of OAR For All Levels of GSEX When OAR is Regressed Onto Factor Subscale Scores For Each GSEX Subgroup .	113

1

CHAPTER I
INTRODUCTION

Personnel selection decisions have used multiple assessment procedures for almost forty years (O.S.S. Assessment Staff, 1948). One of these procedures, the assessment center, first appeared in the United States at Station S of the Office of Strategic Services (O.S.S.) during World War II. This center was used to select personnel for assignments behind enemy lines. Its theoretical foundations and actual implementation were largely the work of Henry Murray (1938). Over the years many variations have been developed (Bray, 1982), but recently, a common definition has been established. The Task Force on Development of Assessment Center Standards (1975, cited in Moses & Byham, 1977) developed the following seven points as the minimal requirements of an assessment center:

1. Multiple assessment techniques must be used. At least one of these techniques must be a simulation.
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 pre-tested prior to use to insure that the techniques provide reliable, objective, and relevant behavioral information for 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. (pp. 304-305 cited in Moses & Byham, 1977, and reprinted in Appendix A)

Today it is estimated that over 1,000 organizations employ assessment centers in the evaluation of managerial skills and abilities. The same source estimates over 50,000 people are being assessed each year (Byham, 1977).

A vast amount of research has been conducted on assessment centers. This research has examined a variety of relationships among three categories of variables. These categories are:

1. the skill and ability dimensions evaluated by assessment center activities (see Murray, 1938, for the original explication of situational tests and exercises used);
2. the overall rating (OAR) of the candidates' capacity to perform the job; and
3. the various criterion measures of performance.

Major literature reviews examine the relationship of the dimensions to criteria and the OAR to criteria (Cohen, Moses, and Byham, 1974; Dunnette, 1971; Huck, 1973; Howard, 1974; MacKinnon, 1975; Finkle, 1976; Norton, 1977; and Norton and Edinger, 1978). One finding is reported throughout this literature. Managerial personnel selection decisions reached through the use of assessment center procedures are consistently related to a number of subsequent job related criteria.

Another stream of research was promoted by the Civil Rights Act of 1964. Since that Act, professionally developed tests used in making employment decisions have come under judicial scrutiny. A review of legal cases and opinions addressing the use of assessment centers indicates tentative judicial support (Byham, 1979).

Validity is the evaluation of the appropriateness of an inference from a test score or measurement procedure (American Psychological Association, 1974; American Psychological Association Division of Industrial and Organizational Psychology, 1980; Guion, 1980). At least three types of inferences can be made concerning the OAR and/or the procedures used to arrive at it.

First, Bray and Campbell (1968) and Huck and Bray (1976) have argued that the OAR is a direct reflection of the managerial skills and abilities exhibited by a job candidate in the assessment center. Consequently, one inference that can be made from the OAR is that it accurately measures the construct 'managerial skill and ability' (i.e., that it is a construct valid measure of managerial skill and ability).

Second, Huck (1974) and Howard (1974) have argued that one of the major contributions of assessment centers is the use of multiple procedures (e.g. situational exercises and simulations) which provide a representative sample of behaviors necessary for managerial performance. Consequently, a second inference that can be made is that the behaviors sampled in the assessment center are representative of those required for effective managerial performance.

Finally, as noted above, research has consistently demonstrated a strong empirical relationship between the OAR

and various criterion measures of managerial performance (e.g. subsequent number of promotions, salary increases, and supervisor's subjective performance evaluations). The inference evaluated in this research is that the OAR is a predictor of future managerial performance.

The way in which assessors use the information gathered in an assessment center to arrive at an OAR is crucial to the evaluation of these inferences. The performance appraisal literature has long been concerned with the cognitive processes of evaluators (Smith, 1976). Levy (1960, cited in Dunnette, 1967) found that department heads who were highly rated by their superiors valued different things in their subordinates than less highly rated department heads. Assessment centers are making an inference regarding the same construct as performance appraisals, the only difference being that the individuals are candidates for jobs and not incumbents. An understanding of the way in which assessors weigh and combine the information gathered in an assessment center is a prerequisite to the inference that the OAR reflects a candidates' managerial skill and ability.

Guion (1974, 1977a, 1977b, 1978a, 1978b, 1980) has argued that the evaluation of the inference that a behavioral sample is representative of a job domain (e.g. that an assessment center representatively samples behavior required in the job domain of manager) is addressed by two

questions. First, are the stimuli used to elicit the behavior in the sample representative (deficient/contaminated) of the stimuli found in the job domain? Second, are the frequencies and types of behaviors available to an individual in the sample representative of the array of behaviors available in the job domain? In addition, Guion (1980) stated:

If the inference to be drawn from a score on a content sample is to be an inference about performance on an actual job, then it is drawn at the end of a series of inferential steps, any one of which can be a serious misstep. The most serious misstep may occur in defining the scoring system. (p. 392)

Job analysis would appear to be the technique most appropriate in establishing the congruence of the sample stimuli and response opportunities to the job domain of interest (Prien, 1977). However, the scoring system used in assessment centers is the clinical judgments of assessors. Consequently, prior to the use of inferences concerning the content validity of assessment centers (c.f. Norton, 1977), an understanding of the way in which assessors 'score' the sample of behavior observed in an assessment center is needed.

Finally, an understanding of assessor cognitive processes would be useful in evaluating the predictive validity of assessment centers. As noted above, a large body of research indicates the CAR is a consistent predictor of subsequent performance criteria. However, Klimoski and Strict-

land (1977) have argued that there has "been a curious homogeneity in the criteria used" in assessment center validity research (p. 354). They further invoke a warning made by Wallace (1974) about the difference between prediction of an individual's job performance and prediction of what people say about an individual's job performance. To date, only one assessment center validity study has used a criterion of behavioral observations (Bray & Campbell, 1968). This ability of assessment centers to predict what people say about an individual's job performance (i.e. as reflected in salary growth, rate of promotion, increase in managerial responsibilities, performance interviews, rankings and ratings of overall performance, and ratings of potential for advancement) is very different from the ability to predict subsequent managerial behavior. The identification of candidates who can elicit high marks from assessors and performance appraisors is not the same as the identification of candidates who can engage in behaviors required of high job performance. An understanding of the cognitive processes employed by assessors in arriving at an OAR would provide insight into whether it was a candidate's career ladder climbing skills or managerial skills that led to that specific CAR.

Hence, an understanding of the cognitive processes used in the clinical judgments of assessors is an important pre-

requisite to the inferences that have been drawn from assessment center outcomes.

Purpose

The present study is directed at obtaining an understanding of the cognitive processes employed by the assessor in arriving at an overall rating of candidates' managerial potential. Specifically, the following two research questions shall be evaluated:

1. Research Question 1: How is the information gathered in an assessment center used by assessors to arrive at an overall rating for a particular assessee?
2. Research Question 2: How does the sexual composition of the assessor group and the sex of the assessee moderate the relationship delineated in Research Question 1 above?

Definition of Key Terms

A number of key terms to be used throughout this document shall be defined. This list is not meant to be all inclusive. As additional terms arise in subsequent chapters they shall be defined.

1. Overall assessment rating (OAR) refers to the individual assessor's rating of the likelihood that a candidate has the requisite skills to successfully perform the job of entry level manager.

2. Dimension shall be used to describe any one of the eighteen sets of behaviors, skills, or attributes being evaluated in the assessment center.
3. Category shall be used to describe an a priori or administrative grouping of dimensions.
4. Factor shall be used to describe an empirically derived grouping of dimensions.
5. Staff Assessor shall refer to the junior members of the assessment staff (see Appendix B for a job description).
6. Staff Director shall refer to the senior member of the assessment staff (see Appendix C for a job description).
7. Assessee shall refer to the candidate for a managerial position under evaluation.

Limitations

There are at least three limitations in the current research. First, only part of the series of judgments made with regard to assessment center performance is being examined. The relevant stages between the assessee's performance and the assignment of an overall rating are:

1. The attributional process by which an assessor arrives at a rating of the assessee's behavior on each of the eighteen dimensions.

2. The group decision making activities used to arrive at a consensus among the assessors concerning each dimension's rating.
3. The integration of the eighteen consensus ratings on the dimensions into an overall rating by each individual assessor.
4. The group decision making activities used to arrive at a consensus among the assessors concerning the overall rating.

This study shall examine data resulting from a combination of steps three and four above. However, other decision points in the assessment center have to be understood for a complete picture of assessors' cognitive processes.

Second, the results of the analyses presented here, at best, will generalize only to assessors operating identical centers and given identical training.

Finally, although the analyses presented herein are an attempt at obtaining an isomorphic representation of assessor decision processes, the success of such an effort can never be known with certainty. At best, one can only hope that the results of this investigation will modify our prior probabilities as to which is the true set of cognitive processes involved.

CHAPTER II
REVIEW OF RELATED RESEARCH

Criterion Validity Research

A number of authors have reviewed research examining the use of assessment centers in selection (Cohen, Moses, and Byham, 1974; Dunnette, 1971; Huck, 1973; Howard, 1974; MacKinnon, 1975; Finkle, 1976; Norton, 1977; and Norton and Edinger, 1978). Cohen et. al. (1974) have reviewed the validity studies performed on assessment center selection techniques published between 1956 and September, 1972. They report a median correlation between assessment center OARs and number of subsequent promotions of .40 and a median correlation of .63 between OARs and subsequent manager's ratings of promotion potential. They conclude that the OAR has been a consistently valid predictor with predictive accuracy generally being highest for ratings of job potential followed by job progress and job performance respectively. They also point out that the OAR is consistently a better predictor than paper-and-pencil tests and other traditional methods of assessment. The other seven reviews of the literature cited above arrive at almost identical conclusions.

Concurrent (Byham and Wettengel, 1974; Thoreson and Jaffee, 1973; and Worbois, 1975) and predictive (Huck and Bray, 1976; and Mitchel, 1975) validity studies performed since the Cohen et. al. review have all yielded consistent positive results. However, as noted in chapter 1, Klimoski and Strickland's (1977) warning concerning the "curious homogeneity in the criteria" (p. 354) used in previous research has yet to be addressed. Until such time as it is, the interpretation of the impressive array of criterion validity studies is in doubt.

Literature Related to Assessor Cognitive Processes

In light of the number of inferences that can be made from an assessment center, the value in the examination of the judgment processes in assessment centers was argued in chapter 1. One step in evaluating the judgment processes in an assessment center is to compare the scoring 'rules' used in measuring the 'sample' behavior of the assessment center and in measuring the behavior in the job 'domain' (Guion, 1980). The scoring 'rule' used in the assessment center is the clinical judgment of the assessors.

Clinical judgments minimally occur at three points in the scoring of the behavioral sample generated by an assessment center. First, a judgment is made concerning the presence or absence of behaviors indicative of skills and abili-

ties required by the job. This judgment usually yields a numerical rating on a set of skill and/or ability requirements. Second, the dimensional ratings are combined to arrive at an overall rating of the candidates performance in the assessment center (c.f. Bray & Grant, 1966; and Huck, 1974). Each of the preceding two steps is usually performed by individual assessors followed by group consensus reaching activities. Third, the overall rating is used in combination with any other pertinent information by some manager (who is not active in the assessment process) to arrive at a selection/placement decision. The assessment center literature bearing on these three decision points shall be reviewed in this chapter.

Perception of Assessment Dimensions

The first assessor judgment noted above is the perceptual attribution of the presence or absence of behavior demonstrated by the candidate. Few studies in the assessment center literature directly address this issue. Byham (1977) provides the most detailed description of the training received by assessors in this area. He stated that, after being trained in how to observe and record behavior, assessors "categorize observed behavior under the dimensions determined important by the organization" (p. 101).

Thompson (1970) performed a multitrait-multimethod matrix analysis of dimensional ratings made by assessors (on behavior observed in an assessment center) and dimensional ratings made by the assessee's supervisor (on behavior observed on the job). The purpose of his analysis was to evaluate the convergant and discriminant validity of methods used in an assessment center. The analysis indicated that the supervisors failed to discriminate among the dimensions. Thompson speculates that the supervisors' ratings of a managers' job behaviors on thirteen assessment center traits mean something different relative to the assessor's ratings of the traits. This difference in meaning may be due to the fact that assessors were trained in the observation and recording of the relevant behaviors while the supervisors were not. Further, the assessors arrived at their dimensional ratings by observing behaviors in a standardized environment. The supervisors faced all the observational irregularities encountered in the actual working environment.

Cohen and Sands (1978) examined the impact of exercise sequencing on nine skill ratings and five summary ratings of exercises. They found that differential sequencing of first day exercises in a two day assessment center had no significant effect on these ratings.

Borman (1982) examined the relationship of assessee ratings on physical attractiveness and likability to their performance in six assessment center exercises. The only significant relationship found for a sample of 57 military recruiter candidates was between physical attractiveness and performance on a simulated interview with a concerned parent ($r = .28, p < .05$).

Group Influences

Sacket and Wilson (1982) examined what makes some assessors more influential in the assessment staff consensus reaching discussion. They also examined how well assessor consensus ratings on the assessment center dimensions can be predicted from pre-discussion ratings. Extreme ratings tended to have more influence on the final consensus rating than assessor sex, assessee sex, or the chairpersonship of the assessor group in a middle management assessment center. Averages of individual assessors ratings predicted consensus ratings better than modes. However, a combined rule (taking modes in some patterns of disagreement and means in others) was found to predict best.

Future Directions

Research is needed to model the cognitive processes used by the assessor to arrive at ratings on the assessment dimensions. The influence of exercise sequencing, assessee

physical attractiveness and likability, assessee sex, assessor sex, and extreme ratings prior to assessor consensus reaching activities on the outcomes of individual and group dimensional perceptions has been examined. However, these findings do not yield a complete understanding of the rating processes.

Integration of the Dimensions

The next step in the sequence of decisions made by the assessors is the integration of the assessment dimensions to arrive at an overall assessment rating (OAR). Few researchers have examined the relative weights and combinatorial processes applied to the assessment dimensions by the assessor in arriving at an OAR. Those who have, limited the scope of their search to additive models using least squares linear regression (cf. Mitchel, 1975; Moses, 1972; Sackett & Hakel, 1979; or Wollowick & McNamara, 1969).

Prior to 1975, the research results were at best inconclusive. Simple correlation coefficients between the dimensional ratings and OAR were being used to describe assessors' underlying decision processes (Moses, 1972, 1973). Indeed, Eyham (1977), in describing assessor training procedures, notes that the information integration and decision process are given very little emphasis. He stated:

Assessors seem to be able to (integrate information and arrive at decisions very) well and therefore there is very little training given out-

side of a brief lecture on the relative importance of the various dimensions, their interrelationships, and the role of the assessor in the final discussion of each individual. ...Because of the interrelatedness of the dimensions and the fact that the dimensions are interactive, research studies have tended to show that assessor weightings of the dimensions for a particular individual are superior predictions over a mathematical interpretation of the data derived from empirically defined weights (Huck & Bray, 1976; Moses 1973). (pp. 111-112)

However, Shanteau (1979) has shown that training can increase decision quality. Moreover, Huck and Bray (1976) and Moses (1973) do not present a mathematical interpretation of empirically derived weights. (See Slovic and Lichtenstein, 1971, for examples of mathematical interpretations.) They do provide zero order correlation coefficients between the various exercises and tasks performed in the assessment center and the respective job performance criteria employed. If, as Byham (1977) suggests above, the assessors do take into account the interrelationships of the assessment center dimensions, the zero order correlation coefficients cannot be interpreted as indices of the relative importance of the dimensions. If Huck and Bray (1976) had regressed their criteria onto the assessment center dimension scores using least squares multiple regression, at least a tentative conclusion regarding the merit of a "mathematical interpretation" (Byham, 1977, p. 112) over assessor interpretation of the dimensions could be made.

Bray and Grant (1966) alluded to the decision processes used by assessment staffs in discussing the results of a hierarchical factor analysis of the OAR and variables rated in an assessment center. They speculate that a higher order factor (third order for a college graduate sample and second order for a nongraduate sample) "could be described as reflecting the assessment staff's 'model' for managerial potential (the loadings of the {OAR} being highest on this factor)." (p. 7) They further speculate that the first order factors represent the "more specific judgments of the assessment staff." (p. 7) In essence, they are arguing that the first order factor analysis reflects the way in which assessors group the assessment dimensions into categories prior to arriving at an OAR. The higher order factor analyses are interpreted as representing the subsequent steps in the assessors' decision processes where the categories are integrated into an OAR. These speculations were not, however, addressed in the subsequent results reported by Bray and Grant (1966). There is no empirical or theoretical reason to believe that a set of linear equations summarizing a correlation matrix represents the way in which assessors use the information gathered in an assessment center.

Moses (1972) regressed a criterion onto the assessment center dimensions. However, he again used zero order correlation coefficients to interpret the relative importance of

the dimensions. After ranking the zero order correlations between the assessment dimensions and the criterion and then ranking the corresponding correlations between the assessment dimensions and the OAR, he argues that because these rank orders are identical support is shown for the validity of the assessor's judgment. Moses (1972) reports no significant difference in the multiple correlation coefficient between the assessment center dimensions and the criterion ($R = .463$, $N = 5943$) and the zero order correlation between the OAR and the criterion ($R = .44$, $N = 5943$). Moses (1972) also regressed the OAR onto the assessment center dimensions ($R = .824$, $N = 5943$). No regression coefficients were reported.

Huck (1974) found no difference between the OAR and a statistical combination of the assessment center dimensions in their ability to predict an overall job performance rating made by immediate supervisors ($r = .42$ and $R = .42$, $p < .01$ in both cases). When regressed onto ratings of potential for future advancement made by the immediate supervisor, the OAR and statistical combination of assessment dimensions did not differ meaningfully in their predictive power ($r = .59$ and $R = .56$, $p < .01$ in both cases).

Wollowick and McNamara (1969) found a large difference between the predictive power of the OAR ($r = .33$, $p < .01$) and that of a statistical combination of the components ($R =$

.62, no significance level reported) in predicting change in position level.

Mitchel (1975) found that the statistical combination of the assessment center components for three groups taken at five points in time yielded an average multiple R of .42 in predicting salary growth. The OAR yielded an average predictive value of $r = .22$. When the regression equations generated for one group were applied to the same group at a different lag period, the average multiple R dropped to .40. However, when the equations were applied between groups between time lags, the average multiple R dropped to .28.

Schmitt (1977) segmented judgments made by an assessment team into three time periods. Factors were derived from dimension ratings over the entire rating period. The OAR's were then regressed onto the factor scores for each period. The multiple R derived from data in the first time period was lower than the relationship found in the other periods.

Cohen and Sands (1978) examined the influence of exercise sequencing on the subsequent OAR. No significant effect was found.

Sackett and Hakel (1979) addressed four questions arising from previous research: "a) Do the interrelationships among dimension ratings made by individual assessors change over time, i.e., in the course of a 6-month assignment as an

assessor? b) To what extent do assessors differentiate among dimensions? c) How much information is used in reaching the overall decision? and d) Do perceptions of dimension importance match models of decision making?" (p. 121). In examining the dimensional ratings and OAR's for 719 individuals prepared by four assessment teams they found no changes in OAR over time. Factor analyses indicated similarities and differences among assessors and assessment teams. Two factors were common to almost all of the assessors and teams (i.e. leadership and organizing/decision-making). With an average interrater reliability of $r = .69$, Sackett and Hakel concluded that these two factors form the common base of agreement between raters while differences in factor structure beyond these two accounted for error variance. Eighty percent of the variance in CAR's could be accounted for by three dimensions regardless of the total number of dimensions under consideration (as indicated by the factor analysis specific to that assessor group). However, no attempts were made to fit multiplicative disjunctive, conjunctive, or averaging models to these data (Goldberg, 1968, 1970; Birnbaum, 1973, 1974). Finally, although they found statistical and subjective rankings of dimension importance not to be highly related, the dimensions of leadership, organizing and planning, and decision making were consistently ranked among the most important by all the assessors.

The Sackett and Hakel (1979) study is the most sophisticated attempt at gaining an understanding of the underlying decision processes. Alternative models of information integration need to be examined. Such tests would not involve the use of significance levels in evaluating multiple correlation coefficients (Birnbaum, 1973, 1974; Shanteau, 1977). Anderson (1974) describes the iterative nature of the evaluation of a decision model and some effects that may cause an investigator to conclude, at first glance, that an hypothesized model is inappropriate when it is not (e.g. serial integration of information and the averaging model).

Possibility of a Nonlinear Model

One of the least parsimonious models is suggested in the previous quote from Byham (1977). He argues that, because of their ability to take into account the interactive nature of the assessment center components, assessors are "superior predictors over a mathematical interpretation of the data..." (p. 112). In short, Byham is arguing that the integration of the assessment components into an OAR by the assessor is better than a mathematical interpretation because it is nonlinear. This argument needs to be investigated on two counts.

First, Dawes and Corrigan (1974) have shown that linear models will outperform (in terms of variance explained) intuitive judgment when (a) the cues have a conditionally monotone relationship to the criteria, (b) there is error in the criteria, (c) there is error in the cues, and (d) deviations from the optimal weighting make little practical difference. In this study, the judgment cues are the consensus ratings on the eighteen dimensions. Assumptions (b) and (c) are surely met in the assessment center literature.

Second, assume for the moment that the assessors do combine the assessment center components in some nonlinear fashion. Specifically, the "interactive" nature Byham proposed suggests a multiplicative or conjunctive model as described by Goldberg (1968). The interest in such models is not in the relative accuracy of predictions, but rather in gaining an understanding of the assessor's cognitive decision processes.

Future Directions

The importance of understanding these processes cannot be overemphasized. In order to live up to the tentative confidence expressed by the courts in the content validity of assessment centers (Firefighters Institute v. City of St. Louis, 1977), a better understanding of assessor information integration processes is required. It cannot be

assumed that just because the parameters of a linear multiple regression equation do not differ across protected subgroups of the population that differential treatment is not occurring. The linear compensatory model may not be describing what assessors are doing despite high correlation coefficients (Anderson & Shanteau, 1977; Dawes & Corrigan, 1974; Birnbaum, 1973, 1974; Shanteau, 1977). Further, an understanding of the decision processes may increase decision quality and reduce assessor group conflict (Dawes, 1980; Hammond, Rohrbaugh, Munpower, & Adelman, 1977).

Use of the Overall Rating

The final step in the use of the information gathered from an assessment center is in the actual promotion/selection decision made by some line manager. No research has addressed the way in which line decision makers subsequently use the assessment center OAR in arriving at final promotion/selection decisions. However, some interesting data are reported by Huck (1974).

Table 1 presents data originally gathered by Huck (1974) at Michigan Bell Telephone. Particular attention should be paid to the column labeled Black % / White %. This column contains the rejection rates used to determine the extent of disparate impact on protected subgroups of the population (Griggs v. Duke Power Co., 1971; and Albemarle

Paper Co. v. Moody, 1975). Using the 4/5 rule as outlined in the Uniform Guidelines on Employee Selection Procedures (1978), one can see that blacks are being promoted at a disproportionately higher rate relative to whites. However, at the same time it should be noted that blacks are falling into the high OAR rating at a disproportionately low rate and into the low OAR rating at a disproportionately high rate. It should also be noted that the 'promoted' row consists only of those women promoted to one of two first-level management positions. It does not include those women promoted into higher levels of management or into other first-level positions.

If it can be assumed that movement of black and white women into other management positions occurred in approximately the same proportions reported for these two positions, at least one hypothesis can be made concerning the selection decision processes. While Huck (1974) found that (a) when the OAR is regressed onto a measure of overall job performance the slopes and intercepts for whites and blacks do not differ and (b) the assessment center dimensions predictive of overall job performance were the same for both groups, it would appear that the race of the assessee is influencing the line manager's ultimate promotion decision. It is obvious that alternative explanations (e.g. the blacks had greater job tenure) and influences need to be examined

before any evaluation of an assessment center's compliance with equal employment opportunity regulations can be made.

In this vein, it is of interest to note comments made in the final report on AT&T's compliance with a consent decree it signed in 1973 (U.S. v. AT&T, 1978). In the decree, assessment centers were cited as one method with which college graduate females might be evaluated for employment in management positions. The final report states:

The test batteries used by the Bell System in the clerical, craft and management areas generally screen out minorities or women at a substantially disproportionate rate. Under the Decree the use of such tests was acceptable so long as intermediate targets were being met, thus precluding the necessity of a lengthy and complicated validation inquiry. Absent the affirmative requirements of the Decree, the tests could become a serious impediment to the Bell System's progress and could be unlawful. (p. 124{o})

Future Directions

The data presented by Huck (1974) (which were collected prior to the signing of the consent decree) and the quote cited above suggest that research is needed on the influence of sex and race of the assessee on decision processes within and outside of the assessment center. The OAR is not the only piece of information that is used to arrive at a final placement/selection decision.

Sex Effects

Title VII of the 1964 Civil Rights Act requires that the large majority of private sector organizations provide equal employment opportunity to certain subgroups of the population. Males and females are two of those protected subgroups.

One study has examined the impact of sex and race on the observation of behavior in a work sample (Hamner, Kim, Baird, and Bigoness, 1974). They had white and black females and males rate the overall performance of video taped samples of grocery store shelve stocking behavior. The actor in the tape varied in terms of sex, race, and objective level of performance. While they found 30% of the variance in performance ratings explained by the objective level of performance exhibited by the actor, there were also strong race and sex effects. Female actors were given higher performance ratings than males. In addition, raters tended to give higher ratings to actors of the same race. An additional 25% of the variance in performance ratings was explained by various sex-race combinations.

No research has examined the impact of sex of the assessee on the decision processes of assessors. Three studies have examined the relationship of sex of the assessee to the outcome of an assessment center (i.e. the CAF).

Moses (1973) examined the relationship of performance in an Early Identification Assessment Program (a short, one day, assessment designed to evaluate a high volume of assessees) to performance in a Personnel Assessment Program (a longer, more extensive, assessment center) for male and female employees of AT&T. The purpose of the study was to "evaluate the effectiveness of the judgments derived in the (Early Identification Assessment Program)." (p. 571) Both centers were designed for the selection of managers. Moses (1973) reports the means, standard deviations, and correlations between OARs generated by each center on 39 men and 46 women. No significant differences were found, although females' mean OAR for each center and correlation between center OARs were lower than males'.

Moses and Boehm (1975) examined the relationship between assessment and subsequent progress in management for 4,846 females and 8,885 males who attended the Personnel Assessment Program at AT&T. The distribution of OARs received by males and females were strikingly similar. Percentages of males and females found in the four OAR categories did not differ by more than 3.7%. Further, the four predictors correlated most highly with management level subsequently obtained were the same for males and females (OAR, leadership, decision making, and organizing and planning). The OAR correlated .37 for females and .44 for males with

managerial level obtained, though, since their superiors were aware of their assessment center performance, the possibility of criterion contamination existed.

Finally, Schmitt and Hill (1977) examined the effects of the race-sex composition of the assessee group on the OAR and seven dimensions rated in an assessment center. Their sample consisted of 73 white males, 146 white females, 11 black males, and 76 black females. While overall influence of the sex-race composition of the groups was found to be negligible, the zero order correlations were consistent with cultural stereotypes. For example, black females were considered less forceful as the number of white males in their assessment group increased. The ratings on five of seven dimensions increased for white males as the number of white males in the group increased.

Future Directions

The research examining the effects of assessee and assessor sex on the way in which information gathered in an assessment center is used has hardly begun. There are a number of ways in which the sex of the assessee could influence a candidates' performance in an assessment center. Examples of some of these are listed below:

1. The skills, abilities, and behaviors identified in the job analysis to be evaluated in an assessment

center may reflect attributes necessary to perform in a male dominated occupation. They may not be required for actual performance of the job.

2. A male and female who engage in the exact same set of behaviors in the assessment center may receive different dimensional ratings.
3. Males and females with the exact same profiles of ratings on the assessment center dimensions may receive different CARs.
4. Males and females with the exact same OARS may be selected/promoted at different rates.

The gross influences of sex of the assessee and sexual composition of the assessee group on assessment center performance evaluated above barely begin to address these issues.

Summary

The argument was made that alternative explanations to the predictive validity of assessment centers remain. One way to address this issue is to evaluate the content validity of assessment centers. A first step in this evaluation is the examination of the 'scoring rule' used to arrive at an overall assessment rating.

Three steps were identified in the clinical judgments of assessors that make up this scoring procedure. Literature bearing on the perceptual and integrative processes of

assessors was examined. Further, literature bearing on the way in which the OAR is subsequently used by line decision makers was summarized. Finally, the influence of sex of the assessee on these processes was examined in the literature.

The overriding conclusion that can be drawn from this literature is that very little is known about the cognitive processes used by assessors to arrive at the OAR. What little research exists examines i) the impact of various manipulations of the decision setting on the decision outcome (e.g. the influence of sequence of exercises on OAR) or ii) simple linear models of the decision process.

CHAPTER III
METHODOLOGY

The Assessment Center

The assessment center from which data for this research were obtained was established to measure the managerial abilities required of a candidate for first level supervisor. The procedures employed in this center were developed from those used in a center described by Huck (1974). The assessment center to be examined here was used by a large midwestern public service organization between December, 1979 and March, 1981. Confidentiality of subject and company identity was a prerequisite to obtaining the data.

During this period a total of 70 assessors evaluated 2191 assessees. Attendance in this assessment center was required to be considered for promotion to first-level supervisor. Candidates could either nominate themselves or be nominated by their supervisor to attend the program. Because of the geographical dispersion of the firm, centers operated in five states (each with a separate staff). Candidates from various departments were gathered in groups of six for the one day assessment activities.

Dimensions

The assessment center yields ratings on eighteen managerial dimensions. These dimensions are then grouped into four categories of abilities. These categories, based on intraorganizational literature, and the dimensions in each are listed below.

- i) Personal Qualities: The ability to perform well under pressure or when conditions are ambiguous. In addition, the ability to perceive self weaknesses and strengths.
 1. Energy: To what extent can the individual maintain a continuous high level of work activity?
 2. Self Objectivity: To what extent does this individual realize his or her own strengths or weakness?
- ii) Interpersonal Skills: The ability to engage in the various behaviors required to successfully interact with different people under different circumstances.
 1. Impact: To what extent does the individual make an impression on others?
 2. Leadership: To what extent can the individual effectively lead a group to accomplish a task without incurring hostility?

3. Awareness of Social Environment: To what extent can the individual perceive subtle cues in the behavior of others toward him or her?
4. Behavior Flexibility: To what extent can the individual, when motivated, modify his or her behavior to reach a goal?
5. Autonomy: To what extent does the individual take independent action?

iii) Problem Solving Skills: The ability to organize and plan their own work and the work of subordinates in order to gather, interpret and analyze facts necessary to make decisions.

1. Fact Finding-Oral: To what extent can the individual effectively interact with another person to obtain information regarding a problem?
2. Fact Finding-Written: To what extent can the individual effectively obtain information regarding a problem from available written sources?
3. Organizing: To what extent can the individual effectively structure his or her own work and the work of others for the most efficient accomplishment of a task?

4. **Interpreting Information:** To what extent can the individual distinguish relevant information and utilize that information when making decisions?
5. **Planning:** To what extent can the individual effectively plan his or her own work and the work of others?
6. **Decisiveness:** To what extent is the individual willing to make decisions when required?
7. **Decision Making:** To what extent can the individual make decisions of high quality?

iv) **Communication Skills:** The ability to prepare and defend their thoughts clearly on paper and when speaking to small groups.

1. **Oral Presentation:** To what extent can the individual effectively present an oral report to an individual and small group?
2. **Oral Defense:** To what extent can the individual, when presenting ideas orally, effectively respond to challenges and questions raised by others?
3. **Writing Fluency:** To what extent can the individual demonstrate knowledge of writing fluency in terms of the standard grammatical mechanics of the English language?

4. Scholastic Aptitude: How does the individual compare to other individuals in his or her ability to learn new things?

Techniques

The exercises used in rating an assessee's performance in each of these dimensions are listed in Table 2. The assessee takes on the role of a retail manager of a franchise department store. He/she is one of a team of managers (in this case the other assesseees in the particular group) who have been called in to help straighten out one of the franchises' other stores where the manager turned over with little notice. The four assessors take on the role of district managers who are on hand to initiate and oversee the activities of the team. Each assessee then engages in the following four exercises (taken from intracorporational literature).

- i) Fact Finding Interview: Given an initial description of the store, the assessee is given the opportunity to interview one of the 'district managers' (assessors) to gain more information concerning the current condition of the store.
- ii) In-Basket: Each assessee is given a set of materials that had been left behind by the previous manager and/or accumulated in his absence. The mater-

ial consists of letters from customers, internal memos, stationary, etc. The assessee is asked to review the material and take appropriate action on each item. After completion, each assessee is interviewed by an assessor concerning his/her approach to the task and reasons behind his/her actions.

- iii) Case Material Proposal: Each assessee prepares and presents a proposal to the assessee group dealing with one aspect of the problems faced by the store.
- iv) Group Meeting: The assessees are told as a group that the initial budget available to address the problems at the store has been cut in half. The assessees as a group are to discuss the different problems faced by the store and to redistribute the available resources as they think appropriate. Each assessee is told that it is his/her task to defend their specific proposal in the face of this resource redistribution.
- v) Paper and Pencil Instruments: A scholastic aptitude test and background questionnaire are given to each assessee. These are designed to measure general mental ability or learning ability and obtain biographical data.

All of the techniques are administered according to standard instructions by the assessor(s) in charge of the exercise. All tests are given to groups of assessees while all interviews are conducted by one assessor with one assessee. The group exercises involve six assessees at one time.

Assesseees and Assessment Staff

Non-management employees assessed at the center were either self nominated or selected by their supervisor. The large majority of the assesseees were taken from clerical or blue collar positions within the firm. Since this is an assessment center for selection into entry level managerial positions, no managerial personnel were assessed.

The group of assessors is made up of second level supervisors, all of whom had been first level supervisors at one time. Assessors receive four weeks of training in the observational and rating skills required to evaluate performance in the assessment center. This training is very intensive, with the assessors spending ten to twelve hours per day honing their observation and recording skills.

Staff assessors, the more junior members of the assessment team, are on a three to six month rotation (the firm uses a form of matrix management structure). Staff directors, the more senior members of the assessment team, are on a two to three year rotation. They are recruited out of the

ranks of staff assessors. See Appendix B and C for job descriptions of these two positions. Each assessment team consists of one staff director and three staff assessors.

Rating Procedures

One or more assessors write up a summary report on each assessee's performance in each exercise or interview. The paper and pencil tests are scored and summarized by clerical personnel.

One week after the assessee has completed the exercises, the four assessors meet to review and evaluate each assessee. The review of each assessee takes from one to two and a half hours. Each assessor reads the reports summarizing the exercises they observed, the paper and pencil test results are distributed, and discussion ensues. Each assessor has a photograph of the assessee in front of him or her to aid recall. When the discussion is completed, each assessor individually rates the assessee on the eighteen dimensions listed above (from 1 {low} to 5 {high}). The assessors then go around the table and, for one dimension at a time, announce their rating. When differences occur, assessors discuss the evidence pertaining to that dimension and are allowed to change their ratings. Discussion must continue until all of the assessor ratings are within one scale unit of each other (i.e. a dimension receiving ratings

of one and three must be discussed until the assessors change to ratings that are no more than one unit apart). Since there are only four assessors, it is possible to have a tie. In this case the staff director (the senior assessor) is given two votes and the 'majority' rating is recorded as a consensus.

After a consensus has been reached on each dimension, the assessors individually consider the profile of eighteen dimensional ratings to arrive at an overall assessment rating (OAR). Staff directors taught assessors to give equal weight to the assessee's performance on each of the dimensional categories. Assessors arrived at category 'scores' for an assessee by subjectively assigning an internal central tendency index to the dimensions in that category (a kind of gestalt). In addition, the staff directors teach assessors to use an interaction term between the Interpersonal Skills and Problem Solving Skills categories. The staff directors indicated that the Problem Solving Skills category was to receive more weight in arriving at the OAR when the assessee had scored high than when the assessee had scored low on the Interpersonal Skills category. The staff directors argued that because a person could learn problem solving skills relatively easily while their interpersonal skills were relatively fixed, assessee's with high interpersonal skills and low problem solving skills should receive

an OAR higher than an individual who is high on problem solving skills and low on interpersonal skills.

If all the assessors independently arrive at the same OAR, it is recorded as a consensus. If there is disagreement, the same consensus reaching activities engaged in for the dimensional ratings are repeated until a consensus or 'majority' OAR is obtained. There are four possible OARs. They are:

- i) High (4): This individual is seen as having substantial potential to perform well as a first-level manager based on the qualities evaluated in assessment. The probability of this person succeeding at that level is very high.
- ii) Good (3): This individual is seen as having potential to perform well as a first-level manager based on the qualities evaluated in assessment. The probability of this person succeeding at that level is good.
- iii) Limited (2): This individual is seen as having limited potential to perform well as a first-level manager based on the qualities evaluated in assessment. The probability of this person succeeding at that level is limited.
- iv) Low (1): This individual is seen as having low potential to perform well as a first-level manager

based on the qualities evaluated in assessment. The probability of this person succeeding at that level is significantly limited.

Data Collection

The data for this research was collected from record sheets maintained by the firm. Each sheet contained the following information:

- i) The name, social security number, age, job tenure, race, and sex of the assessee.
- ii) The eighteen individual ratings made by each of the four assessors on the dimensions.
- iii) The consensus ratings of the assessor group on each of the eighteen dimensions.
- iv) The overall assessment rating made by each of the four assessors.
- v) The consensus overall rating made by the assessor group.
- vi) The identities of the four assessors.

The sex of the assessors was obtained from their personnel files. The data were key punched directly from the record sheets into the computer. This initial transmission was then subjected to verification. There were 976 males and 1205 females candidates assessed in this data set. There were 32 male assessors and 38 female assessors.

A number of Staff Directors and Staff Assessors were interviewed individually concerning the way in which they use the dimensional ratings to arrive at an OAR. In addition, the in house training manual used in assessor training was examined for reference to the assessor judgment processes. The assessors and training manual uniformly indicated that the dimension categories were given equal weight. Further, the interaction term between the Interpersonal Skills and Problem Solving Skills categories was described by all sources.

Method

The following series of analyses were organized around subquestions designed to address Research Question 1 below.

Research Question 1

How is the information gathered in an assessment center used by assessors to arrive at an overall rating for a particular assessee?

All regression analyses used the algorithm found in the General Linear Models procedure in the Statistical Analysis Systems software package (SAS Institute, 1979). The stepwise regression analysis used the Stepwise procedure in SAS. Finally, all factor analyses performed used the algorithms found in the Statistical Package for the Social Sciences (SPSS) software package (Kim, 1975).

Prior to performing any of the analyses designed to address Research Question 1 a multiple least squares linear regression was performed on the half of the sample whose last digit of their social security was five or greater. The OAR was regressed onto the eighteen dimensions. This was repeated for the second half of the data and a double cross validation performed. This was done in order to examine the extent to which shrinkage would occur in the multiple correlation coefficients due to the sampling error. Because there was evidence of minimal shrinkage, all subsequent analyses were performed on the entire sample. The results of the cross validation are presented in chapter IV.

How well can the OAR be predicted by the eighteen dimensions?

The OAR was regressed onto the eighteen assessment dimensions in order to examine their predictive power. A stepwise least squares multiple regression was performed to determine how well a smaller complement of dimensions would predict the OAR. The alpha level used as a criterion for a dimension's inclusion or removal from the equation was .01. Because of the extreme sample size, an alpha level of .05 was feared to allow for the inclusion of the majority of the dimensions, hence precluding any interpretation as to how predictable the OAR might be by a smaller complement of dimensions.

Is there a simple structure underlying the eighteen dimensions corresponding to the a priori categories?

Principal components analyses with orthogonal and oblique rotations were performed on the eighteen dimensions to see if the dimensions load onto the a priori categories listed above. Initial analyses retained all factors with eigen values greater than one for rotation. Subsequent analyses forced the loadings into four factors in order to confirm the a priori categories. Finally, the loadings were forced into some 'appropriate' number of factors as indicated by a skree test.

Do the assessors arrive at the OAR in the way they say they do?

The assessors were trained to give equal weight to the four a priori categories in arriving at an OAR. Category 'scores' were the assessors subjective impression of the central tendency of the dimensional ratings in each category. The scores used here were simply the arithmetic averages of the dimensional ratings found in each category and each factor analytically derived factor. Further, the assessors were trained to use an 'interaction term' between the categories of Interpersonal Skills and Problem Solving Skills. Consequently, the OAR was regressed onto subscale scores and the two-way interaction term between Interpersonal Skills and Problem Solving Skills categories. This

was also done for the factor analytically derived subscale scores (with the two-way interaction term consisting of those factors dominated by the Interpersonal Skill and Problem Solving Skill dimensions).

Research Question 2

How does the male to female ratio of the assessor group and the sex of the assessee moderate the relationship delineated in Research Question 1?

In order to examine the relationship of sex of the assessee to the OAR a chi-square test was performed on a OAR by assessee sex contingency table (chi-square = $[\text{observed} - \text{expected}]^2 / \text{expected}$, squared and summed over all cells of the contingency table). This was followed by two sets of multiple least squares linear multiple regressions. Because of the large sample size and the large number of dimensions used in this assessment center, the examination of the effects of assessee sex and male to female ratio of the assessor group was limited to the relationship between subscale scores and their two-way interactions to the CAR (as delineated in the last subquestion asked in addressing Research Question 1). It assumed that because the assessors were trained to use their central impressions of the factors' dimensional ratings in arriving at the OAR, any effects of the male to female ratio of the assessor group or effects of the assessee's sex would be manifested in the relationship between the subscale scores and the OAR.

Effects of sex of the assessee

In the first regression analysis, the independent variables consisted of the subscale scores for the a priori categories, the two-way interaction term between Problem Solving Skills and Interpersonal Skills, a dummy variable representing the sex of the assessee, and all interactions between the dummy coded sex variable and the subscale scores (including the assessee sex by Interpersonal Skill by Problem Solving Skill interaction). This was repeated for subscale scores calculated from factor analytically derived factors (again, using subscale score interactions for those factors which are dominated by Interpersonal Skill and Problem Solving Skill dimensions).

Effects of male to female ratio of assessor group

Second, the previous regression analyses were repeated with a substitution of a dummy coded male to female ratio of the assessor group variable replacing the dummy variable for the sex of the assessee.

CHAPTER IV

RESULTS

In this chapter the results of the analyses bearing on Research Question 1, Research Question 2, and the subquestions delineated in chapter 3 are presented. When indicated by these results, additional analyses were performed and their results reported.

Research Question 1

How is the information gathered in an assessment center used by assessors to arrive at an overall rating for a particular assessee?

Double Cross Validation

Squared multiple correlations of .811 and .796 were obtained when the OAR was regressed onto the eighteen dimensions for the two randomly sampled halves. When the regression coefficients generated by these analyses were used to predict the OAR for the individuals in the other half of the data, the squared multiple correlation coefficients decreased by .02 and increased by .009 respectively. Consequently, as noted in chapter III, all subsequent analyses were performed on the entire sample.

How well can the OAR be
predicted by the eighteen
dimensions?

The matrix of intercorrelations among the eighteen assessment dimensions and the consensus overall rating appear in Table 3. The best single predictor of the CAR is dimension 11, Interpreting Information ($r = .69$). The weakest single predictor is dimension 13, Decisiveness ($r = .32$). All the correlations with the OAR are significant at $p < .0001$, due primarily to the large sample size ($N = 2191$).

The results of six regression analyses, in which the OAR is regressed on to various combinations of the eighteen dimensions, are presented in Table 4. The first set of regression coefficients was generated by regressing the OAR onto all eighteen dimensions. Energy, Impact, Awareness of Social Environment, and Autonomy yield nonsignificant regression coefficients. All the rest of the dimensions yield coefficients significantly different from zero at $p < .001$ or less.

The second set of regression coefficients was generated by a stepwise multiple regression analysis. The purpose of performing the stepwise regression analysis was to determine how well a smaller complement of dimensions would predict the OAR. With a criterion of $p < .01$ for entry or exit from the equation, fifteen of the eighteen dimensions remained.

All the dimensions that failed to yield significant regression coefficients in the full equation, with the exception of Impact, also failed to remain in the equation resulting from the stepwise procedure. The multiple correlations generated from these two equations were identical.

With fifteen variables entering the equation using $p < .01$ as a criterion, the objective of evaluating the predictive power of a smaller compliment of dimensions has only been partially accomplished. Consequently, the results of four additional analyses are reported.

It has already been noted that dimension 11, Interpreting Information, is the one best predictor of the assessor group judgment. The last four regression analyses reported in Table 4 (labeled MaxR) present the best predicting groups of two, three, four, and five dimensions. For example, of all possible pairs of dimensions, the two with the greatest predictive power are Interpreting Information and Leadership. The best predicting set of three dimensions adds Organizing to the previous two. It is of interest to note that 92% of the variance explained by all eighteen dimensions is explained by but five dimensions (Leadership, Behavior Flexibility, Organizing, Interpreting Information, and Decision Making). It would appear that, at least for the assessors being examined here, a substantial proportion of the OABs could be obtained using only five of the eigh-

teen dimensions rated in the assessment center. The answer to this subquestion would have to be that the OAR can be predicted quite accurately by the dimensional ratings.

Is there a simple structure
underlying the eighteen
dimensions?

Factor analyses were performed to determine if there is a smaller set of constructs underlying assessees' ratings on the eighteen dimensions. The factor analyses generated linear combinations of dimensions in an empirical attempt to identify any possible underlying constructs. No theory or model exists which specifies the degree of correlation (if any) between linear combinations of the dimensional ratings obtained in an assessment center or their corresponding constructs. Consequently, the correlation matrix reported in Table 3 was subjected to common factor analysis and subsequently rotated to varimax and oblimax solutions respectively. The angle of the oblique rotation performed using the oblimin procedures was limited to two different minimums ($\delta = 0$ or -1), yielding two different factor pattern matrices. The factor pattern matrices generated by these three procedures are reported in Tables 5, 6, and 7 (only the factors with eigen values greater than one are presented). These matrices contain the weights used to estimate the original dimensions from the factors. From these

weights, any underlying structure to the dimensions can be interpreted (Kim, 1975). Factor structure matrices were not interpreted or reproduced here because simple correlations between the factors and dimensions do not lend themselves to interpretation for simple structure (Nunnally, 1978). A skree test indicated that three factors were dominant in terms of the relative size of their eigen values (indeed, only three factors had eigen values greater than one).

For the varimax rotation, the criterion for inclusion of a dimension in the factor was that it load at least .45 or greater on one factor and no greater than .40 on any other factor. For the oblimin rotation, $\delta = 0$, the criterion for inclusion of a dimension was that it load at least .45 or greater on one factor and no greater than .25 on any other factor. With $\delta = -1$, the criterion for inclusion was .40 or greater on one factor and no greater than .30 on any other factor. A summary of the dimensions loading in the three factors using these criteria is given in Table 8.

In each rotation, Fact Finding-Oral did not load on any factor. In fact, the loadings of Fact Finding-Oral are almost equal for the three factors. Self Objectivity did not load in either of the two oblimin rotations. The only reason it loads in Factor 1 using the varimax rotation is the use of a less rigorous loading rule (.45/.40). However,

Self Objectivity did load highest on Factor 1 in the two oblimin rotations. Further, the loadings from the varimax rotation appear to be parsimonious in that a) Personal Qualities, Interpersonal Skill, and the two oral communication dimensions load in Factor 1, b) Problem Solving Skills load in Factor 2, and c) the dimensions measured by paper and pencil instruments load in Factor 3. It does not seem counter intuitive that Personal Qualities, Interpersonal Skills, and the two communication skills dimensions would be highly related to each other while Problem Solving Skill dimensions would be a highly intrarelated yet independent group. The fact that the dimensions measured by paper and pencil instruments load on a separate dimension confirms assessors' statements noted earlier that they have the least amount of confidence in these dimensional ratings. Consequently, the loadings described in Table 8 for the varimax rotation were used in calculating the factor subscale scores (averages) to be used through out the rest of this chapter.

Three additional factor analyses were performed to determine if the a priori category groupings would be confirmed by the loadings of the dimensions. Other than forcing the analysis to yield four factors, each was identical to the varimax and oblimin rotated common factor analyses described above. The factor pattern matrices for these analyses are reported in Tables 9, 10, & 11. Two points are of interest

in these matrices. First, the dimensions from the a priori categories did not load highest within the same factor. Second, the two dimensions measured by paper and pencil instruments consistently loaded by themselves on the third or fourth factor. None-the-less, since the assessors were trained to use their general impressions of a candidate's performance in each category to arrive at an OAR, the a priori category groupings were also examined in subsequent analyses.

In answer to this subquestion, there does appear to be a simple structure underlying the eighteen dimensions. In this structure, Factor 1 loads with dimensions concerned with and/or related to Interpersonal Skills, Factor 2 loads with dimensions dealing with the observation of Problem Solving Skills, and Factor 3 contains the two dimensions rated by paper and pencil instruments. This structure does not correspond with the a priori category groupings.

Do assessors arrive at the CAR
in the way they say they do?

The assessors were trained to arrive at the OAR by giving each of the a priori categories equal consideration. In addition, they were to use the Interpersonal Skill and Problem Solving Skill categories interactively. This contention concerning the assessors' decision processes is examined in the following analyses. Because factor analysis yielded a

three factor solution with the first two factors dominated by Interpersonal Skill and Problem Solving Skill dimensions (respectively), a model in which the assessors give equal weight to the empirically derived factors was examined. The relationship of the interaction of the factors dominated by the Interpersonal Skill and Problem Solving Skill a priori categories to the OAR was also examined.

Exact factor scores were not calculated due to the assessors' contentions that they formed a "wholistic" impression of each a priori category of dimensions. Consequently, subscale scores were derived by averaging the variables selected for each factor. Category subscale scores were calculated in the same way.

Table 12 presents the results of three multiple regression analyses in which the OAR is regressed onto category subscale scores, sex of the assessee, and interaction terms. Columns A and B were used to examine the assessors' use of the a priori categories. Column A gives the regression coefficients obtained when the OAR is regressed onto the category subscale scores. All coefficients are significantly different from zero at $p < .0001$. The category subscale scores accounted for 79% of the variance in the OAR. Column B adds the Interpersonal Skill by Problem Solving Skill interaction term that the assessors indicated they used. None of the coefficients for the category subscale

scores change dramatically. The interaction term is not significant, indicating that the assessor groups were not using the interaction of the Interpersonal Skills and Problem Solving Skills categories as they had previously indicated. Column B accounts for 79% of the variance in the OAR. Column C will be discussed in the subsequent presentation of the analyses for Research Question 2.

Table 13 presents the results of three multiple regression analyses in which the OAR is regressed onto factor subscale scores, sex of the assessee, and interaction terms. Again, equations A and B are of interest in examining the way in which assessors' use the information contained in the factor subscale scores (column C is again pertinent to Research Question 2). Column A gives the regression coefficients obtained when the OAR is regressed onto the factor subscale scores (the subscale scores have been named after the dimensions of the category which dominates that factor). All coefficients are significantly different from zero at $p < .0001$. These factor subscale scores predict 78% of the variance in the OAR. Column B includes an interaction term between the two factors dominated by categories the assessors were ostensibly using interactively in arriving at an OAR. This interaction term is negative and significant ($p < .02$), exactly opposite of the way in which the assessors interviewed stated they used it. Again, the coefficient for

the original factor subscale scores did not change dramatically while the variance explained stayed exactly the same.

These findings seem to indicate that assessors are not using the categories the way they are trained to. Further, it could be tentatively concluded that, if the factor subscale scores are more representative of the way in which assessors use the dimensional ratings, the assessors are using the Interpersonal Skill by Problem Solving Skill interaction term in a fashion which is opposite the way in which they are trained.

Research Question 2

How does the sex of the assessee and the male to female ratio of the assessor group moderate the relationship delineated in Research Question 1?

The way in which assessee sex and assessor group male to female ratio moderate the relationship between the CAR and i) the a priori category subscale scores and ii) the empirically derived factor scale scores are to be reported here.

Prior to running analyses to examine this moderating effect, a contingency table was calculated showing the frequency with which assessees received each rating on the OAR by their sex (see Table 14). The sample size decreased by ten because of missing values for assessee sex. The chi-square statistic generated from this table is significant (p

< .0001). Examination of the cell frequencies indicates that males tend to receive higher overall assessment ratings than females. Cells for OAR ratings of 1 and 4 would seem to be the source of greatest difference between men and women.

Regressions Involving Assessee Sex

Column C in Tables 12 and 13 evaluate the influence of assessee sex when the OAR is regressed onto category and factor subscale scores. Table 12 shows that when a dummy coded sex variable is entered into the equation as a main effect and in interaction with the previous category subscales in the equation, there is minimal increase in variance explained (.001). However, while only the main effect for Personal Qualities becomes nonsignificant, the coefficients for Interpersonal Skills and Problem Solving Skills increase meaningfully (almost double from column A). The Interpersonal Skills by Problem Solving Skills interaction remains negative and becomes significant ($p < .01$). Finally, the effects of sex, sex by Interpersonal Skills, sex by Problem Solving Skills, and sex by Interpersonal Skills by Problem Solving Skills are significant ($p < .01$).

Table 13 shows that when a dummy coded sex variable is entered into the equation as a main effect and in interaction with the previous factor subscale scores in the equa-

tion, there is minimal increase in variance explained (.002). However, while all the main effects remain significant (Problem Solving Skills almost triple and Interpersonal Skills almost double from column A), the Interpersonal Skill by Problem Solving Skill Interaction changes from $-.046$ ($p < .02$) to $.895$ ($p < .0001$). There is a negative main effect for sex ($p < .02$) and the interaction effects sex by Interpersonal Skills ($p < .02$), sex by Problem Solving Skills ($p < .0001$), and a positive sex by Interpersonal Skills by Problem Solving Skills effect ($p < .0001$). Examination of marginal means for males and females indicates Interpersonal Skills and Problem Solving Skills receive more weight for males while the Interpersonal Skill by Problem Solving Skills interaction receives a negative weight for males and no weight for females.

Finally, the OAR was regressed onto a) the factor subscale scores and b) the factor subscale scores and Interpersonal Skill by Problem Solving Skill interaction separately for the male and female assesses in the sample. The predicted mean OAR by sex for each of the two equations derived from both the male and female subsamples is presented in Table 15.

The negative main effect for sex found in Tables 12 and 13 confirms the results of the chi-square test on the assess sex by OAR contingency table. What is most interesting

is the reversal in sign and dramatic increase in the regression coefficient for the sex by Interpersonal Skills by Problem Solving Skills interaction term in Table 13. This would seem to indicate that the assessors are using the Interpersonal Skill by Problem Solving Skill interaction term in a beneficial way for males and a harmful way for females. However, when examining the predicted mean CARs of the various equations in Table 15 it becomes clear that whatever effect this interaction is having, it is being counteracted by the negative sex by Interpersonal Skill and sex by Problem Solving Skill factors. There is very little change in male and female predicted CARs when equations derived from only male and female assesseees are used.

Assessor Group Sexual Composition

The male to female ratio of the assessor group was dummy coded (0 = all males, 1 = one female/three males, 2 = two females/two males, 3 = three females/one male, and 4 = all females) and entered into Column B from Tables 12 and 13 as a main effect and interacting with all previously entered variables. Table 16 reports the regression coefficients derived when sexual composition of the assessor group (GSEX) was entered with the category subscale scores. The only main effect not to be significantly different from zero is Interpersonal Skills. The only interactions to reach signi-

ificance were GSEX by Problem Solving Skills and GSEX by Problem Solving Skills by Interpersonal Skills ($p < .05$).

Table 17 reports the regression coefficients derived when GSEX was entered into the equation with the factor subscale scores and Interpersonal Skills by Problem Solving Skills interaction effects. Only Interpersonal Skills and the paper and pencil dimensions reached significance as main effects ($p < .0001$ and $.01$ respectively). GSEX interacted positively with Problem Solving Skills ($p < .05$) and negatively with the Interpersonal Skills by Problem Solving Skills interaction ($p < .01$).

Finally, the OAR was regressed onto a) the factor subscale scores and b) the factor subscale scores and Interpersonal Skill by Problem Solving Skill interaction for the five different GSEX levels in the sample. The predicted mean OAR by GSEX level for each of the five equations derived from the five GSEX subsamples (i.e. assesses assessed by assessor groups consisting of one, two, three, or four males) is presented in Table 18.

Without an understanding of the group interaction processes taking place among the assessors, it is difficult to interpret these regression coefficients. Examination of the predicted OAR means seems to indicate that the female dominated assessor groups tend to give higher OARs.

CHAPTER V
SUMMARY AND DISCUSSION

The purpose behind the present research was to investigate the way in which assessors use the information gathered in an assessment center to arrive at an overall rating of a candidate's potential as an entry level manager. An additional purpose was to investigate the relationship of the sex of assessees and assessors to the usage of this information. To do this, data were gathered from an ongoing assessment center concerning 2191 assessees' i) overall assessment rating, ii) ratings on eighteen dimensions thought to be relevant to success as an entry level manager, iii) sex, and iv) assessor group male to female ratio.

Preliminary analyses indicated that the eighteen dimensional ratings predicted the OAR very accurately. It was further noted that a much smaller set of dimensions have substantially the same predictive accuracy within this data set. Factor analyses of the eighteen dimensions failed to provide support for the a priori categorical groupings of dimensions assessors claimed they were using. Support was found for a three factor simple structure of the eighteen

dimensions. The OAR was subsequently regressed onto subscale scores for the a priori category groupings and the three factor analytically derived groupings. Sex of the assessee and the male to female ratio of the assessor group were examined for their moderating effects on these regressions.

Discussion of Research Question 1

Predictive Power

The ability of various combinations of the eighteen assessment dimensions to predict the OAR was summarized in Table 4. Interpretation of the regression coefficients (other than whether they are significantly different from zero) shall not be attempted here due to the multicollinearity among the dimensions as indicated in Table 3. Two of the results are of particular merit. First, over 80% of the variance in the OAR is accounted for by the consensus ratings on the eighteen dimensions. A multiple correlation of .90 is very large compared to those found in most behavioral science research. One explanation of the predictive power, however, would seem to be the simple fact that the assessors were trained to arrive at the CAR through the evaluation of the dimensional ratings. It should come as no great surprise, then, that these dimensions predict the majority of the variance in the OAR.

Second, fifteen dimensions account for the same amount of variance in the OAR as all eighteen dimensions. Further, over 92% of the variance accounted for by all eighteen dimensions is accounted for by only five dimensions (four of which have the highest simple correlations with OAR in Table 3). These five dimensions use only the In-Basket, Case Proposal, and Group Meeting as primary sources of information (see Table 2). At first glance, it would appear as if the OAR could be arrived at with almost equal accuracy by disregarding the Fact Finding Interview and paper and pencil tests. In addition, it would appear that the assessors could omit the effort necessary to rate thirteen of the dimensions.

It must be noted, however, that five of the eighteen dimensions (needing primarily only three of the five exercises) predict over 92% of the variance in an OAR from an assessment center in which all eighteen dimensions have been rated and all five exercises have been required. An alternative explanation yet to be eliminated is that the predictive power found in the five dimensions may only be obtained in an assessment center that is 'rich' in opportunities to observe a variety of behaviors (i.e. one that has five exercises and requires that eighteen dimensions be rated). Prior to making the interpretation that equal information can be obtained from an assessment center that

has only three exercises in which five dimensions are rated, an examination of the importance of the other thirteen dimensions in arriving at accurate ratings for the five remaining dimensions would be required.

Underlying Structure of the Eighteen Dimensions

The factor analyses reported were performed in an attempt to disclose any simple structure underlying the eighteen dimensions. A priori it would be hypothesized that the structure would conform to the four categories of dimensions purportedly given equal weight by the assessors in arriving at an OAR. When the dimensions were forced to load onto four factors, only the Problem Solving Skills category loaded cleanly onto one factor (in Tables 9, 10, & 11 all the dimensions except for Fact Finding- Oral load onto either factor 2, 1, or 4, respectively). Neither of the two dimensions making up the Personal Qualities category load together in any of the rotations. Awareness of Social Environment and Behavior Flexibility consistently load together on a factor different from the rest of the dimensions in the Interpersonal Skills category. Finally, Oral Presentation and Oral Defense always load together on a factor different from the Written Communications and Scholastic Aptitude paper and pencil scored dimensions of the Communications Skills category. It can be concluded that

the underlying empirical relationships between the eighteen dimensions do not reflect the category groupings the assessors are being trained to use.

When the dimensions were allowed to load only on those factors with eigen values greater than one, a more parsimonious structure emerges. As noted in Table 8, the varimax rotation and oblimin rotations found the dimensions rated by paper and pencil instruments loading on Factor 3 and the Problem Solving Skills dimensions (with the exception of Fact Finding-Oral) loading on Factor 2. Fact Finding-Oral loaded almost equally on all three factors. This can be interpreted as meaning that oral fact finding skills are required for i) engaging in interpersonal interactions, ii) solving problems, and iii) responding to paper and pencil instruments. The Communications Skills dimensions not rated by paper and pencil instruments (Oral Presentation and Oral Defense), the Interpersonal Skills dimensions, and the Personal Qualities dimensions all loaded onto Factor 1 when varimax rotation was used. Oblimin rotation left out one of the Personal Quality dimensions (Self-Objectivity), though it loaded highest on Factor 1. As noted in chapter 4, it seems intuitive that a person's personal qualities, interpersonal skills, and oral communication skills would be highly related. The loading of the dimensions rated by paper and pencil instruments on the weakest factor by themselves con-

forms with the assessors' statements in their interviews. Further, it does not seem counter intuitive to speculate that the Factors 1 & 2 might represent the 'categories' the assessors maintain they are using interactively, since Interpersonal Skills dominates Factor 1 and Problem Solving Skill dimensions are only found in Factor 2.

It does not seem unreasonable to speculate that the three factor solution represents a close approximation to any simple structure that might be underlying the ratings on the eighteen dimensions. Acting on this speculation, subscale scores (averages of dimensions) were generated for the dimensional loadings presented in Table 8 for the varimax rotation.

Use of the Subscale Scores

Regressions of the CAR onto i) the category subscale scores and ii) the category subscale scores and Interpersonal Skill by Problem Solving Skill interaction are presented in Table 12 (columns A & B). Category subscale scores were used because, even though the category groupings do not reflect the empirical relationships between dimensions, the assessors could still be using the categories to reach OARs. There is no difference in variance accounted for, the regression coefficients for the category subscale scores do not change substantially, and the regression coe-

ficient for the interaction term is negative and nonsignificant across the two equations. Thus, though a substantial proportion of variance is explained by the category subscale scores, the way in which the assessors maintain they integrate the categories is not supported. Indeed, the interaction term, though nonsignificant, was predicted to have been positive.

Regressions of the OAR onto i) the factor subscale scores and ii) the factor subscale scores and Factor 1 by Factor 2 interaction are presented in Table 13 (columns A & B). Again, the OAR variance accounted for does not differ between the two equations. Further, the coefficients for the factor subscale scores do not differ substantially between the two equations. However, the interaction term between the factors dominated by the Interpersonal Skills and Problem Solving Skills dimensions is significant and negative. The interaction term should have been positive if it was to be congruent with the assessors' contentions. One should note, however, that the absolute value of the regression coefficient for the interaction term is not large. In light of the large degrees of freedom in the error term, the significance of this coefficient may not be too meaningful.

Discussion of Research Question 2

Assessee Sex Effects

Regression equations relevant to the relationship of assessee sex to the way in which assessors use the dimensional ratings are found in Tables 12 & 13. The coefficients generated by equation C in Table 12 indicate that sex of the assessee and its interaction with Interpersonal Skills, Problem Solving Skills, and the Interpersonal Skills by Problem Solving Skills interaction are significantly different from zero. Further, the sex of the assessee by Interpersonal Skills by Problem Solving Skills interaction is significant. The Interpersonal Skills by Problem Solving Skills interaction term, that was not significant in equation B, attains significance in equation C at $p < .01$. Finally, the regression coefficient for the Personal Qualities category increases but becomes nonsignificantly different from zero at $p < .05$.

At this point it would be extremely speculative to infer that the sex of the assessee was influencing assessors' uses of dimensional ratings. Equation B yields no support for the assessors' contention that they use equal weights for the four categories and one interaction term. The one thing that can be inferred from equation C is that the sex of the assessee i) is related to the OAR and ii) is related to the way in which the category subscale scores are

related to the OAR. Whether this finding is indicative of the effect of assessee sex on the internal cognitive processes of the assessor is unclear. The policy implications of these results shall be discussed in the last section of this chapter.

The coefficients generated by equation C in Table 13 (using factor subscale scores) indicate that sex of the assessee and its interaction with the Interpersonal Skills factor, the Problem Solving Skills factor, and the Interpersonal Skill by Problem Solving Skills factor interaction are significantly different from zero. The Interpersonal Skill factor by Problem Solving factor changes sign and increases greatly in size (while remaining significantly different from zero at $p < .0001$).

Since assessee sex was dummy coded 1 for male and 2 for female, an interpretation of the one-way interactions with assessee sex is that Interpersonal Skills and Problem Solving Skills are given greater weight by the assessor for males in arriving at their OAR than for females. In agreement with Table 14, the main effect for assessee sex is negative, indicating that males tend to receive higher OARs than females. The assessee sex by Interpersonal Skills by Problem Solving Skills interaction indicates that the female OARs increase more as a function of the Interpersonal Skills by Problem Solving Skills interaction than males. In order

to determine exactly how the OAR was related to this interaction, the OAR was regressed onto the factor subscale scores and the Interpersonal Skills by Problem Solving Skills interaction separately for each sex. All of the regression coefficients were significantly different from zero at $p < .0001$ except for the Interpersonal Skills by Problem Solving Skills interaction for females ($p = .35$). The coefficient for this interaction term for males was negative and significant. Consequently, it seems that, while the females' OARs are not related to the interaction, males' OARs are lower as a result of it.

Table 15 contains the predicted mean OAR and standard deviation from four separate regression equations. The OAR was regressed onto i) the factor subscale scores and ii) the factor subscale scores and the Interpersonal Skills by Problem Solving Skills interaction separately for males and females. Then, the four regression equations were used to generate predicted OAR means for both males and females in the sample. These means yield insight into how great the impact of assessee sex is on the way in which assessors use the factor subscale scores.

On average, the means in Table 15 indicate that females' OARs are lower by .02 (on a four point scale) because of sex differences in the way in which assessors use the subscale scores (with or without the interaction term).

On average, the means in Table 15 indicate that males' OARs are greater by .01 because of sex differences in the way in which assessors use the subscale scores (with or without the interaction term). These small differences indicate that the real effect of differential factor subscale score usage for male and female assessees, on average, is minimal. However, a comparison of the predicted means for male and female assessees for any one regression equation reflects a much larger difference (.23 to .24 depending on which equation is used). This difference is due to male and female assessees receiving different dimensional ratings and hence different factor subscale scores. Frequency tables of assessee sex by dimensional rating indicate that for fifteen of the eighteen dimensions, females receive significantly fewer high ratings and more low ratings than males ($p < .001$ for all chi-square statistics).

Finally, it is of interest to note the change in direction of the regression coefficient for the Interpersonal Skills by Problem Solving Skills interaction term in equation C of Table 13 when assessee sex and its interactions are included. There are two possible explanations for this change. First, with the addition of the assessee sex by Interpersonal Skills by Problem Solving Skills interaction, the negative moderating influence of assessee sex may have been taken out of the original two-way interaction term. In

the absence of this moderating effect, the 'true' way in which the Interpersonal Skills by Problem Solving Skills interaction is used by the assessors (as per their contentions) may now be being reflected in its regression coefficient.

Second, as noted in reference to Tables 3 and 4, there is a great deal of multicollinearity among the dimensions. In fact, the range of correlations among the factor subscale scores is from .625 to .372. An alternative explanation for the change in the Interpersonal Skills by Problem Solving Skills interaction is that it reflects nothing more than the increased regression coefficient variation found under conditions of multicollinearity.

Assessor Group Sex Effects

Table 16 contains the coefficients resulting from the regression of the OAR onto the category subscale scores, the interaction between the Interpersonal Skills and Problem Solving Skills categories, a dummy coded variable reflecting the number of females on the four person assessor team (GSEX), and all interactions between GSEX and the previous variables entered into the equation. The sample size has decreased by 15 due to missing assessor sex information. As in equation C of Table 12, it is difficult to ascribe much meaning to these coefficients. In addressing Research Ques-

tion 1 it was concluded that the underlying simple structure of the eighteen dimensions did not reflect the a priori categories. Further, when the OAR was regressed onto the category subscale scores, no support was found for the assessors' contentions that they used an interaction term between the Interpersonal Skills and Problem Solving Skills dimensions. Consequently, at best it can be said that the number of females on the assessor team is related to the assessors' use of the Problem Solving Skills category and the Interpersonal Skills by Problem Solving Skill interaction term.

Table 17 contains the coefficients resulting from the regression of the OAR onto the factor subscale scores, the interaction between the Interpersonal Skills and Problem Solving Skills categories, a dummy coded variable reflecting the number of females on the four person assessor team (GSEX), and all interactions between GSEX and the previous variables entered into the equation. Of the three factors and one interaction term the assessors contend that they use to arrive at the OAR, in the presence of GSEX and the GSEX by factor subscale score interaction terms, the Problem Solving Skills factor and the Interpersonal Skills by Problem Solving Skills factor regression coefficients are not significantly different from zero. The only other variables to yield significant regression coefficients were the GSEX by Interpersonal Skills interaction, the GSEX by paper and

pencil instruments interaction, and the GSEX by Interpersonal Skills by Problem Solving Skills interaction.

It is as difficult to interpret the coefficients in Table 17 as it was those in Table 16. The assessors' contentions concerning their use of the categories which dominate the three factors were not confirmed. However, it does appear that the male to female ratio of the assessor group has some real impact on the OAR. To examine this impact, the OAR was regressed onto i) the three factor subscale scores and ii) the three factor subscale scores and Interpersonal Skills by Problem Solving Skills factor interaction for each of the five possible combinations of males and females in the assessor group. Table 18 contains the predicted mean OARs and standard deviations for the resulting ten equations. It is of interest to note how the predicted means change as the sexual composition of the assessor group on which the equation was derived changes. In the two sets of five by five matrices of means, there are forty possible paired comparisons between the rows. Comparing any pair of these means amounts to a comparison of the predicted OAR for an assessor group with X females in it to the predicted OAR for an assessor group with $X-1$ females in it. Of the forty such comparisons, five are ties (the predicted means are equal). In only three of these comparisons are the predicted OARs from assessor groups with more females greater

then the predicted OARs from assessor groups with less females. Using a binomial test of the hypothesis that the male to female ratio of the assessor group is not related to the predicted value of the OAR ($p = .5$, $n = 35$, $a = 3$), the probability of finding only three comparisons in which the female dominant assessor group receives a higher predicted OAR is less than .00001.

Consequently, it can be concluded that the male to female ratio of the assessor group is having a real impact on the OAR received by the assessee. On average, this impact will not be large (rounding all of the predicted values to the nearest integer rating would yield a matrix containing nothing but twos). The main effect of GSEX in Table 17 is not significant, reinforcing the conclusion that this impact will be minimal.

CHAPTER VI
LIMITATIONS AND IMPLICATIONS

Limitations

The limitations of the study were presented in Chapter 1. At this point, it is important to examine again the third limitation discussed earlier.

The regression analyses performed to address Research Questions 1 & 2 were attempts at obtaining a mathematical representation of the way in which assessors use the information gathered in an assessment center to arrive at an OAR. The study started with the staff directors' descriptions of the way in which they train assessors to integrate and combine the assessment dimensions. A mathematical representation of this was examined as a first step towards understanding assessors' usage of information. Regression coefficients were examined in an attempt to evaluate the assessors' contentions concerning their information usage. Herein lies the limitation.

Because this data set was obtained from an assessment center operating in a real organization, assessors were not faced with an equal number of all the possible profiles of

assessee dimensional ratings. Consequently, multicollinearity precludes any test other than the test of the hypothesis that the regression coefficients are significantly different from zero. Further, the study is necessarily conforming to the dictums of Egon Brunswick and his lens model. The phenomena of assessors' decision making behavior was examined in its' natural environment, not in a laboratory. Brunswick contends that when human behavior is examined outside of its natural setting, any phenomena that are documented shall be limited in their generalizability to the laboratory setting in which they occur (Hammond, 1980). This contention was recently given substantial support by Ebbesen and Konecni (1975).

Ebbesen and Konecni (1975) compared the way in which real felony court judges used information to arrive at a bail bond decision in actual court cases to the way they used information in simulated cases. Substantial differences were found.

On the other hand, another camp of researchers contend that it is impossible to determine the true way in which information is used by decision makers unless some control is maintained over the decision environment. They contend that without equal cell sizes (an equal number of assessee dimensional profiles in each of the 18 X 5 possible cases) the true way in which the information is used will be masked

by the lack of independence among the 'treatments' or dimensions (Anderson, 1974).

This contrast in approaches to the examination of human decision process is brought up here in order to underscore the fact that this study examined assessors' usage of information in real decision situations that were of necessity nonorthogonal in design. Because of this, whether conclusions can be drawn concerning the relative importance of the information available to the assessor is problematical. This includes conclusions concerning the impact of sex of the assessee and the male to female ratio of the assessor group.

Implications

Future Research

The current study examined one stage of a sequential decision process. Assessors must i) decide what rating to give each assessee on any particular dimension, ii) come to a consensus regarding that rating, iii) individually decide upon an OAR based on the eighteen dimensional ratings, and iv) arrive at a consensus regarding that OAR. The relationship between the consensus dimensional ratings and the consensus OAR (a combination of iii and iv) was examined here. Future research must examine all four steps and the components of the decision environment (e.g. assessee sex) that influence them.

It is important that these four steps in the decision process be examined in order that the way in which information is used in the assessment center be made congruent with the way in which information is used in evaluating job performance. This is the argument, made in Chapter 1, that the content validity of an assessment center cannot be evaluated until the decision processes (scoring rules) used in rating behavior in the assessment center can be compared to those used in performance appraisal.

Another argument for examining the decision processes used in the assessment center is its importance in evaluating the construct validity of an assessment center. Content validity is the evaluation of the inference that procedures by which behaviors are sampled and scored in an assessment center are representative of those used to sample and score actual job performance behaviors. Construct validity of an assessment center is the evaluation of the inference that the center is measuring some construct, in this case effective managerial performance.

The final argument for examining the cognitive processes of assessors in assessment centers is for the evaluation of the predictive validity of assessment centers. The two ways to ensure that assessment centers are not measuring career ladder climbing skills are to i) increase the quality of the performance criteria in predictive validity studies

and/or ii) examine the cognitive process of the assessors and performance appraisers.

Implications for Practice

There is one major implication for the actual conduct of assessment centers in the future that comes out of this study. The way to ensure that the information gathered in an assessment center is appropriately combined into an overall rating is to take the decision out of the hands of the assessors and, instead, use an arithmetic scoring rule. After the assessors had arrived at a consensus rating for all eighteen dimensions, the assessee's profile of ratings could be given to a clerical worker who would simply combine the ratings by some predetermined rule to arrive at the OAR. In the assessment center under examination here, the averages for the dimensions in each category could be added to the product of the Interpersonal Skills by Problem Solving Skills interaction (adjusted to a 1 - 5 point scale), the sum converted to a four point scale, and the OAR would be determined by rounding to the nearest integer.

There are at least two advantages in the substitution of a decision rule in place of assessors arriving at the OAR. First, human error in the integration of assessment center information to arrive at an OAR would be all but eliminated. All that would remain would be the possibility of

an arithmetic error on the part of the clerical worker. This would eliminate one potential source of error for which the firm could be penalized under the law (e.g. sex, race, and age bias).

Second, it would decrease the cost of running the assessment center. The less costly clerical worker's time would be substituted for that of the four assessors. This would amount to approximately fifteen to twenty minutes per assessee in the current assessment center.

In light of this recommendation, any number of scoring rules could be used. Any of the equations presented in this study in which the OAR is regressed onto some combination of the assessment dimensions would do. Given the relative complexity of these equations (imagine having to calculate a predicted OAR by hand with any of these equations), the OAR was regressed onto a simple arithmetic average of the eighteen dimensions. The squared correlation coefficient for this regression was .782. The substitution of a simple averaging model for assessor integration of the consensus dimensional ratings would compare favorably with any other regression performed in this study (the highest squared correlation previously obtained is .802).

APPENDIX A
ASSESSMENT CENTER DEFINED

To be considered as an assessment center, the following minimal requirements must be met:

- i) 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.
- ii) Multiple assessors must be used. These assessors must receive training prior to participating in a center.
- iii) Judgments resulting in an outcome (i.e., recommendation for promotion, specific training or development) must be based on pooling information from assessors and techniques.

- iv) An overall evaluation of behavior must be made by the assessors at a separate time from assessors and techniques.
- v) Simulation exercises are used. These exercises are developed to tap a variety of predetermined behaviors and have been pre-tested prior to use to insure that the techniques provide reliable, objective, and relevant behavioral information for the organization in question.
- vi) The dimensions, attributes, characteristics, or qualities evaluated by the assessment center are determined by an analysis of relevant job behaviors.
- vii) The techniques used in the assessment center are designed to provide information which is used in evaluating the dimensions, attributes, or qualities previously determined.

In summary, an assessment center consists of a standardized evaluation of behavior based on multiple inputs. Multiple trained observers and techniques are used. Judgments about behavior are made, in part, from specially developed assessment simulations.

These judgments are pooled by the assessors at an evaluation meeting during which all relevant assessment data are reported and discussed, and the assessors agree on the eval-

uation of the dimensions and any overall evaluation that is made.

APPENDIX B
STAFF-ASSESSOR JOE DESCRIPTION

Job Summary

Serves as a team member to individually and collectively evaluate potential of non-management employees for promotion to first-level management assignments through utilization of [an assessment center process].

Conducts a variety of assessment techniques and evaluates effectiveness of observed behavior. Records and categorizes behavioral information from multiple sources and prepares detailed reports on assessees behavior in a variety of activities. Coordinates and conducts in-depth feedback interviews and post-assessment action plan with candidates and their first through third level managers.

Job Duties and Responsibilities

- i) (25%) Observes and records candidate behavior
(Average of 12 candidates each week).
 1. Observes candidates in a group problem exercise. Takes comprehensive notes of candidate's behaviors. This requires an understanding of group dynamics, verbal and nonverbal behavior,

the ability to manually record conversations, and knowledge of relationships to the dimensions evaluated.

2. Assures that the group exercise process is uniformly administered and that the candidates adhere to standardized instructions, time frames, and use of materials.
3. Observes candidates while conducting one-on-one role playing interviews. Interacts with candidates in a uniform manner, providing needed facts, answering questions, clarifying answers and challenging responses, while simultaneously observing and recording behavior.

ii) (25%) Prepares written reports.

1. Prepares comprehensive, narrative reports on behaviors observed during the group problem.
2. Prepares reports documenting the behaviors observed during the one-on-one interviews.
3. Prepares reports on results and climate of feedback.

iii) (25%) Participates in evaluation of candidates, arriving at a determination of the advancement potential.

1. Orally presents reports of observed candidate behavior while categorizing and synthesizing

information presented by three other assessors, along with results of paper and pencil tests.

2. Makes independent judgments on 18 managerial dimensions for each candidate. These 18 factors encompass Communication Skills, Problem-Solving Skills, and Interpersonal as well as Personal Skills.
3. Explains, elaborates, and defends all judgments on individual rating of assessee, which must culminate with a group consensus rating for each dimension.
4. Independently arrives at an overall rating or prediction of success in a first-level assignment. This requires mentally assimilating a large quantity of behavioral information obtained during the assessment day and integrating and comparing it with first-level performance requirements.

iv) (25%) Schedules, prepares, and conducts detailed, in-depth feedback interviews with candidates and their supervisors. Also presents Post Assessment Action Plan.

1. Explains philosophy and use of assessment, use of techniques, definitions of dimensions, and candidates's performance.

2. Defends all staff judgments and backs them up with specific examples of the individual's performance at the assessment center.
3. Conducts feedback sessions with second- and third- level supervisors. Speaks about the assessment process in detail. Presents the candidate's performance to the supervisor, explains its implications, use of the results, and prepares the supervisor to discuss results with the candidate. Deals constructively with developmental and career movement issues.
4. Presents post assessment action plan. This provides a tool for candidate to consider in determining how to go about strengthening their individual areas of weakness.
5. Prepares comprehensive reports to management summarizing the candidate's performance at the assessment center and documenting the reasons for the final rating. These reports are used by management to make promotion and development decisions and must be a clear, concise, and accurate reflection of the candidate's strengths and weaknesses.

Scope and Nature of Supervision

- i) Reports to a Staff Director along with 3 or 4 other assessors. No subordinates report to the Staff Assessor.
- ii) The assessment process is quite structured, with little allowance for deviation from prescribed administration. A four week training period is provided to qualify as an assessor.
- iii) Some line management experience is desired, as the assessor is making judgments regarding an individual's probability of succeeding in first-level management jobs. Judgments are based on knowledge of the performance required for a wide variety of first-level jobs.

APPENDIX C
STAFF DIRECTOR JOB DESCRIPTION

Job Summary

Serves as both team director and as team member to individually and collectively evaluate potential of nonmanagement employees for promotion to first-level management assignments through utilization of the assessment center process. Trains subordinate team members in administration of this process.

Conducts a variety of assessment techniques and evaluates effectiveness of observed behavior.

Directs the provision of in-depth feedback interviews and post-assessment action plan with candidates and their first through third level managers.

Job Duties and Responsibilities

- i) (10%) Conducts 4 weeks of training for the assessment staff. Each team is made up of one director and 3 staff members of a varied race/sex mix. A team will serve in assesment for approximately 12 months; a director's rotation is 1-2 years.

1. Provides training in each phase of the center's operation. Makes opening and closing remarks to the candidates. Demonstrates techniques for conducting one-on-one interviews. Observes candidates and compiles reports of their participation throughout the group activities. Trains staff-assessors in the readout and evaluation process, the writing of summary reports, providing of feedback to candidate and supervisors, administering a development tool for candidates who have been through assessment, and the writing of feedback reports.
 2. Observes teams in the process of assessment to determine effectiveness of training and standardization of application. Adjusts training program as appropriate after follow up.
- ii) (35%) Serves as team member as assessees are being observed (average of 12 candidates each week).
1. Observes candidates in group problem solving exercises. This requires an understanding of group dynamics, verbal and nonverbal behavior, and knowledge of relationships to the dimensions evaluated.
 2. Assures that the group exercise process is uniformly administered and that candidates adhere

to standardized instructions, time frames, and use of materials.

3. Observes candidates and team members during one-on-one role playing interviews. Interacts with candidates, providing needed facts and answering questions, while simultaneously observing behavior.
4. Provides any needed direction to staff subordinates as they prepare comprehensive, narrative reports on behaviors observed during the group problem and/or the one-on-one interview.

iii) (35%) Participates in evaluation of candidates, arriving at a determination of the advancement potential.

1. Presides over evaluation session as team subordinates orally present reports of observed candidate behavior. Categorizes information presented by team members, along with results of paper and pencil tests, to independently arrive at an overall rating or prediction of success in a first-level assignment. This requires mentally assimilating a large quantity of behavioral information obtained during the assessment day and integrating and comparing it with first-level performance requirements.

Eighteen managerial dimensions are considered for each candidate. These 18 factors encompass Communication Skills, Problem Solving Skills, Interpersonal Skills, and Personal Skills.

2. Explains or defends all judgments on individual rating of each assessee, which must culminate with a group consensus rating for each dimension. Directors rating is the deciding factor in resolution of any "deadlock" on scoring each dimension.

iv) (20%) Provides direction in administration of the feedback process and post assessment action plan.

Provides education to the field on objectives and purpose of the assessment center.

1. Presents programs of education to Company managers. This may involve speaking at departmental seminars or meetings, or it may be on a one-to-one basis. (Includes history of assessment, assessment components, objectives, target populations, use of results, future direction, etc.).

APPENDIX D
TABLES

Table 1

Sample Representation of Total Number
Assessed by Overall Rating and Occurance
of Promotion*

	Total**		Promoted		High		Moderate		Low	
	B	W	B	W	E	W	B	W	B	W
N	198	715	35	91	34	226	57	217	107	272
%	-	-	18	13	17	31	29	30	54	38
B _W /W _B			1.39		.55		.93		1.43	

* Taken from Huck, J. R. Determinants of assessment center ratings for white and black females and the relationship of these dimensions. Unpublished doctoral dissertation, Wayne State University, Detroit, Michigan, 1974.

** Figures include only those females assessed from July 1966 through June 1971 at the Michigan Bell Personnel Assessment Program.

Table 2
Dimension Sources

Dimensions by Category	Exercises				
	Fact Finding Interview	Case Materials In-Basket	Case Material Proposal	Group Meeting	Paper and Pencil Instruments
Personal Qualities					
Energy	B	B	B	B	
Self-Objectivity					A
Interpersonal Skills					
Impact	B		B	A	
Leadership				A	
Awareness of Social Environment	B		B	A	B
Behavior Flexibility	B		B	A	
Autonomy			B	A	
Problem-Solving Skills					
Fact Finding--Oral	A			B	
Fact Finding--Written		A	A	A	
Organizing		A	B		
Interpreting Information		E	A	B	
Planning		A	B		
Decisiveness		A	B		
Decision Making		A	B		
Communication Skills					
Oral Presentation			B	A	
Oral Defense			A	B	
Written Communication					A
Scholastic Aptitude					A

A = Primary source for observing relevant behavior.

B = Additional source for observing relevant behavior.

Table 3

Correlation Matrix of Consensus Ratings

	Overall Rating	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10	D11	D12	D13	D14	D15	D16	D17	D18	
D1 Energy	.62		.44	.67	.57	.44	.53	.59	.39	.41	.49	.43	.45	.22	.39	.51	.46	.25	.34	
D2 Self-Objec- tivity	.53			.39	.42	.43	.39	.31	.36	.39	.38	.39	.35	.22	.37	.37	.31	.21	.28	
D3 Impact	.56				.61	.43	.53	.66	.31	.32	.33	.36	.31	.18	.32	.46	.45	.15	.27	
D4 Leadership	.66					.55	.60	.58	.31	.35	.35	.41	.33	.16	.37	.46	.47	.15	.29	
D5 Awareness of Social Environment	.59						.61	.38	.40	.37	.34	.46	.38	.13	.41	.42	.42	.21	.28	
D6 Behavior Flexibility	.64							.48	.38	.33	.35	.46	.33	.13	.36	.44	.49	.17	.29	
D7 Autonomy	.54								.29	.36	.30	.37	.33	.22	.34	.41	.53	.12	.25	
D8 Fact Finding-- Oral	.52									.37	.35	.42	.36	.12	.36	.32	.36	.24	.31	
D9 Fact Finding-- Written	.59										.49	.64	.47	.24	.52	.35	.43	.27	.39	
D10 Organizing	.60											.51	.62	.28	.50	.37	.37	.25	.33	
D11 Interpreting Information	.69												.54	.25	.58	.42	.51	.31	.44	
D12 Planning	.61													.29	.63	.39	.41	.24	.33	
D13 Decisiveness	.32														.38	.13	.19	.06*	.12	
D14 Decision Making	.63															.37	.44	.22	.32	
D15 Oral Presenta- tion	.57																	.43	.25	.36
D16 Oral Defense	.61																		.19	.32
D17 Written Communication	.36																			.52
D18 Scholastic Aptitude Test	.50																			

*All correlations are significant at the .0001 level except this one, where $p = .0072$.

Table 4

Multiple Regression Analyses:
Dependent Variable = Overall
Assessment Rating

N=2191	All 18 Dimensions	Stepwise (p=.01)	Regression Coefficients			
			2	3	4	5
D1	Energy	.013				
D2	Self-Objec- tivity	.069**	.071*			
D3	Impact	.023	.036*			
D4	Leadership	.188**	.195*	.416**	.376**	.351**
D5	Awareness of Social Environment	.021				
D6	Behavior Flexibility	.143**	.152*			.222**
D7	Autonomy	.017				
D8	Fact Finding --Oral	.093**	.096*			
D9	Fact Finding --Written	.054*	.056*			
D10	Organizing	.092**	.092*	.271**	.217**	.206**
D11	Interpreting Informaticn	.141**	.141*	.553**	.429**	.337**
D12	Flanning	.085**	.087*			
D13	Decisiveness	.060**	.061*			
D14	Decision Making	.109**	.110*			.209**
D15	Oral Presenta- tion	.081**	.085*			
D16	Oral Defense	.102**	.107*			
D17	Written Communication	.045**	.045*			
D18	Scholastic Aptitude Test	.066**	.065*			
R Square		.802**	.802**	.646**	.694**	.717**

* = significant at $p \leq .001$

** = significant at $p \leq .0001$

Table 5

Principal Component Analysis varimax
rotated Factor Pattern matrix: Three
Factor Rotation*

	Factors			
	1	2	3	
D1	.68050	.31020	.19337	Energy
D2	.41133	.33662	.21230	Self-Objectivity
D3	.78649	.15237	.07567	Impact
D4	.75435	.19870	.11157	Leadership
D5	.55217	.29264	.21589	Awareness of Social Environment
D6	.67667	.22288	.17483	Behavior Flexibility
D7	.70698	.21955	.03586	Autonomy
D8	.33370	.31926	.30258	Fact Finding--Oral
D9	.26720	.57014	.30993	Fact Finding--Written
D10	.26589	.61571	.22408	Organizing
D11	.32834	.60494	.36419	Interpreting Information
D12	.23516	.70807	.19147	Planning
D13	.11719	.40480	-.02728	Decisiveness
D14	.24953	.73132	.15918	Decision Making
D15	.50966	.26642	.27671	Oral Presentation
D16	.51908	.36899	.18374	Oral Defense
D17	.08247	.11498	.65736	Written Communications
D18	.21512	.20574	.68494	Scholastic Aptitude Test
Eigen- values	7.48	1.63	1.22	

* Communalities were used on the diagonal.

Table 6

Principal Components Analysis Oblimin Rotated
 Factor Pattern Matrix: Three Factor Rotation
 ($\delta=0$)*

	Factors			
	1	2	3	
D1	.68884	.09235	.05017	Energy
D2	.34532	.22090	.11071	Self-Objectivity
D3	.89684	-.11818	-.07087	Impact
D4	.83432	-.05628	-.03366	Leadership
D5	.53435	.11345	.09934	Awareness of Social Environment
D6	.71996	-.01064	.04536	Behavior Flexibility
D7	.78002	.00282	-.11397	Autonomy
D8	.24249	.20942	.22710	Fact Finding--Oral
D9	.06503	.54356	.20241	Fact Finding--Written
D10	.05956	.61835	.09894	Organizing
D11	.11859	.55280	.24683	Interpreting Information
D12	-.00794	.75099	.05149	Planning
D13	-.00450	.46408	-.12022	Decisiveness
D14	.00595	.78119	.00908	Decision Making
D15	.48257	.08409	.17869	Oral Presentation
D16	.46990	.22704	.05558	Oral Defense
D17	-.04416	-.02432	.69544	Written Communications
D18	.08027	.03333	.69266	Scholastic Aptitude Test
Eigen- values	7.48	1.63	1.22	

* Communalities were used on the diagonal.

Table 7

Principal Components Analysis Oblimin Rotated
 Factor Pattern Matrix: Three Factor Rotation
 ($\delta = -1$)*

	Factors			
	1	2	3	
D1	.67934	.11330	.07782	Energy
D2	.36562	.21436	.13517	Self-Objectivity
D3	.84948	-.06740	-.04916	Impact
D4	.79819	-.01435	-.01027	Leadership
D5	.53453	.12515	.12292	Awareness of Social Environment
D6	.69640	.02117	.06729	Behavior Flexibility
D7	.74997	.63792	-.08765	Autonomy
D8	.26929	.19805	.24635	Fact Finding--Oral
D9	.13843	.49411	.23813	Fact Finding--Written
D10	.13846	.56305	.14043	Organizing
D11	.19307	.50431	.28444	Interpreting Information
D12	.08789	.68115	.09999	Planning
D13	.04883	.42277	-.08872	Decisiveness
D14	.10345	.70970	.06043	Decision Making
D15	.48389	.69527	.19789	Oral Presentation
D16	.48468	.22603	.08498	Oral Defense
D17	-.01867	-.03241	.68945	Written Communications
D18	.10853	.02546	.69052	Scholastic Aptitude Test
Eigen- values	7.48	1.63	1.22	

* Communalities were used on the diagonal.

Table 8

Subscale Loadings from
Common Factor Analyses

I. Varimax Rotation: Loading Rule of 45/40

Factor 1	Factor 2	Factor 3
Energy	Fact Finding--Written	Written Communication
Self-Objectivity	Organizing	Scholastic Aptitude
Impact	Intrepreting Information	
Leadership	Planning	
Awareness of Social Environment	Decisiveness	
Behavior Flexibility	Decision Making	
Autonomy		
Oral Presentation		
Oral Defense		
Omitted: Fact Finding--Oral		

II. Oblimin Rotation (Delta = -1, 0): Loading Rule of 40(45)/30(25)
Respectively

Factor 1	Factor 2	Factor 3
Energy	Same as Varimax	Same as Varimax
Impact		
Leadership		
Awareness of Social Environment		
Behavior Flexibility		
Autonomy		
Oral Presentation		
Oral Defense		
Omitted: Self-Objectivity Fact Finding--Oral		

Table 9

Principal Components Analysis Varimax
Rotated Factor Pattern Matrix: Four Factor
Rotation*

	Factors				
	1	2	3	4	
D1	.67394	.32684	.19511	.21677	Energy
D2	.30138	.31843	.32351	.18649	Self-Objectivity
D3	.80425	.16419	.19021	.10180	Impact
D4	.61667	.18025	.44057	.09034	Leadership
D5	.30521	.21978	.68368	.13675	Awareness of Social Environment
D6	.47736	.16720	.59621	.11635	Behavior Flexibility
D7	.72357	.23832	.16037	.05838	Autonomy
D8	.20845	.29538	.33936	.26904	Fact Finding--Oral
D9	.19037	.56143	.23581	.29111	Fact Finding--Written
D10	.22584	.61857	.16113	.22090	Organizing
D11	.19076	.58202	.37084	.32512	Interpreting Information
D12	.18352	.70375	.18190	.18238	Planning
D13	.14197	.42184	-.01276	-.01539	Decisiveness
D14	.16517	.71467	.25140	.13734	Decision Making
D15	.42525	.26093	.29018	.26593	Oral Presentation
D16	.42495	.35865	.31416	.16856	Oral Defense
D17	.05839	.12054	.08131	.66256	Written Communications
D18	.17664	.21036	.13982	.69777	Scholastic Aptitude Test
Eigen- values	7.48	1.63	1.22	.90	

* Communalities were used on the diagonal.

Table 10

Principal Components Analysis Oblimin Rotated
 Factor Pattern Matrix: Four Factor Rotation
 (delta=0) *

	Factors				
	1	2	3	4	
D1	.13854	-.66430	.13470	.00109	Energy
D2	.22303	-.16239	.09165	-.25318	Self-Objectivity
D3	-.06534	-.86795	.02677	-.00785	Impact
D4	-.01908	-.53884	-.02727	-.36480	Leadership
D5	.07671	-.05317	-.01379	-.72762	Awareness or Social Environment
D6	-.01923	-.31043	-.01877	-.58964	Behavior Flexibility
D7	.05840	-.76542	-.02625	.01443	Autonomy
D8	.20039	-.04458	.18943	-.28354	Fact Finding--Oral
D9	.53010	.00018	.19089	-.11321	Fact Finding--Written
D10	.61318	-.06154	.11414	-.01502	Organizing
D11	.53264	.05792	.20188	-.27312	Interpreting Information
D12	.73280	.01628	.05453	-.04471	Planning
D13	.47160	-.08142	-.08804	.11108	Decisiveness
D14	.75359	.06617	-.01018	-.14240	Decision Making
D15	.10605	-.33678	.19271	-.17730	Oral Presentation
D16	-.24263	-.30853	.06219	-.20971	Oral Defense
D17	-.02959	.01518	.71570	.04798	Written Communications
D18	.03168	-.08461	.72657	.02223	Scholastic Aptitude Test
Eigen- values	7.48	1.63	1.22	.90	

* Communalities were used on the diagonal.

Table 11

Principal Components Analysis Oblimin Rotated
 Factor Pattern Matrix: Four Factor Rotation
 ($\delta = -1$)*

	Factors				
	1	2	3	4	
D1	.04550	-.63117	.15877	.16982	Energy
D2	.27940	-.17254	.12489	.19114	Self-Objectivity
D3	.04541	-.81240	.04170	-.00102	Impact
D4	.38642	-.51750	.00372	-.01139	Leadership
D5	.73133	-.07792	.03352	.01086	Awareness of Social Environment
D6	.59901	-.31004	.01977	-.04695	Behavior Flexibility
D7	.02666	-.72241	-.00609	.10789	Autonomy
D8	.30462	-.06092	.22005	.15780	Fact Finding--Oral
D9	.15917	-.03140	.23406	.47060	Fact Finding--Written
D10	.06912	-.09128	.15909	.56075	Organizing
D11	.31576	.01761	.25289	.45484	Interpreting Information
D12	.10272	-.02604	.10788	.66321	Planning
D13	-.07162	-.09939	-.05934	.44770	Decisiveness
D14	.19735	.01612	.04908	.67156	Decision Making
D15	.20586	-.32693	.21721	.10094	Oral Presentation
D16	.24406	-.30977	.09762	.22347	Oral Defense
D17	-.03217	.02203	.70964	-.03787	Written Communications
D18	.00335	-.07591	.72841	.02253	Scholastic Aptitude Test
Eigen- values	7.48	1.63	1.22	.90	

* Communalities were used on the diagonal.

Table 12

Regression of Overall Assessment Rating Onto
Category Subscale Scores, 2-Way Interactions,
Dummy Coded Assessee Sex Variable, and Sex by
Subscale Score Interactions

Categories	Regression Coefficients*		
	A	B	C
C1: Personal Qualities	.071	.071	.099 NS
C2: Interpersonal Skills	.441	.519	.864
C3: Problem-Solving Skills	.636	.722	1.222
C4: Communication Skills	.267	.266	.275
C2 x C3		-.033 NS	-.209**
SEX			.692**
SEX x C1			-.017 NS
SEX x C2			-.234**
SEX x C3			-.330**
SEX x C4			-.006 NS
SEX x C2 x C3			.119**
Intercept	-1.429	-1.621	-2.529
R Square	.786	.786	.737
N	2191	2191	2181

* All coefficients significant at $p \leq .0001$ unless otherwise noted.

** Significant at $p \leq .01$.

NS Non-significant.

Table 13

Regression of Overall Assessment Rating Onto
Factor Subscale Scores, 2-Way Interactions,
Dummy Coded Assessee Sex Variable, and Sex by
Subscale Score Interactions

Factors	Regression Coefficients*		
	A	B	C
F1: Interpersonal Skills	.724	.832	1.318
F2: Problem-Solving Skills	.551	.680	1.322
F3: Paper & Pencil Instruments	.124	.122	.134
F1 x F2		-.046**	.895
SEX			-.267**
SEX x F1			-.323**
SEX x F2			-.420
SEX x F3			-.007 NS
SEX x F1 x F2			.146
Intercept	-1.447	-1.734	-3.103
R Square	.779	.779	.781
N	2191	2191	2181

* All coefficients significant at $p \leq .0001$ unless otherwise noted.

** Significant at $p \leq .02$.

NS Non-significant ($p \leq .742$)

Table 14
Frequencies: Sex x OAR

	Overall Assessment Rating				
Sex	1	2	3	4	Total
Male	235	308	375	58	976
Female	<u>430</u>	<u>370</u>	<u>369</u>	<u>36</u>	<u>1205</u>
Total	665	678	744	94	2181

Chi-Square = 44.493, $p < .0001$

Table 15

Predicted OAR Means for Males and Females by
 Variables Included in Regression Equation
 and Sex of Subsample the Regression Equation
 Was Derived From

Variables	Equations Sex of Subjects Calculated On	Predicted Means (Standard Deviations)	
		Males	Females
Main Effects	Males	2.26 (.73)	2.03 (.77)
	Females	2.25 (.79)	2.01 (.78)
Main Effects & Interaction	Males	2.26 (.78)	2.03 (.80)
	Females	2.25 (.80)	2.01 (.78)

Table 16

Regression of Overall Assessment Rating Onto
Category Subscale Scores, 2-Way Interactions,
Assessor Group Sex Composition (GSEX), and GSEX by
Subscale Score Interactions

Categories	Regression Coefficients
C1: Personal Qualities	.155**
C2: Interpersonal Skills	.260
C3: Problem-Solving Skills	.367*
C4: Communication Skills	.243***
C2 x C3	.090
Assessor Group Sex (GSEX)	-.243
C1 x GSEX	-.036
C2 x GSEX	.103
C3 x GSEX	.141*
C4 x GSEX	.008
C2 x C3 x GSEX	-.048*
Intercept	-.998**
R Square	.786***
N	2176

* Significant at $p \leq .05$.

** Significant at $p \leq .01$.

*** Significant at $p \leq .0001$.

Table 17

Regression of Overall Assessment Rating Onto
 Factor Subscale Scores, 2-way Interactions,
 Assessor Group Sex Composition (GSEX), and GSEX
 by Subscale Score Interactions

Factors	Regression Coefficients
F1: Interpersonal Skills	.623***
F2: Problem-Solving Skills	.304
F3: Paper & Pencil Instruments	.110**
F1 x F2	.086
Assessor Group Sex (GSEX)	-.268
F1 x GSEX	.076
F2 x GSEX	.143*
F3 x GSEX	.005*
F1 x F2 x GSEX	-.050*
Intercept	-1.020**
R Square	.779***
N	2176

* Significant at $p < .05$.

** Significant at $p < .01$.

*** Significant at $p \leq .0001$.

Table 18

Predicted Means and Standard Deviations of
OAR For All Levels of GSEX When OAR is
Regressed Onto Factor Subscale Scores For
Each GSEX Subgroup

Equations: Main Effects	PREDICTED MEANS (STANDARD DEVIATIONS)				
	GSEX				
	0	1	2	3	4
GSEX GROUP CALCULATED ON					
GSEX = 0	1.96 (.80)	2.25 (.69)	2.22 (.70)	2.32 (.74)	2.34 (.73)
GSEX = 1	1.82 (.85)	2.19 (.77)	2.15 (.78)	2.25 (.83)	2.27 (.82)
GSEX = 2	1.75 (.88)	2.12 (.79)	2.09 (.80)	2.19 (.85)	2.21 (.84)
GSEX = 3	1.74 (.84)	2.09 (.74)	2.06 (.76)	2.16 (.80)	2.17 (.79)
GSEX = 4	1.76 (.77)	2.10 (.70)	2.06 (.71)	2.15 (.75)	2.17 (.75)
GSEX GROUP CALCULATED ON					
Equations: Main Effect & Interaction	0	1	GSEX 2	3	4
GSEX = 0	1.96 (.80)	2.24 (.74)	2.21 (.75)	2.31 (.82)	2.34 (.82)
GSEX = 1	1.81 (.87)	2.19 (.77)	2.15 (.78)	2.25 (.82)	2.27 (.81)
GSEX = 2	1.75 (.88)	2.12 (.79)	2.09 (.80)	2.19 (.85)	2.21 (.84)
GSEX = 3	1.73 (.86)	2.09 (.75)	2.06 (.76)	2.16 (.80)	2.17 (.79)
GSEX = 4	1.73 (.83)	2.10 (.72)	2.06 (.73)	2.15 (.76)	2.17 (.75)
N =	26	290	880	817	163

REFERENCES

- Albemarle Paper Company v. Moody, 10 FEP 1181, (1975).
- Anderson, N.H. Information integration theory: A brief survey. In D.H. Krautz, R.C. Atkinson, R.D. Luce, and P. Suppes (Eds.), Contemporary developments in mathematical psychology. (Vol. 2) San Fransisco: Freeman, 1974.
- Anderson, N.H. & Shanteau, J. Weak inferences with linear models. Psychological Bulletin, 1977, 84, 1155-1179.
- American Psychological Association, Division of Industrial-Organizational Psychology. Principles for the validation and use of personnel selection procedures. (2nd ed.). Berkeley, CA: Author, 1980.
- American Psychological Association. Standards for educational and psychological tests. Washington, D. C.: American Psychological Association, 1974.
- Birnbaum, M.H. The devil rides again: Correlation as an index of fit. Psychological Bulletin, 1973, 79, 239-242.
- Birnbaum, M.H. Reply to the devil's advocates: Don't confound model testing and measurement. Psychological Bulletin, 1974, 81, 854-859.
- Borman, W.C. Validity of behavioral assessment for predicting military recruiter performance. Journal of Applied Psychology, 1982, 67, 3-9.
- Bray, D.W. The assessment center and the study of lives. American Psychologist, 1982, 37, 180-189.
- Bray, D. W. & Campbell, R. J. Selection of salesmen by means of an assessment center. Journal of Applied Psychology, 1968, 52, 36-41.
- Bray, D.W. & Grant, D.I. The assessment center in the management of potential for business management. Psychological Monographs, 1966, 80, (17, Whole No. 625).

- Byham, W.C. Assessor selection and training. In J.L. Moses and W.C. Byham (eds.), Applying the Assessment Center Method. New York: Pergamon Press, 1977.
- Byham, W.C. Review of legal cases and opinions dealing with assessment centers and content validity: Monograph IV. Pittsburgh, PA: Development Dimensions Press, 1979.
- Byham, W.C. & Thornton, G.C. III Assessment centers: A new aid in management selection. Studies in Personnel Psychology, 1970, 2, 21-35.
- Byham, W.C. & Wettengel, C. Assessment center for supervisors and managers: An introduction and overview. Public Personnel Management, 1974, 3, 352-364.
- Cohen, B.M., Moses, J.L., & Byham, W.C. The validity of assessment centers: A literature review. Monograph II. Pittsburgh, PA: Development Dimensions, 1974.
- Cohen, S.L. & Sands, I. The effects of order of exercise presentation on assessment center performance: One standardization concern. Personnel Psychology, 1978, 31, 35-46.
- Dawes, R.M. A case study of graduate admissions: Application of three principles of human decision making. American Psychologist, 1971, 26, 180-187.
- Dawes, R.M. & Corrigan, B. Linear models in decision making. Psychological Bulletin, 1974, 81, 95-106.
- Dunnette, M.D. Criteria for evaluating executive effectiveness: A critique. In F. R. Wickert and D. E. MFarland (eds.) Measuring Executive Effectiveness. New York: Appleton-Century-Crofts, 1967.
- Dunnette, M.D. Multiple assessment procedures in identifying and developing managerial talent. In F. McReynolds (ed.) Advances in Psychological Assessment, vol. 2, Palo Alto, CA: Science and Behavior Books, 1971.
- Finkle, R.B. Managerial assessment centers. In M.D. Dunnette (ed.) Handbook of Industrial and Organizational Psychology. Chicago: Rand-McNally, 1976.
- Firefighters Institute v. City of St. Louis, 14 FEP 1486 (1977).

- Goldberg, L.R. Simple models or simple processes? Some research on clinical judgments. American Psychologist, 1968, 23, 483-496.
- Goldberg, L.R. Man versus model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. Psychological Bulletin, 1970, 73, 422-432.
- Griggs v. Duke Power, 3 FEP 175 (1971).
- Guion, R.M. Open a new window: Validation and values in psychological measurement. American Psychologist, 1974, 29, 287-296.
- Guion, R.M. Content validity-the source of my discontent. Applied Psychological Measurement, 1977, 1, 1-10. (a)
- Guion, R.M. Content validity: Three years of talk - what's the action? Public Personnel Management, 1977, us:6, 407-414. (b)
- Guion, R.M. Scoring of content domain samples: The problem of fairness. Journal of Applied Psychology, 1978, 63, 499-506. (a)
- Guion, R.M. "Content validity" in moderation. Personnel Psychology, 1978, 31, 205-213. (b)
- Guion, R.M. On trinitarian doctrines of validity. Professional Psychology, 1980, 11, 385-398.
- Hammond, K.R. Introduction to Brunswikian theory and methods. New Directions for Methodology of Social and Behavioral Science, 1980, 3, 1-13.
- Hammond, K.R., Rohrbaugh, J., Munpower, J., & Adelman, L. Social judgment theory: Application in policy formation. In M.F. Kaplan & S. Schwartz (Eds.), Human Judgment and Decision Processes in Applied Settings, New York: American Press, 1977.
- Hamner, A., Kim, J.S., Baird, L., & Bigness, W.J. Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. Journal of Applied Psychology, 1974, 59, 705-711.
- Howard, A. An assessment of assessment centers. Academy of Management Journal, 1974, 17, 115-134.

- Huck, J.R. Assessment centers: A review of the external and internal validities. Personnel Psychology, 1973, 26, 191-212.
- Huck, J.R. Determinants of assessment center ratings for white and black females and the relationship of these dimensions to subsequent performance effectiveness. (Doctoral Dissertation, Wayne State University, 1974). University Microfilms No. 74-29,814.
- Huck, J.R. & Bray, D.W. Management assessment center evaluation and subsequent performance of white and black females. Personnel Psychology, 1976, 29, 13-30.
- Kim, Jae-On Factor analysis. In N.H. Nie, C.H. Hull, J.G. Jenkins, K. Steinbrenner, and D.H. Bent (Eds.) SPSS: Statistical Package for the Social Sciences. New York: McGraw-Hill, 1975.
- Kleinmuntz, B. The processing of clinical information by man and machine. In B. Kleinmuntz (Ed.) Formal Representation of Human Judgment. New York: Wiley, 1968.
- Klimoski, R.J. & Strictland, W.J. Assessment centers - valid or merely precient? Personnel Psychology, 1977, 30, 353-361.
- MacKinnon, D.W. An overview of assessment centers. Center for Creative Leadership Technical Report No. 1, 1975.
- Moses, J.L. Assessment center performance and management progress. Studies in Personnel Psychology, 1972, 4, 7-12.
- 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. & Boehm, V.R. Relationship of assessment-center performance to management progress of women. Journal of Applied Psychology, 1975, 60, 527-529.
- Moses, J.L. & Byham W.C. (eds.) Applying the Assessment Center Method, New York: Pergamon Press, 1977.
- Mitchel, J.O. Assessment center validity: A longitudinal study. Journal of Applied Psychology, 1975, 60, 573-579.
- Murray, H.A. Explorations in personality. New York: Oxford Press, 1938.

- Norton, S.D. The empirical and content validity of assessment center vs. traditional methods for predicting managerial success. Academy of Management Review, 1977, 2, 442-453.
- Norton, S.D. & Edinger, J.A. Assessment centers versus traditional methods for predicting managerial success: Empirical and content validity. Catalog of Selected Documents in Psychology, 1978, 8, M.S. 1786.
- Nunnally, J. Psychometric Theory. New York: McGraw-Hill, 1978.
- O.S.S. Assessment Staff. The assessment of men. New York: Rinehart, 1948.
- Prien, E.P. The function of job analysis in content validity. Personnel Psychology, 1977, 30, 167-174.
- Sackett, P.R. and Hakel, M.D. Temporal stability and individual difference in using assessment information to form overall ratings. us;Organization Behavior and Human Performance, 1979, 23, 120-137.
- Sackett, P.R. and Wilson, M.A. Factors affecting the consensus judgment process in managerial assessment centers. Journal of Applied Psychology, 1982, 67, 10-17.
- SAS Institute. SAS User's Guide 1979 Edition. Raleigh, N.C.: Author, 1979.
- Schmitt, N. Interrater agreement in dimensionality and combination of assessment center judgments. Journal of Applied Psychology, 1977, 62, 171-176.
- Schmitt, N. & Hill, T.E. Sex and race composition of assessment center groups as a determinant of peer and assessor ratings. Jcurnal of Applied Psychology, 1977, 62, 261-264.
- Shanteau, J. Correlation as a deceiving measure of fit. Bulletin of the Psychometric Society, 1977, 10 (2), 134-136.
- Shanteau, J. Information integration theory applied to the training of experts. Paper presented at the National Meeting of the American Institute of Decision Sciences, New Orleans, 1979.

- Smith, P. C. Behaviors, results, and organizational effectiveness: The problem of criteria. In M.D. Dunnette (ed.) Handbook of Industrial and Organizational Psychology. Chicago: Rand-McNally, 1976.
- Slovic, P. &, Lichtenstein, S. Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 1971, 7, 649-744.
- Thompson, H.A. Comparison of predictor and criterion judgments of managerial performance using multitrait-multimethod approach. Journal of Applied Psychology, 1970, 54, 496-502.
- Thoresen, J.D. &, Jaffee, C.L. A unique assessment center application with some unexpected by-products. Human Resource Management, 1973, 12, 3-7.
- Uniform Guidelines on Employee Selection Procedures, In BNA's FEP Manual, 1978, 401, 2231-2238.
- U.S. v. AT&T, 8 FEP, 431, 124 (1), 1978.
- Wallace, S.R. How high the validity? Personnel Psychology, 1974, 27, 397-407.
- Wollowick, H.B. &, McNamara, W.J. Relationship of the components of an assessment center to management success. Journal of Applied Psychology, 1969, 52, 348-352.
- Worbois, G.M. Validation of externally developed assessment procedures for identification of supervisory potential. Personnel Psychology, 1975, 28, 77-91.