



University of Tennessee, Knoxville

TRACE: Tennessee Research and Creative Exchange

Doctoral Dissertations

Graduate School

6-1986

The Effect of Job Involvement and Organizational Commitment on the Psychometric Characteristics of Job Performance Ratings

Steven Ronald Gordon
University of Tennessee - Knoxville

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



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

Recommended Citation

Gordon, Steven Ronald, "The Effect of Job Involvement and Organizational Commitment on the Psychometric Characteristics of Job Performance Ratings. " PhD diss., University of Tennessee, 1986. https://trace.tennessee.edu/utk_graddiss/3038

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

To the Graduate Council:

I am submitting herewith a dissertation written by Steven Ronald Gordon entitled "The Effect of Job Involvement and Organizational Commitment on the Psychometric Characteristics of Job Performance Ratings." I have examined the final electronic copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.

John W. Lounsbury, Major Professor

We have read this dissertation and recommend its acceptance:

John M. Larsen, Gerald H. Whitlock, & H. Dudley Dewhirst

Accepted for the Council:

Carolyn R. Hodges

Vice Provost and Dean of the Graduate School

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

THE EFFECT OF JOB INVOLVEMENT AND ORGANIZATIONAL COMMITMENT
ON THE PSYCHOMETRIC CHARACTERISTICS OF JOB PERFORMANCE RATINGS

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

Steven Ronald Gordon

June 1986

DEDICATION

This dissertation is dedicated to my wife

Leslie Acomb Gordon.

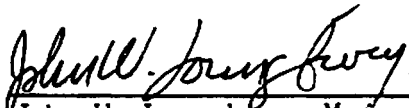
Her love, patience, understanding,

and support gave me the

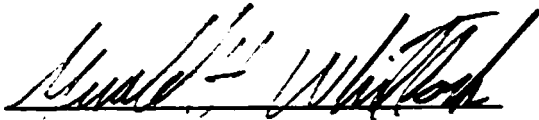
strength to complete this task.

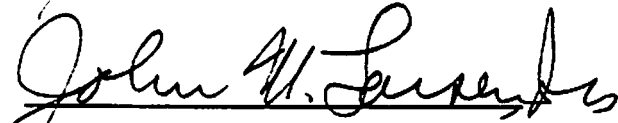
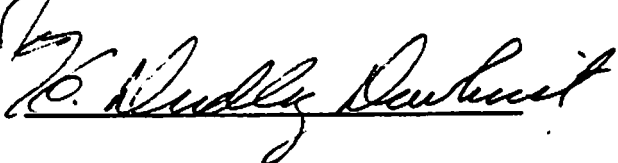
To the Graduate Council:

I am submitting herewith a dissertation written by Steven Ronald Gordon entitled "The Effect of Job Involvement and Organizational Commitment on the Psychometric Characteristics of Job Performance Ratings." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.



John W. Lounsbury, Major Professor

We have read this dissertation
and recommend its acceptance:



Accepted for the Council:


Vice Provost
and Dean of the Graduate School

ACKNOWLEDGMENTS

I wish to thank my committee chairman Dr. John W. Lounsbury for his guidance, patience, support, and friendship throughout the course of my graduate studies. I am also grateful for the assistance and valuable comments of Dr. John M. Larsen, Dr. Gerald H. Whitlock, and Dr. H. Dudley Dewhirst who served as members of my dissertation committee.

Special thanks are extended to Mr. Cole Piper for permission to obtain the data which served as the basis of this study. I also wish to thank Dr. E. Craig McGee and Dr. Terry Guth for their assistance during the initial stages of this project.

I hereby acknowledge with thanks the permission granted by the editor of the Addison-Wesley Publishing Company to reprint Figure 1.2 on page 16 of Belief, Attitude, Intention and Behavior: An Introduction to Theory and Research by M. Fishbein and I. Ajzen (1975). I similarly acknowledge the permission granted by the editor of the Academy of Management Review for permission to reprint Figure 1 from Volume 7, Number 3, page 420 (1982).

Finally, I wish to extend my sincere appreciation to my parents, Ted and Muriel Gordon for their continual love and support. The values they taught me are undoubtedly the most important lessons I have ever learned.

ABSTRACT

The purpose of this study was to examine a theoretical model developed to predict the influence of rater job involvement and organizational commitment on the process of appraising employee job performance. The model is based upon Fishbein's (1967) theory relating attitudes to behavioral intentions and specific behaviors. Organizational commitment is incorporated as an indication of the subjective norm and the performance -- self-esteem definition of job involvement represents the attitudinal component. Hypotheses proposed to test the model addressed the psychometric characteristics of ratings assigned by individuals with varying levels of a composite measure of job involvement and organizational commitment. A field study was conducted with subjects employed at three locations of a department store. Ratings of 199 sales clerks were made by 38 section leaders on nine dimensions of a graphic rating scale and a behavioral checklist. Measures of rater job involvement and organizational commitment were obtained. Multiple operational procedures were used to assess the extent of leniency, central tendency, restriction of range, and halo in the ratings. Contrary to the predictions of the model, scores on the composite attitude variable were directly related to the prevalence of the four rating errors. The correspondence between rater attitude scores and the amount of each error was not significantly different for evaluations made on either rating format. Rater ability levels moderated the relationship between composite scores and the amount of leniency and central tendency in ratings on

the graphic rating scale. Scores on the composite variable were not predictive of measures of rating reliability or convergent and discriminant validity. Suggestions were made for revisions of the model. Alternative explanations proposed for the contrary experimental results included inconsistencies between rater attitudes and the behavioral criteria considered, the influence of alternative performance -- self-esteem contingencies, and failure to incorporate assignment of error-free ratings into the subjective norm. The results of the study imply that the reward structures and values of organizations must be altered if performance appraisal programs are to succeed.

TABLE OF CONTENTS

CHAPTER	PAGE
I INTRODUCTION	1
II REVIEW OF THE LITERATURE	8
Introduction	8
Rater Personal Characteristics	11
Cognitive and Psychological Attributes	11
Job Performance and Attitudes	14
Rater Demographic Characteristics	17
Race and Sex	17
Age, Education and Experience	20
Attitudes and Behavior	23
Fishbein's Behavioral Intentions Model	23
Wiener's Model of Organizational Commitment	29
Job Involvement and Instrumental Motivation	34
III THEORETICAL MODEL AND HYPOTHESES	39
The Model	39
The Hypotheses	46
IV METHOD	61
Subjects	61
Job Performance Measures	61
Job Involvement and Organizational Commitment	67
Rating Errors	70
Procedure	72
V RESULTS	76
Introduction	76
Reliability of Performance Measures	79

CHAPTER	PAGE
Research Design, Test Procedures, and Power	83
Leniency	86
Central Tendency	115
Restriction of Range	135
Halo	150
Reliability	168
Convergent and Discriminant Validity	175
VI DISCUSSION	192
Discussion of Results	192
Limitations of the Study	210
Directions for Future Research	213
Conclusions and Implications of the Study	214
LIST OF REFERENCES	218
APPENDICES	231
APPENDIX A. Sales Clerk Performance Appraisal Form . .	232
APPENDIX B. Mean and Standard Deviation of the Importance Ratings for Items Included on the Behavioral Checklist	246
APPENDIX C. Section Leader Performance Appraisal Form .	256
APPENDIX D. Job Involvement Scale.	259
APPENDIX E. Organizational Commitment Scale.	262
APPENDIX F. Procedures for Computation of Variance Components and Indexes from Multitrait- Multimethod Matrix Data.	265
VITA	267

LIST OF TABLES

TABLE		PAGE
1	Demographic Characteristics of the Sales Clerk Sample .	62
2	Demographic Characteristics of the Section Leader Sample	63
3	Summary Statistics for Section Leader Attitude Scale Scores	77
4	Summary Statistics for Rating Dimension Scores on the Graphic Rating Scale and the Behavioral Checklist for Sales Clerks	78
5	Summary Statistics for Rating Dimension Scores on the Graphic Rating Scale for Section Leaders	80
6	Reliability Coefficients for the Sales Clerk Perfor- mance Ratings	81
7	Dimension and Overall Means of the Leniency Measure for the Sales Clerk Sample Dichotomized on the Basis of the Median Values of Section Leader Attitude Scale Scores	88
8	Analysis of Variance with Section Leader Composite Score Group and Rating Scale Dimensions as the Independent Variables and the Leniency Measure as the Dependent Variable	89
9	Analysis of Variance with Section Leader Job Involve- ment Score Group and Rating Scale Dimensions as the Independent Variables and the Leniency Measure as the Dependent Variable	90
10	Analysis of Variance with Section Leader Organiza- tional Commitment Score Group and Rating Scale Dimensions as the Independent Variables and the Leniency Measure as the Dependent Variable	91
11	Comparison of Mean Dimension Skewness (Leniency) in Ratings Assigned by Section Leaders Scoring Above and Below the Median on the Three Attitude Measures	95
12	Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Leniency Measures Across Performance Dimensions on the Graphic Rating Scale and the Behavioral Checklist With Sales Clerks as the Unit of Analysis	98

TABLE	PAGE
13 Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of the Leniency Measures Across Performance Dimensions on the Graphic Rating Scale and the Behavioral Checklist with Section Leaders as the Unit of Analysis	100
14 Analysis of Variance with Section Leader Attitude Score Groups and Rating Format as the Independent Variables and Dimension Skewness as the Dependent Variable	102
15 Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Leniency from Section Leader Composite Attitude Scores With Sales Clerks as the Unit of Analysis	106
16 Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Leniency from Section Leader Composite Attitude Scores With Section Leaders as the Unit of Analysis . . .	108
17 Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Leniency Measures Across Performance Dimensions with Sales Clerks as the Unit of Analysis	111
18 Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Leniency Measures Across Performance Dimensions and Across Sales Clerks with Section Leaders as the Unit of Analysis	112
19 Summary of Results for Tests of Leniency Hypotheses 1A - 1D	114
20 Dimension and Overall Means of the Central Tendency Measure for the Sales Clerk Sample Dichotomized on the Basis of the Median Values of Section Leader Attitude Scale Scores	116
21 Analysis of Variance with Section Leader Composite Score Group and Rating Scale Dimensions as the Independent Variables and Central Tendency Measure as the Dependent Variable	118

TABLE

PAGE

22	Analysis of Variance with Section Leader Job Involvement Score Group and Rating Scale Dimensions as the Independent Variables and Central Tendency Measure as the Dependent Variable	119
23	Analysis of Variance with Section Leader Organizational Commitment Score Group and Rating Scale Dimensions as the Independent Variables and the Central Tendency Measure as the Dependent Variable	120
24	Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Central Tendency Measures Across Performance Dimensions on the Graphic Rating Scale and the Behavioral Checklist with Sales Clerks as the Unit of Analysis	123
25	Comparison of Correlations Between Section Leader Attitude Scale Scores and Mean of the Central Tendency Measures Across Performance Dimensions on the Graphic Rating Scale and the Behavioral Checklist with Section Leaders as the Unit of Analysis	125
26	Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Central Tendency from Section Leader Composite Attitude Scores with Sales Clerks as the Unit of Analysis	127
27	Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Central Tendency from Section Leader Composite Attitude Scores with Section Leaders as the Unit of Analysis . . .	129
28	Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Central Tendency Measures Across Performance Dimensions with Sales Clerks as the Unit of Analysis	131
29	Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Central Tendency Measures Across Performance Dimensions and Across Sales Clerks with Section Leaders as the Unit of Analysis	132
30	Summary of Results for Tests of Central Tendency Hypotheses 2A - 2D	134

TABLE	PAGE
31 Comparison of Restriction of Range in Ratings Assigned by Section Leaders Scoring Above and Below the Median on the Three Attitude Measures	136
32 Comparison of Mean Dimension Kurtosis (Restriction of Range) in Ratings Assigned by Section Leaders Scoring Above and Below the Median on the Three Attitude Measures	139
33 Comparison of Correlations Between Section Leader Attitude Scale Scores and a Measure of Restriction of Range of the Ratings Assigned by Each Section Leader on the Graphic Rating Scale and Behavioral Checklist . .	142
34 Analysis of Variance with Section Leader Composite Attitude Score Group and Rating Scale as the Independent Variables and Dimension Kurtosis as the Dependent Variable	143
35 Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Restriction of Range from Section Leader Composite Attitude Scores with Section Leaders as the Unit of Analysis	146
36 Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Restriction of Range in Ratings Assigned on the Graphic Rating Scale and the Behavioral Checklist	148
37 Summary of Results for Tests of Restriction of Range Hypotheses 3A - 3D	149
38 Comparison of the Measurement of Halo in Ratings Assigned by Section Leaders Scoring Above and Below the Median on the Three Attitude Measures	152
39 Multitrait - Multimethod Correlation Matrix for Sales Clerks Rated by Section Leaders Scoring Above the Median on the Job Involvement - Organizational Commitment Composite Variable	154
40 Multitrait - Multimethod Correlation Matrix for Sales Clerks Rated by Section Leaders Scoring Below the Median on the Job Involvement - Organizational Commitment Composite Variable	155

TABLE	PAGE
41 Comparison of the Measurement of Halo in Ratings Assigned on the Graphic Rating Scale and Behavioral Checklist by Section Leaders Scoring Above and Below the Median on the Composite Attitude Measure	157
42 Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Halo on the Graphic Rating Scale and the Behavioral Checklist with Sales Clerks as the Unit of Analysis	160
43 Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Halo on the Graphic Rating Scale and the Behavioral Checklist with Section Leaders as the Unit of Analysis	161
44 Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Halo from Section Leader Composite Attitude Scores with Sales Clerks as the Unit of Analysis	163
45 Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Halo from Section Leader Composite Attitude Scores with Section Leaders as the Unit of Analysis	164
46 Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Halo on the Graphic Rating Scale and the Behavioral Checklist with Sales Clerks as the Unit of Analysis	166
47 Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Halo on the Graphic Rating Scale and the Behavioral Checklist with Section Leaders as the Unit of Analysis	167
48 Summary of Results of Results for Test of Halo Hypotheses 4A - 4D	169
49 Comparison of Reliability Coefficients for Sales Clerk Performance Ratings Assigned by Section Leaders Scoring Above and Below the Median on the Job Involvement - Organizational Commitment Composite Measure	172
50 Generalizability Study for Sales Clerks Rated by Section Leaders Scoring Above the Median on the Composite Attitude Scale	176

TABLE

PAGE

51	Generalizability Study for Sales Clerks Rated by Section Leaders Scoring Below the Median on the Composite Attitude Scale	177
52	Analysis of Variance of Correlations from the Multitrait - Multimethod Matrices for Sales Clerks Rated by Section Leaders Scoring Above or Below the Median on the Composite Attitude Scale	182
53	Comparison of Variance Components and Indexes from the Multitrait-Multimethod Matrices for Section Leaders Scoring Above and Below the Median on the Composite Attitude Scale	184
54	Correlation of Section Leader Demographic Variables with Work-Related Attitudes	187
55	Relative Importance of Graphic Rating Scale Dimension Scores for Predicting Overall Performance Assessments Assigned by Section Leaders Scoring Above and Below the Median on the Composite Attitude Scale	189
56	Relative Importance of Behavioral Checklist Dimension Scores for Predicting Overall Performance Assessments Assigned by Section Leaders Scoring Above and Below the Median on the Composite Attitude Scale	190

LIST OF FIGURES

FIGURE	PAGE
1	Schematic presentation of conceptual framework for prediction of specific intentions and behaviors. 26
2	Wiener's model representing relationships between organizationally related behaviors, beliefs concerning those behaviors, and commitment and instrumental motivation 32
3	Schematic diagram of the theoretical model for prediction of rating behavior on the basis of rater job involvement and organizational commitment levels. 40

THE EFFECT OF JOB INVOLVEMENT AND ORGANIZATIONAL COMMITMENT ON THE PSYCHOMETRIC CHARACTERISTICS OF JOB PERFORMANCE RATINGS

CHAPTER I

INTRODUCTION

Few sources of information available for management decision making have had potential to serve as wide a spectrum of purposes as performance appraisal. Performance measurement is a critical element in making organizational decisions regarding promotions, transfers and layoffs, establishing and evaluating training programs, determining wages and salaries, counseling employees, and serving as criteria for validation studies. The widespread prevalence of formal performance appraisal programs in both the private and public sectors (Locher & Teel, 1977; Feild & Holley, 1975) leaves little doubt that the need to evaluate employee performance on the job is an organizational reality.

Ideally, a complete assessment of performance would include both objective indices such as production levels and personnel data (e.g. absences, accidents), as well as judgmental evaluations (Landy & Trumbo, 1981). Although the first two of these components are desirable because of their lack of subjectivity, they are often contaminated by factors beyond the control of the individual and reflect behavioral outcomes rather than employee behavior (Cascio, 1978). Thus, despite the well documented problems inherent in the use of performance ratings, most individuals rely upon a judgmental index for making personnel decisions.

However, the problems with the use of performance ratings are myriad and very resistant to solution. These difficulties have been the source of the prevalent dissatisfaction with these measures felt by both practitioners and researchers in the field. The principal problems with judgmental evaluations stem from the fact that they are extremely vulnerable to both intentional and inadvertent bias as manifested in the form of rating errors. Consequently, performance ratings have been the focus of extensive research efforts to study and improve their psychometric characteristics.

The task of appraising an individual's performance is essentially a two-part process consisting of first, the observation of behavior and second, the rating of that behavior against a standard. As such, both phases of the process are susceptible to certain characteristic problems. Borman (1978) has classified the causes of problems with performance ratings into four categories: 1) limitations of the rater's opportunity to observe job-related ratee behavior; 2) insufficient knowledge of the nature of common rating errors on the part of the rater and a lack of experience with the performance appraisal process; 3) problems inherent in the rating instrument; and 4) organizational constraints put upon the rater which serve to systemically affect judgments. An additional source of potential bias which has received considerable attention concerns the personal characteristics of the rater and the possible interaction of these factors with characteristics of the ratee.

In recent years, research in the field has been dominated by studies attempting to reduce the prevalence of rating errors either

through the implementation of rater training programs or the development of bias resistant rating scale formats. Rater training has generally proven to be effective in reducing common rating errors (Bernardin, 1978; Bernardin & Walter, 1977; Brown, 1968; Latham, Wexley, & Purcell, 1975). However, many questions remain unanswered in this domain regarding the content of rater training programs (Ivancevich, 1979; Warmke & Billings, 1979); the required length of the programs (Bernardin, 1978; Borman, 1975); the effects of error reduction training on the accuracy and validity of ratings (Bernardin & Pence, 1980; Dunnette & Borman, 1979); and the longitudinal effects of rater training on error reduction (Bernardin, 1978a; Latham, Wexley & Purcell, 1975; Warmke & Billings, 1979).

Research regarding the rating instrument has proceeded on the premise that if the format used to structure observation and record the evaluations could be designed to reduce the subjectivity of the rating process, then the psychometric quality of the ratings would improve. The primary thrust of this body of research has been the development and refinement of behaviorally-based rating scales (Latham & Wexley, 1977; Smith & Kendall, 1963). A voluminous amount of research has been conducted to contrast various behavioral rating scales with traditional (e.g. trait-based) formats. The outcome leaves the question of scale format superiority largely unresolved (Bayroff, Haggerty & Runquist, 1954; & DeCotiis, 1977; Landy & Farr, 1980; Schwab, Heneman & DeCotiis, 1975). This work has shown that regardless of the particular format, scales superior in terms of reducing rating errors will result from psychometric rigor in

development (Bernardin, 1977; Bernardin, Alvares & Cranny, 1976) and by having individuals who will be using the scales participate in the development process (Friedman & Cornelius, 1976). However, the available research evidence suggests that even when these suggestions are followed, only about 4% to 8% of the variance on ratings can be accounted for by the rating format employed (Borman & Dunnette, 1975; Landy & Farr, 1980).

In light of these conclusions, there is a growing consensus in the field that further efforts to identify the elusive ultimate rating format should be halted until greater insight is gained into the rating process itself and the organizational and individual variables that affect this process (Bernardin, 1977; Cooper, 1981; DeCotiis, 1977); Feldman, 1981; Landy & Farr, 1980). Among the process-related factors which have been suggested for study are the purpose of the ratings (Borresen, 1967; Sharon & Bartlett, 1969), structural and organizational climate variables (Kavanaugh, 1973), the means by which raters make their judgments (Christal, 1968; Zedeck & Kafry, 1977), as well as individual differences in raters that systematically affect the rating process (e.g., Kirchner & Reisberg, 1962; Mandell, 1956; Taft, 1955).

The last of these areas has been the subject of a considerable amount of research primarily directed toward identifying the characteristics of superior raters in terms of their ability to accurately observe and evaluate the behavior of others and minimize constant rating errors. The individual difference variables most frequently studied have tended to be secondary characteristics such as race, sex,

tenure, and job performance rather than potential direct influences on rating behavior such as cognitive processes and attitudes. Borman (1978), Landy & Farr (1980), and Feldman (1981), among others, have proposed the application of theoretical frameworks from the study of personality and social psychology for conceptualizing the mental processes involved in evaluating the behavior of others. Two of the models proposed for this purpose include Kelly's (1955) theory of personal constructs and implicit personality theory (Bruner & Tagiuri, 1954). However, no theoretical models have been proposed to explain the impact of relevant rater attitudes on the performance appraisal process. A few studies that have examined the effect of raters' attitudes toward the use of various rating forms and the rating process (Rothe, 1949; Schneier, 1977) and the effects of rater-ratee similarity in background and attitudes. But no published research has investigated the influence of the work-related attitudes of raters on the psychometric characteristics of ratings made to evaluate ratee job performance.

The primary objective of the present study is to investigate the influence of two widely researched variables, job involvement and organizational commitment, on the process of making performance appraisal ratings. This line of research addresses the study of the rating process by examining one means by which organizational characteristics and contextual factors directly impinge upon the process of appraising individual performance - through their impact on rater attitudes which in turn influence behavior. It is suggested that organizational and contextual factors lead to the formation of beliefs

which have a direct effect on the formation of these work-related attitudes. These attitudes then influence rater behavioral intentions and in turn, rater behavior during the observational and rating portions of the appraisal process. In doing so, these attitudes may systemically influence the occurrence of constant errors in rating as well as the reliability and validity of performance appraisal ratings.

A theoretical model is proposed to explicate the means by which job involvement and organizational commitment affect the rating process. This model represents an attempt to integrate existing models of the rating process suggested by Borman (1978), Landy and Farr (1980), and Cooper (1981) with a model of work attitudes as behavioral determinants presented by Wiener (1982) and based upon Fishbein's behavioral intentions model (Fishbein, 1967; Fishbein & Ajzen, 1975). Research hypotheses are proposed to test the predicted relationships between various components of the model. These research questions address the nature of attitudinal influences on the psychometric characteristics of performance appraisal ratings made in a field setting with two different rating scale formats.

This study is intended as a response to the call for greater understanding of the process of rating job performance. The proposed model depicts a theoretical view of the influence of job involvement and organizational commitment on ratings and represents an attempt to provide new insight into the impact of organizational and contextual factors upon the performance appraisal process. The study of the behavioral correlates of these attitudes is an important element in the performance rating puzzle which has previously been overlooked by

researchers working in this area. This study was conducted as a first step in rectifying this oversight.

CHAPTER II

REVIEW OF THE LITERATURE

Introduction

Rating errors in performance appraisal persist regardless of the type of rating format or training program used or the organizational context in which the evaluations are made. This suggests that there are systematic differences in the way people rate the behavior of others which are a function of individual differences in raters. Research in this area has primarily focused on identifying correlates of accurate performance ratings in order to isolate the characteristics of superior raters (Borman, 1979a; Schneider & Bayroff, 1953, Taft, 1955). Personal and demographic variables which have been demonstrated to affect subjective judgments in other interpersonal observation situations (e.g., cognitive complexity) have been applied to performance appraisal in an attempt to determine extraneous factors which influence ratings and to examine components of the decision-making process (Schneider, 1977; Zedeck & Kafry, 1977). Although such investigations have not resulted in definitive criteria for selecting superior raters as yet, some insights have been gained concerning the nature of the judgment process and the influence of extraneous factors.

The rater characteristics investigated with respect to their effect on performance appraisal ratings may be grouped into two general categories: 1) personal characteristics of the rater; and 2) demographic characteristics. Research conducted on factors falling

within both of these classifications is summarized to integrate findings and serve as a foundation for theoretical development.

Most of the research conducted on rater characteristics has focused on the reliability of ratings and the prevalence of rating errors (Thornton & Zorich, 1980). However, an increasing number of studies in this area have begun to consider the equally important issues of validity and accuracy (e.g., Bernardin & Pence, 1980; Borman, 1979a). The issue of validity concerns whether the systematic source of variance is related to the attribute being measured. In contrast, accuracy has been defined as the degree to which ratings are relevant to or correlated with true criterion scores (Dunnette & Borman, 1979).

The distinction between these concepts is critical in the study of ratings. Most early studies of performance ratings focused exclusively on reliability and constant rating errors. The assumption implicit in these investigations was that the level of rating errors in the judgments is inversely related to the accuracy and validity of the ratings (Bernardin, 1978). This belief has since been seriously questioned on the basis of the results of several laboratory studies which found no significant relationship between rating errors and accuracy (Bernardin & Pence, 1980; Bernardin, Cardy & Carlyle, 1982).

Although reliable ratings are highly desirable, reliability alone is not sufficient to demonstrate that performance appraisals are providing the information that interpretation typically ascribes to them. It is critical to also investigate the accuracy and validity of ratings especially in light of their widespread use as the basis for

important personnel and organizational decisions. However, both of these latter rating characteristics have proven difficult to study. Studies of interpersonal perceptual accuracy are complicated by problems of establishing true measures of ratee traits or dimension scores, methodological difficulties in selecting proper scoring indices to assess rater accuracy, and the instability of accuracy estimates across situations (Cronbach, 1955; Borman, 1979a). On the other hand, the problem of validating ratings, as with any immediate criteria, becomes one of determining the relevance of the measure to the appropriate intermediate or ultimate criterion (Thorndike, 1949). Kavanaugh (1973) has suggested that this process may best be conceptualized from the approach taken in establishing construct validity.

Finally, a comment is in order regarding the nature of rater bias. This phrase is frequently and inconsistently used in the discussion of performance appraisals. There are three general ways in which the phrase has been applied to the rating process: 1) in relation to characteristics of the rater (e.g., Klimoski & London, 1974; Klores, 1966); 2) in relation to characteristics of the ratee (e.g., Cascio & Calenzi, 1978); and 3) in terms of the interaction between rater and ratee characteristics (e.g., DeJung & Kaplan, 1962; Hamner, Kim, Baird & Bigoness, 1974). The first of these sources of bias is examined in the present study in regard to the influence exerted on the rating process by rater work-related attitudes.

Rater Personal Characteristics

Cognitive and Psychological Attributes

Campbell, Dunnette, Lawler and Weick (1970) noted that "observers differ considerably in their ability to do an effective, reliable, and valid job of observing and recording the job behavior of other persons" (p. 115). Traditionally, studies of the ability to judge have focused on correlates of rating behavior and demonstrated the existence of such an ability by summarizing the personal characteristics of the so-called "accurate" rater. Although the research concerning cognitive characteristics is fragmented, a pattern of results is discernible which suggests that some individual difference variables may substantially influence rater judgments.

In 1955, Taft conducted a literature review of research on the ability of persons to judge the traits, motives, and abilities of others. Intelligence, self-insight, and social skills were concluded to be significantly related to judgmental accuracy. Additional personal characteristics of accurate raters identified by Borman (1979) are as follows: dependability, stability maturity, patience, and verbal fluency as well as intelligence and personal adjustment. These individual differences accounted for approximately 17% of the total variance in performance rating accuracy. Borman concludes that these results "suggest that individual differences probably play a significant role in determining a person's accuracy in evaluating other's performance" (p.113).

A wide variety of psychological factors have been investigated in relation to their impact on performance ratings. However, in most

cases, the consideration of each variable has been limited to one study which makes it difficult to reach any firm conclusions about their effects (Landy & Farr, 1980). For example, the tendency to use extreme response categories in the evaluation of others has been found to be directly related to a person's level of anxiety (Lewis & Taylor, 1955). Mandell (1956) found a significant relationship between raters' self-confidence and the degree of leniency in their ratings of others. In contrast, Rothaus, Morton, and Hanson (1965) found that the psychological distance between the rater and the ratee correlated positively with the severity of ratings given.

Cognitive characteristics of raters have also been studied by examining their relation to consistency of judgments across tasks or settings. Consistency is assessed by determining whether an individual who provides accurate ratings in one situation or on one rating dimension will make similarly accurate assessments in other situations or on other dimensions.

Mullins and Force (1962) conducted an exploratory study to determine whether people who accurately predicted the scores of peers on a vocabulary test would also be accurate in evaluating their scores on a test of carefulness. By considering the correlations between predicted scores and test scores, the authors concluded that there was evidence of a general ability to make accurate ratings. However, the results of this study must be viewed cautiously as the measure of carefulness was a composite of five tests designed to measure this characteristic that were not highly intercorrelated. This makes the

correlation between the ratings and criterion very difficult to interpret.

Borman (1977, 1979b) examined two types of consistency: across task and within task. Across task consistency is demonstrated if approximately the same amount and type of error is present in rating individuals in two or more settings or on different tasks. Within task consistency is an indication of the amount and type of error in ratings in a single setting or on various components of one task. Both of Borman's studies demonstrated only moderate rating consistency across two distinct jobs.

Information processing has also been considered in relation to performance appraisal in order to determine whether the way a rater integrates and interprets performance data contributes to the accuracy of ratings. The ability to discriminate between various dimensions of a stimulus as well as to make distinctions within dimensions has been termed cognitive complexity (see Schneier, 1977). In a widely referenced study, Schneier (1977) found evidence that when the cognitive complexity of raters and the complexity of a particular rating format were compatible, the ratings made exhibited less halo, central tendency, and restriction of range. In addition, cognitively complex raters were found to evidence less halo in their ratings regardless of the scale format used. Unfortunately, these intriguing findings have since been brought into question by the failure of several attempts to replicate Schneier's results (Bernardin & Boetcher, 1978; Bernardin, Cardy & Carlyle, 1982; Lahey & Saal, 1981; Sauser & Pond, 1981).

Zedeck and Kafry (1977) conducted a study to investigate how raters integrate information to form an overall rating of job performance. They first used a regression procedure to determine the relative weights given to various rating dimensions by each rater and next employed a clustering procedure to identify groups of raters possessing similar rating strategies. An analysis of individual differences associated with the various strategies was then conducted. Although the variables considered were some of those which Taft (1955) found to be related to the ability to accurately judge others, no significant differences in these factors were found between groups of raters employing different rating strategies.

Landy and Farr (1980) conclude that "in general, cognitive characteristics of raters seem to hold the most promise for understanding the rating process" (p.72). Much additional research must be conducted before general conclusions concerning the influence of raters' cognitive characteristics on ratings can be made. The hypothesized ability to judge must be operationally defined. Currently, systematic differences in the way people rate others are attributed to a difference in ability. However, these differences may in fact be due to other factors which affect behavior such as the rating situation, the rating strategy used, and rater attitudes towards the rating process or the organizational context in which the ratings are made.

Job Performance and Attitudes

The relationship between the performance level of raters and the way they rate others has been the subject of a small number of

studies. The relation between the validity of ratings and the scores of raters on an aptitude test, their academic achievement levels, and their overall rated value to the army was examined in a study by Schneider and Bayroff (1953). They found that the validity of ratings varied directly with all three performance measures.

Kirchner and Reisberg (1962) divided a sample of raters into three subgroups on the basis of a measure of supervisory effectiveness. They found that the better supervisors showed greater range in their ratings of subordinates. They also placed their greatest emphasis on behaviors requiring independent action on the part of the ratees in arriving at their evaluations. In contrast, Mandell (1956) found that raters' level of performance was not related to the degree of leniency in their ratings. He did conclude though, that supervisors who were rated as poor performers on the job tended to be poor judges of the performance of others. Similarly, Bayroff, Haggerty, and Rundquist (1954) reported that raters who were considered to be better job performers on the basis of written test scores, training class standing, and a criterion ranking gave ratings of peers that were more valid in terms of predicting future job performance.

The relationship between styles of leadership and performance ratings given to subordinates has also been examined directly in two studies. Taylor, Parker, Martens, and Ford (1959) reported that supervisors who were highly production oriented tended to give their subordinates relatively low performance ratings. Klores (1966) found that supervisors who were high on the variable of consideration tended to show greater leniency in their ratings of others. Raters who were

high in initiation of structure evidenced less central tendency in their ratings than those who were low on this factor.

A very limited amount of research has been done regarding the influence of raters' attitudes on the nature of their ratings of others. Rothe (1949) used a quasi-experimental design to find that raters' degree of acceptance and enthusiasm toward the rating process differed systematically with their assessments of co-workers. A number of other studies have measured raters' attitudes in regard to the rating process and the scales used (e.g., Schneier, 1977; Bernardin, Orban, & Carlyle, 1981). However, in none of the studies identified, was the relationship of these attitudes to the psychometric characteristics of the ratings empirically examined.

The attitudes of raters toward ratees may also influence the nature of performance ratings. This tenet has been investigated by examining the extent to which the degree of similarity a rater perceives between the ratee and himself influences the characteristics of his ratings of that person. Vroom (1959) summarized the results of a laboratory study on perceived similarity as follows: "An individual accurately perceives characteristics that are part of his self-concept in a) others towards whom he has a positive attitude and who are similar in these characteristics and b) others towards whom he has a negative attitude and who are dissimilar in these characteristics" (p. 343). Frank and Hackman (1975) considered the influence of rater-ratee similarity in terms of biographical and attitudinal variables on ratings of candidates made by college admissions interviewers. They used a design employing only three raters and found very mixed results

between interviewers indicating considerable individual variation in terms of the influence of this factor. No other studies reported in the literature have examined the relationship of rater work-related attitudes on the psychometric characteristics of their ratings of the job performance of others. Landy and Farr (1980) have suggested that studies of such relationships represents a promising direction for future research into the performance appraisal process.

Rater Demographic Characteristics

Race and Sex

As was the case for most of the personal characteristics thought to influence rater judgments, the hypothesis that a rater's race and sex have an effect on the performance ratings of others has not been unequivocally supported or disconfirmed. In those instances where a significant relationship has been found, the amount of variance in ratings accounted for by rater's race or sex has generally been small.

The existence of a bias in favor of ratees of the same race as the rater has been suggested by several authors. This type of rating bias is said to occur when a rater judges people of his or her own race differently (i.e., more leniently) than individuals of another race. A number of laboratory studies have found qualified support for the belief that raters tend to give somewhat higher ratings to ratees of the same race. DeJung and Kaplan (1962) found evidence of racial bias in performance ratings but only for black raters. Cox and Krumboltz (1958) found similar results but noted a substantial overlap in the distributions of ratings given by minority and non-minority

judges. Also, rating bias was not present for all of the work groups considered and in general, raters agreed on the rank order of ratees. Hamner, Kim, Baird, and Bigoness (1974) employed an objective criterion of performance in a simulated work setting and discovered that although a racial bias was present in the ratings, performance differences were the major determinants of the appraisals given. The race of the rater accounted for only 2% of the variance in ratings.

Two field studies have specifically examined the effect of rater race on the performance ratings of others. Crooks (1972) reported the findings of an investigation conducted over a six-year period. In the majority of instances, race of the rater did affect the ratings given to subordinates as raters were more lenient in evaluating persons of their own race. Wendelken and Inn (1981) conducted a study to examine the effects of rater race and ratee race as well as an index of ratee job performance on ratings of performance on a structured interview for predicting the future performance of police officers in supervisory positions. This study was unique in that it permitted a separate analysis of the variance in ratings accounted for by performance level as well as the nonperformance factors considered. The use of this design enabled the authors to determine whether rating differences were due to real performance variations, rating biases, or both (Dunnette & Borman, 1979). The results of their investigation showed that although the main effects of rater race, ratee race, job performance level, and the rater race by ratee race interaction were all significant, none of these elements accounted for even a moderate portion of the variance in the interview ratings. This study

accentuates the importance of considering indices of the variance accounted for in determining the influence of nonperformance factors on performance ratings (Osborn & Vicars, 1976).

Although evidence has been presented to demonstrate differences in the rating behaviors of males and females, the inconsistency of the findings prohibits making the assumption that the sex of raters systematically influences their evaluations of others. For example, London and Poplawski (1976) found in a study stimulating interview and performance appraisal contexts, that several dimensional ratings made by females were consistently higher than those of the male raters although there were no systematic differences in overall ratings. In contrast, Norton, Gustafson, and Foster (1971) found that in assessing the management potential of subjects, male raters were more lenient than females. Hamner, et al. (1974) reported that females generally evaluated the performance of ratees higher than did male raters and that the strength of this difference covaried with the level of ratee performance. Finally, Goldberg (1968) and Pheterson, Kiesler, and Goldberg (1971) used rating simulations to show that female raters tend to evaluate the performance and accomplishments of women more stringently than do male raters, but only when the rating criteria are highly subjective or are not clearly specified.

The studies reviewed demonstrate that raters' race and sex have only a minimal influence on the performance appraisals of others. It has been suggested that the context in which the rating takes place may offer a more viable explanation for those incidents of rating bias that have been attributed to these nonperformance factors (e.g.,

Goldberg, 1968; Wendleken & Inn, 1981). When information concerning job performance is limited or the characteristics of the target job are not well understood, raters may base their decisions on the information available to them and interpret this data in light of common stereotypes. This interpretation has been supported by studies in which raters were thoroughly familiar with job requirements and were provided with sufficient information to assess ratee performance. Under these conditions, Norton, et al. (1977) found no significant interactions between the sex of the rater and ratee, and following rater training, the sex of the rater no longer influenced the evaluations. Similar findings have been reported in a variety of experimental situations such as evaluating an individual's performance in a social situation (Taynor & Deaux, 1973), rating an applicant for a scholarship program (Deaux & Taynor, 1973), and when students act as raters in a work simulation exercise (Rose & Stone, 1978). Also, Schmidt and Johnson (1973) have demonstrated that rating biases due to the race and sex of raters and ratees may be reduced by training raters to become aware of these nonperformance influences on their evaluation of others.

Age, Education and Experience

Additional demographic variables which have been studied in relation to their effect on rater performance appraisal judgments include rater age, education level, and job experience or seniority.

Mandell (1956) and Klores (1966) arrived at conflicting conclusions regarding the influence of a rater's age on his evaluation processes. Mandell found older supervisors to be more lenient when

appraising the job performance of subordinates. Klores, however, found no consistent or significant relationship between a rater's age and the ratings made of other individuals.

Cleveland and Landy (1981) examined the influence of rater and ratee age on the performance evaluations of a sample of managers on two general and six specific performance measures. They found that younger managers tended to assign lower ratings than older managers on only one of the specific criterion measures - interpersonal skills. Modest significant interactions were also found between rater age and ratee age on the dimensions of self-development and interpersonal skills.

Klores (1966) also considered the raters' level of education in his study and found no significant relationship between this variable and the nature of their performance ratings. Cascio and Valenzi (1977) conducted a study to examine the main and interactive effects of rater and ratee educational levels and tenure on the job performance ratings of police officers. They found evidence of a significant main effect for rater educational level for four of the eight rating dimensions included on a behaviorally anchored rating scale. Although raters with more formal education did tend to give more severe ratings, the proportion of the total variance in the ratings accounted for by this factor was so small, the authors concluded that the result was of no practical significance. The possible interaction of rater and ratee education level also did not attain practical significance based on the eta-squared value for the F ratio.

The results of these studies appear to indicate that rater educational level has no meaningful impact on the rating process. However, this conclusion must be viewed cautiously as the range of educational levels included in the studies cited was very limited. For example, Cascio and Valenzi categorized raters as being either high or low on level of education. The low groups consisted of those individuals who had graduated from high school, but had completed less than two years of college. This severe restriction of range may have prevented detection of a stronger relationship if one did in fact exist.

A number of researchers have also presented findings indicating that a rater's length of service with an organization has no practical effect on performance ratings given to subordinates (Cascio & Valenzi, 1977; Klores, 1966; Stockford & Bissell, 1967; Zedeck & Kafry, 1977). However, Mandell (1956) found that senior supervisors tended to assign more lenient ratings than supervisors with shorter tenures. Also, Jurgenson (1950) reported a study in which he found a direct relationship between raters' length of experience and the reliability of the performance ratings they assigned.

As was the case with the personal and psychological rater characteristics studied, the conflicting results of the investigation of the impact of rater demographic factors on the rating process indicate that our knowledge of this domain is too incomplete to reach firm conclusions. Although additional research on many of the factors discussed is needed, perhaps it is time to redirect attention toward the study of contextual factors as they affect rater behavior and those elements which may be theoretically linked to the different

components of the performance appraisal process. With that in mind, the influence of work-related attitudes on behaviors is next considered before addressing the impact of these factors on the specific organizational behavior of evaluating the performance of others.

Attitudes and Behavior

Fishbein's Behavioral Intentions Model

Fishbein (1967) has proposed a model of motivation which is stated in terms of the prediction of specific behaviors. The model is based upon the assumption that a behavior (B) is a function of the intention (I) to perform that behavior. The theory attempts to predict behavioral intentions and assumes that if external circumstances are compatible with the intentions, the corresponding behaviors will be forthcoming. The external circumstances which determine whether specific behaviors are exhibited include but are not limited to the individual's ability and opportunity to perform the act in question.

Behavioral intentions are considered to be determined by two major factors. The first of these consists of the individual's attitude toward performing the behavior in question (A_B). This attitude is believed to be a function of the expected outcomes of engaging in a particular behavior as well as the subjective value assigned to these outcomes. It is represented as follows:

$$A_B = \sum_{i=1}^n b_i e_i$$

where b represents the belief that performing a certain act will lead to prescribed consequences (i), and e stands for the individual's evaluation of each outcome i (Fishbein & Ajzen, 1975, p. 301). This definition contrasts with the traditional view of an attitude (e.g., Kretch, Crutchfield, & Ballachey, 1962; Jones & Gerard, 1967). In this case an attitude represents the individual's feelings about engaging in a particular behavior rather than about the actual object or target toward which the action is directed.

The second element in the model is the subjective norm (SN) which involves the influence of the external social milieu on behavior. The subjective norm is concerned with a person's perceptions of whether or not the majority of individuals who are considered important think that one should or should not engage in a certain behavior. The theory defines the subjective norm as a summation of the products of an individual's normative beliefs (NB_i) or beliefs regarding what relevant others (i) think should be done and the person's motivation to comply (m) with the wishes of each referent i. This function may be algebraically expressed as:

$$SN = \sum_{i=1}^n NB_{im_i}$$

The specific reference groups and individuals which are considered relevant are expected to vary with different situations. The above equation implies that the products should be summed for each significant other with the sum being viewed as a "generalized normative belief" (Fishbein & Ajzen, 1975, p. 302).

When the two basic components of Fishbein's model are combined the central equation of the theory reads as follows:

$$B \sim I = (AB)w_1 + (SN)w_2$$

where w_1 and w_2 are theoretical importance weightings which are expected to vary according to the type of intention being predicted and the circumstances under which the prediction is made. The model as a whole can be conceptualized as a multiple regression equation where attitude and subjective norm are predictors and behavioral intention is the criterion. The entire model is represented in Figure 1.

Thus behavioral intentions may be thought to serve as intervening variables between the attitudinal and normative components of Fishbein's model and actual behavior. It is stressed that the level of specificity of the intentions dealt with by the theory is highly flexible. However, the ability of the model to predict behavioral intentions is dependent upon the degree of correspondence between the specificity of the intentional criterion and those of the attitudinal and normative elements. Similarly, the magnitude of the relation between intention and behavior rests upon the degree to which the measures of the behavioral intention and the behavior in question are

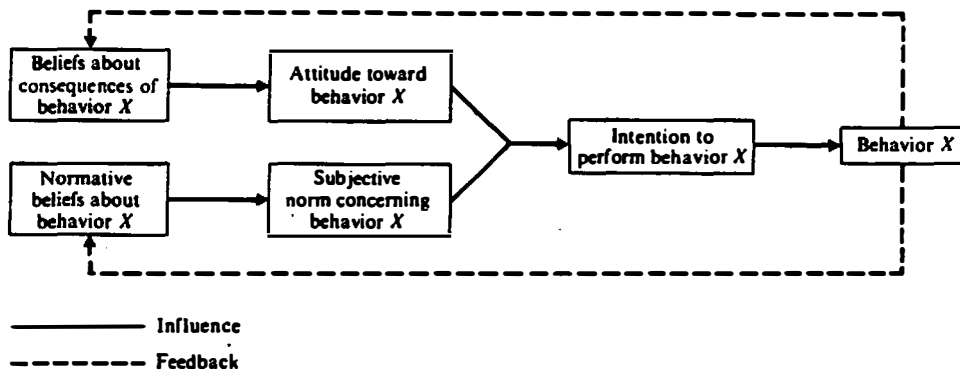


Figure 1. Schematic presentation of conceptual framework for prediction of specific intentions and behaviors.

Note: From Belief, Attitude, Intention, and Behavior (p. 16) by M. Fishbein and I. Aizen, 1975, Reading, Mass.:Addison-Wesley. Copyright 1975 by Addison-Wesley Publishing Company, Inc. Reprinted by permission.

similar in their level of specificity. The strength of such a relationship will also be affected by the stability of the intention and the extent to which realization of the intention is within the volitional control of the individual. Finally, Fishbein suggests that variables external to the model can influence the formation of intentions but only indirectly by bringing about some change in either or both of the two major factors in the model (A_B) and SN) or their relative weights (Fishbein & Ajzen, 1975, pp. 303-307).

Fishbein and his associates have undertaken an extensive program of research based upon the behavioral intention model in which they have attempted to predict intentions to engage in a wide variety of activities (see Ajzen & Fishbein, 1975). They have found considerable evidence supporting their contention that behavioral intentions can be accurately predicted on the basis of measurements of the theory's attitudinal and normative components. Behavior has been found to be predictable from behavioral intentions only if the specificity level of both are equivalent and the appropriate behavioral criteria are considered.

In his review of the research on the relation of attitudes to behavior, Wicker (1969) concluded that there is little evidence to support belief in a direct relationship. Fishbein (1967) has suggested that in order to understand this relationship, the nature of behavioral criteria must be reconceptualized. He differentiates between three types of behavioral criteria that should be considered in the study of the relation between attitudes and behavior (Fishbein, 1973). These three behavioral criteria are: 1) the single act

criterion; 2) the repeated observation criteria; and 3) the multiple-act criterion. The single act criterion is highly specific and involves an observable behavior directed toward a defined target object at a particular place and time. A repeated observation criterion is determined by observing the same behavior directed at different targets in varied contexts and at different times. Finally, a multiple-act criterion is based upon the observation of a variety of behaviors all directed toward the same target object, in the same context and within the same time frame.

Fishbein and Ajzen (1975, p. 358) suggest that the difference between attitude toward a behavior and attitude toward an object is critical to understanding the relationship of attitudes to behavior. They conclude that attitudes toward a behavior can be expected to predict a repeated observation criterion. An attitude toward an object will be predictive of a multiple-act behavioral criterion. In contrast, an attitude towards an object or behavior may or may not correlate with a single act criterion and this is the primary reason for the general lack of association between attitudes and behavior that has characterized past research in this area.

Fishbein and Ajzen (1975, pp. 359-361) re-analyzed a number of published studies of the attitude-behavior relationship in light of their behavioral criteria theory and have generally found support for their conclusions. In addition, they have conducted further tests of the theory and were able to predict repeated observation and multiple-act criteria regarding religious behavior with traditional attitude measures. The Fishbein model has also been effective in predicting

job-related behavior in organizational settings. Hom, Katerberg, and Hulin (1979) used this model to account for a significantly greater proportion of the variance in turnover of a sample of National Guard members than was explained by either of two more traditional attitudinal approaches - a measure of job satisfaction and Porter's organizational commitment model (Porter, Steers, Mowday & Boulian, 1974).

Wiener's Model of Organizational Commitment

Wiener (1982) has presented a model of work behavior which is structured upon the conceptual framework of Fishbein's behavioral intentions model. Wiener's model is built around the construct of organizational commitment which is viewed as a motivational phenomenon intervening between certain antecedents and behavioral outcomes. The basic premise of the model is that a complete explanation of an individual's work behavior requires looking beyond behavior - outcome contingencies to also consider the importance of internalized normative pressures.

Definitions of organizational commitment have centered upon the extent to which individuals personally accept an organization's goals, put a high premium on membership in the organization, and intend to work hard to help the organization fulfill its purpose (Porter, Steers, Mowday and Boulian, 1974; Steers, 1977). Commitment goes beyond simple attachment to the organization. It entails an active identification on the part of the individual with the goals and values of the organization, with the individual's role in the organization,

and to the organization itself apart from its purely instrumental worth (Buchanan, 1974).

Staw (1977) has made the distinction between behavioral commitment and attitudinal commitment. Behavioral commitment is viewed as the result of irrevocable decisions or behaviors which prevent people from backing out of an organization without considerable difficulty. Attitudinal commitment is the extent to which an individual identifies with goals and values of the organization and integrates them into a system of personal goals and values.

Research based upon the attitudinal or identification view of organizational commitment has primarily focused upon determining antecedents and outcomes. Steers (1977) has suggested that the determinants of organizational commitment may be grouped into three main categories. These are: 1) personal characteristics (e.g., needs, values, age, education, central life interests); 2) job characteristics, (e.g., challenge, feedback, task identity, opportunities for social interaction); and 3) work experiences (e.g., organizational dependability and trust, personal importance to the organization, rewards).

Studies addressing the behavioral correlates of commitment have shown that the outcomes having the strongest association are turnover and intention to leave the organization (Hom, Katerberg, & Hulin, 1979; Mowday, Steers, & Porter 1979; Porter et al., 1974; Steers, 1977). Investigations of the relationships between organizational commitment and indices of job performance have generally produced disappointing results (Mowday, Porter, & Dubin, 1974; Steers, 1977;

Wiener & Vardi, 1980). It is possible that ability, role clarity, and whether or not performance was within the volitional control of the individual would moderate this relationship. Another possible explanation for the modest correlations between these factors may be found in Fishbein's suggestion that an appropriate behavioral criterion must first be selected. If performance was based on a multiple-act criteria which could be viewed as a behavioral measure of organizational commitment, it is likely that the strength of the attitude - behavior relation would be significantly enhanced.

Wiener (1982, p. 419) has suggested that the principal reason for the fragmented nature of the results of research on organizational commitment is the lack of a comprehensive theoretical model of this construct linking it to other work attitudes and motivational processes. The model he has presented is an attempt to fill this void. A schematic diagram of the model is shown in Figure 2.

The model suggests that internalized normative beliefs lead to the generation of organizational commitment. Similarly, instrumental beliefs result in instrumental motivation. Together organizational commitment and instrumental motivation determine work-related behavioral intentions and behaviors. Behavioral outcomes are characterized in terms of their specific natures and underlying attributes.

Wiener refers to the attitudinal component of Fishbein's model as instrumental motivation. The instrumental beliefs upon which attitudes are founded correspond to the cognitions involved in expectancy/valence models of motivation, namely that effort will lead to performance and that performance will lead to valued rewards

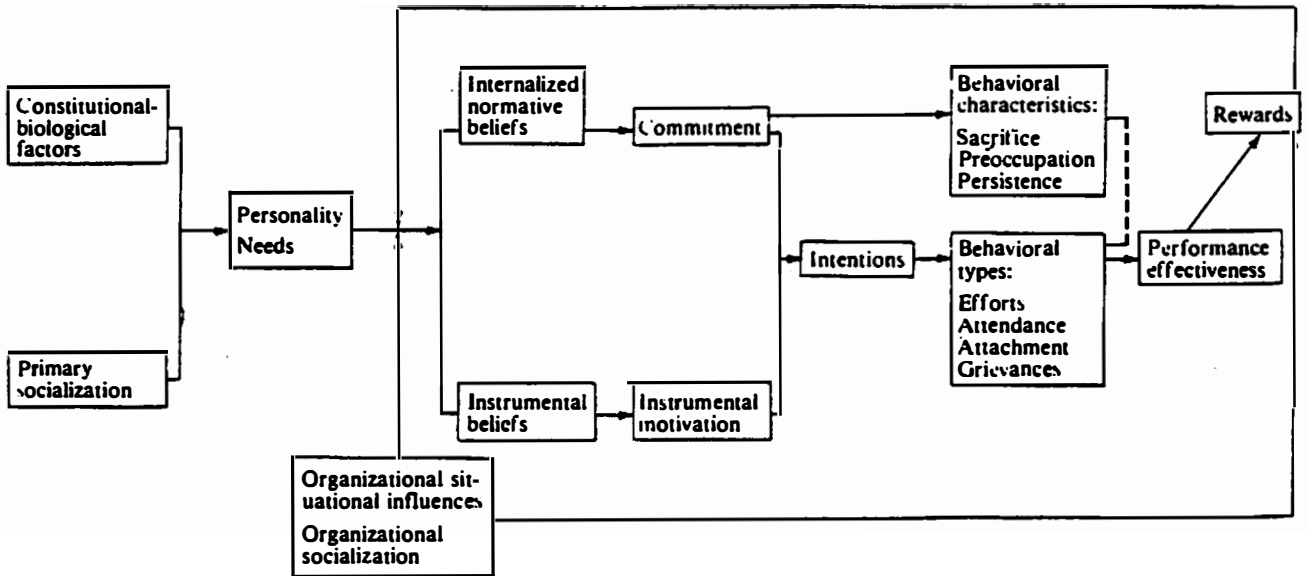


Figure 2. Wiener's model representing relationships between organizationally related behaviors, beliefs concerning those behaviors, and commitment and instrumental motivation.

Note: From "Commitment in organizations: A normative view" by Y. Weiner, 1982, *Academy of Management Journal*, 3, p. 420. Copyright 1982 by the Academy of Management. Reprinted by permission.

(Nadler & Lawler, 1977; Vroom, 1964). The attitudinal component refers to an individual's feelings toward the act of exerting effort to perform a particular behavior that is instrumental in obtaining a desired outcome. Thus, the attitudinal leg of the model represents the utilitarian, self-serving aspect of the entire motivational process.

Wiener uses the subjective norm portion of Fishbein's model as the basis for his conceptualization of organizational commitment. He focuses on the portion of the subjective norm which is based upon personal normative beliefs or personal moral standards in regard to specific behaviors. These personal beliefs and standards develop when expectations of significant others are internalized. Once these norms have been internalized, the behaviors that they guide are no longer motivated by the performance/outcome contingencies that initially caused them to occur. Wiener considers the personal aspect of the subjective norm as the basis of organizational commitment.

"Organizational commitment is viewed as the totality of internalized normative pressures to act in a way that meets organizational goals and interests. The stronger the commitment, the stronger is the person's predisposition to be guided in his actions by such internalized standards rather than by consideration of the consequences of these actions. Thus, committed individuals may exhibit certain behaviors not because they have figured that doing so is to their personal benefit, but because they believe it is the right and moral thing to do." (Wiener, 1982, p. 421)

This definition of commitment strongly emphasizes the point that it is a distinct concept from the instrumental motivation process underlying the attitudinal component of the model. In accord with Fishbein's theory, Wiener believes that the relative weights given to

both components in predicting behavioral intentions and behavior will be a function of the attributes of the situation in which the behavior is performed, the individual performing the behavior, and the behavior itself. He goes on to explain that while commitment and instrumental motivation may precipitate the same forms of behavior in an organizational setting, patterns or sequences of behaviors determined primarily on the basis of commitment should have some unique qualities. Specifically, such behaviors will be characterized by one or more of the following: personal sacrifice for the organization, preoccupation with organizational goals, and persistence in behaviors favored by the organization and that are not dependent upon externally mediated rewards.

The major contribution of Wiener's model is that it provides a sound theoretical basis to support the premise that work-related behavior may be determined not only by instrumental motivation but also by internalized normative beliefs fostered by the organization. Once established, the influence exerted by organizational commitment on behavior may be very strong and it is independent of situational circumstances and reward contingencies.

Job Involvement and Instrumental Motivation

Job involvement, like organizational commitment, is another work-related individual difference factor which has been the subject of considerable theoretical and empirical study (Rabinowitz & Hall, 1977). Perhaps the most formidable problem in researching this topic that is immediately evident from a review of the literature concerns

the lack of agreement over exactly what constitutes job involvement. Though many different terms have been used in connection with this concept (e.g., central life interest, ego-involvement, commitment, etc.) two conceptualizations have emerged as the core elements of the meaning of job involvement. Rabinowitz and Hall (1977) refer to the first of these definitions as the performance - self-esteem contingency. In this view, job involvement is dependent upon the extent to which job performance influences a person's sense of personal worth and may be conceived of as the value an individual puts on superior job performance. In the second view, job involvement is seen as a component of a person's self-image or the degree to which the individual psychologically identifies with work.

Saleh and Hosek (1976) attempted to clarify the theoretical basis of job involvement by factor analyzing items from scales developed by other researchers to measure this construct. Their results indicate that a three factor solution is optimal. Factor I is considered a measure of a person's active participation in his job. Factor II corresponds to the view of work as a central life interest for the individual. Factor III indicates the extent that job performance influences a person's concept of self-worth and corresponds to the performance - self-esteem contingency definition. This study in conjunction with the work of other researchers (e.g., Lawler & Hall, 1970; Lodahl & Kejner, 1965; Smith, 1981), indicates that job involvement may most realistically be conceived of as a multidimensional concept with the individual's self-concept serving as a common underlying bond between facets.

Although generally referred to as an individual difference variable, definitions of job involvement strongly imply an attitudinal disposition toward one's job. This is particularly evident for the performance - self-esteem definition when attitude is conceived of in the terms of Fishbein's model. A person is considered to be job involved to the extent that he perceives performance as being instrumental to enhancing his self-esteem. Hence, according to the behavioral intentions model, job involvement may be conceived of as an attitude toward a behavior, namely, exerting effort in order to perform well on the job. This attitude is determined on the basis of the belief that superior performance will positively influence the individual's sense of self-worth.

The attitudinal aspect of job involvement is closely related to the motivation of job performance to the extent that feelings of enhanced self-esteem are viewed as salient rewards. Accordingly, job involvement may be considered as part of the instrumental motivation component of Wiener's model of work behavior. Of course, increased self-esteem is only one of an array of outcomes that motivate job behavior. However, in many circumstances externally mediated rewards are not linked to specific job behaviors and feelings of improved self-image are the only direct personal consequence of these actions. This would seem to be particularly true for job behaviors which only the individuals themselves are aware they are performing. Also, certain behaviors have no external standards for comparison and their performance may only be assessed by the individual in terms of the amount of .cp 5 effort expended. Under these circumstances, the

motivational impact of job involvement on behavior could be considerable.

A substantial number of research studies have examined the relationship of job involvement to work outcomes. The outcomes most consistently found to be predicted by job involvement are absenteeism and leaving the organization (Farris, 1971; Patchen, 1965; Siegel & Ruh, 1973). However, the results of efforts to find a significant relationship between job involvement and indices of job performance have generally been disappointing (Goodman, Rose, & Furcon, 1970; Lodahl & Kejner, 1965; Lawler & Hall, 1970). Both sets of findings may be explainable in light of Fishbein's model.

According to this model, attitude toward a behavior (i.e., job involvement) is expected to be related to a repeated-observation criterion. Showing up at the job each day may be considered such a measure. However, job performance, as it has typically been conceptualized in the studies cited might be more accurately characterized as a multiple-act criterion. Hence, Fishbein's model would not predict that job involvement would correlate with such a broad measure as overall job performance. In addition, none of these studies considered the impact of the subjective norm in determining the performance outcome. Wiener's model further illustrates the theoretical basis for considering this element in predicting work-related behaviors.

Thus, a strong argument can be made for including job involvement as part of the instrumental motivation component of Wiener's model. The principal objective of the present study is to apply this altered

version of Wiener's model to the prediction of a specific form of work behavior - the observation and rating of employee job performance. The relevancy of this application will be tested by examining the relationship between various components of the new model and different forms of predicted behavior. The specific hypothesized relationships to be investigated are presented in the following chapter.

CHAPTER III

THEORETICAL MODEL AND HYPOTHESES

The Model

The primary purpose of the proposed model is to predict the influence of rater job involvement and organizational commitment on the process of appraising employee job performance in a work setting. A diagram of the model is presented in Figure 3.

Borman (1978, p. 141) has suggested that the making of performance rating judgments consists of a three step process; 1) the observation of work-related behavior; 2) the evaluation of these behaviors; and 3) the weighting of these behaviors and the determination of performance ratings. These three steps serve as the foundation upon which the proposed model is constructed.

The model begins with ratee job behaviors. These behaviors are observed and stored by the raters. At some point in time, these behaviors are recalled and evaluated in terms of the rater's assessment of their effectiveness. At the time of the formal performance review, ratings of the ratee's work behavior are assigned in accordance with the parameters of the appraisal system in use. The outcomes of this process are job performance ratings which are uniquely characterized in terms of their psychometric qualities, viz., reliability, validity, accuracy, and the prevalence of constant rating errors.

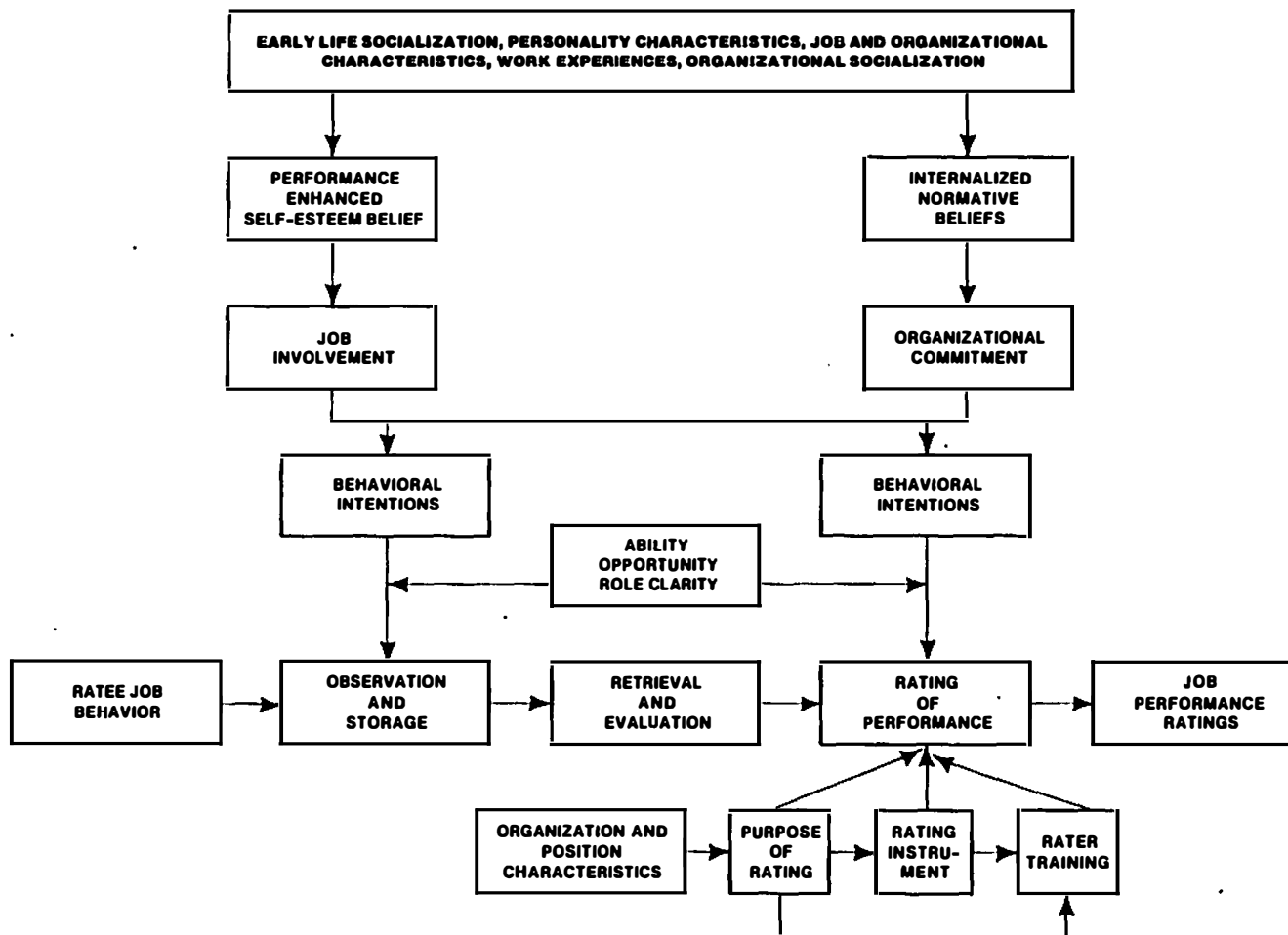


Figure 3. Schematic diagram of the theoretical model for prediction of rating behavior on the basis of rater job involvement and organizational commitment levels.

The model posits that the performance appraisal process is influenced at the observation stage and the rating stage by the rater's level of job involvement and organizational commitment. The mechanism by which this influence occurs follows from Wiener's (1982) application of Fishbein's (1967) behavioral intentions model to the prediction of work-related behavior in an organizational context.

Job involvement is conceptualized in terms of the performance - self esteem contingency definition and represents one particular form of instrumental motivation. The first steps in this process involve the formation of attitudinal beliefs regarding the performance - self-esteem contingency and internalized normative beliefs based on organizational subjective norms. Suggested antecedents of these beliefs include early life socialization (Lodahl, 1964), personality characteristics (Hall & Mansfield, 1971), job and organizational characteristics (Vroom, 1962), work experiences (Steers, 1977), and organizational socialization (Wiener, 1982). Fishbein & Ajzen (1975) propose that these beliefs may possibly be formed by means of three different processes - observation, inference, and acceptance of information. The specific means by which the beliefs in question are formed is not central to understanding how the model works. However, it is critical to understand that the characteristics of the organization and the context in which performance appraisal is conducted are very likely to have a major impact on the process of belief formation.

Considering first the upper-left branch of the model, beliefs about the performance - self-esteem contingency lead to the formation of an attitude towards job performance which has been defined as job

involvement. On the right side of the model, as Wiener suggests, internalized normative beliefs lead to the development of organizational commitment. Next, job involvement and organizational commitment interact to result in the formation of behavioral intentions which pertain to the observation of job behavior of target individuals and the assignment of ratings indicating the quality of their performance. The relative importance of job involvement and organizational commitment in determining the various behavioral intentions will vary according to the situation and the specific nature of the behavior being predicted.

The behavioral intentions formed are theorized to affect rater behavior at either the observation or rating steps in the appraisal process. Although intentions may affect the evaluation of others' performance, it is very unlikely that specific behavioral predictions could be made because of the implicit nature of this step in the process (Borman, 1979a) and the difficulty of conceptually separating this activity from the other two steps.

Behavioral intentions represent the motivational element in the generation of behavior. However, as numerous theorists have suggested (e.g., Fishbein & Ajzen, 1975; Porter & Lawler, 1968; Vroom, 1964) motivation alone will not ensure task performance. A person must possess in addition, the abilities and skills required to perform a particular act. Also, the opportunity to engage in the behavior must exist. Finally, role clarity or an understanding of the nature of the task is also necessary for successful task performance. These three

factors mediate the relationship between behavioral intentions and overt behaviors and they are represented accordingly in the model.

A final sequence of elements is also included in the model because of its direct influence on the process of assigning performance ratings. This sequence is derived from Landy and Farr's (1980) process model of performance rating. The sequence begins with the characteristics of the organization and the position(s) of the individuals whose performance is to be appraised. These factors represent the context in which the rating occurs and directly determine the purpose for which the rating is performed. The organization may require an appraisal of employee performance for administrative purposes, for use in employee counseling, or as a criterion measure against which to validate a selection procedure. Similarly, performance ratings may be used for different purposes at different levels in the organization.

The model assumes that the purpose of the appraisal will have a strong impact on the process of assigning ratings. One indication of this effect is the commonly reported finding that ratings made for research purposes only are considerably different from ratings made in the same setting but for use in making administrative decisions (Borrensen, 1967; Bernardin, 1978).

The purpose of the ratings and the organizational context will influence the choice of the rating instrument to be used. The type of judgments required will be mandated by the nature and level of specificity of the information required from the rating process. For example, global dimensional assessments may be suitable for use in

making salary decisions whereas behavior specific ratings may be more desirable for use in providing detailed performance feedback to employees. The different cognitive processes involved in utilizing various types of appraisal instruments will have a major impact on the process of rating behavior.

Rater training is included as a final element in the model. This type of training focuses upon the use of the particular rating instrument(s) selected as well as the nature of constant rating errors. Training of this sort has generally been found to increase the reliability and accuracy of ratings and decrease the occurrence of rating errors (Borman, 1979b; Warmke & Billings, 1979; Zedeck & Cascio, 1982). However, a number of practical questions regarding rater training remain unanswered including what form of training is most effective, how extensive rater training must be to produce results, and how long the effects of rater training persist. In any event, research has shown that even very brief training programs may improve the quality of ratings (Borman, 1975). Training may be viewed as increasing raters' role clarity with regard to the task of assessing the performance of others.

The proposed model is specifically intended to demonstrate the means by which job involvement and organization commitment are believed to influence the performance rating process. Both the observation and rating phases of the process are viewed as constituting repeated-act behavioral criteria. Thus, according to Fishbein's theory, behavior in both contexts should be predictable on the basis of individuals' attitudes and subjective norms regarding these

actions. This proposition will be empirically tested in the proposed study.

The model presented is not intended to be a complete conceptualization of the rating process. A number of important elements have been excluded because they are not directly relevant to the theoretical issues addressed in the present study. Among these factors are the various personal and demographic characteristics that raters bring to the appraisal process. Also excluded are the nonperformance characteristics of ratees that introduce bias in raters' evaluations. These factors have been empirically demonstrated in many instances to influence performance ratings independently and through certain interactions with each other. However, in order for this study to break new ground in researching the rating process, these other characteristics will be assumed to vary randomly across raters and ratees.

Finally, observation of the behavior of others and the assignment of meaningful performance ratings in a real organizational context are obviously motivated by other outcomes besides enhanced self-esteem. In order to study the impact of this one source of motivation, the externally mediated rewards of the job have also been excluded from the model. This is justified on an intuitive as well as practical basis. Performance of the behaviors in question is rarely, if ever, directly linked to externally mediated rewards in an organizational setting. Typically, there are no standards against which to evaluate how well someone observes and appraises the performance of others. In most real world contexts only the rater can assess how well he or she

has performed these behaviors. Under such circumstances it is not unlikely that feelings of enhanced self-esteem would be a particularly potent motivator of behavior.

The Hypotheses

The hypotheses presented are intended to test the predictions made by the model regarding the effect of raters' job involvement and organizational commitment on the psychometric characteristics of job performance ratings they assign. The model predicts that the degree to which a rater is job involved and committed to the organization will have an effect on the appraisal process during both the observation and rating stages.

The hypotheses are tested in a field setting using data from supervisor assessments of subordinate job performance. Supervisor judgments comprise the vast majority of appraisal ratings completed within organizations (Campbell, Dunnette, Lawler, & Weick, 1970). It is generally assumed that an employee's immediate supervisor is more familiar with the requirements of the individual's job and has a better overview of his/her day-to-day performance than any other person in the organization (Lawler, 1967). In view of this, the model has been conceptualized to apply primarily to supervisory ratings of subordinate job performance. Whether the model is equally applicable in other rating contexts and for raters having a non-supervisory relationship to ratees remains open to speculation and should be investigated in the future.

During the observation phase, a rater's level of job involvement is believed to have a substantial impact on his/her behavior to the extent that monitoring the performance of subordinates is an important part of the supervisor's role. If observing how subordinates perform on their jobs is a critical aspect of the rater's job, then more highly job involved supervisors should endeavor to increase their self-esteem by performing this task as thoroughly as possible. Thus, given ample ability and sufficient opportunity, the model proposes that highly job involved persons will observe the job performance of their subordinates more frequently, in greater detail, and more thoroughly than supervisors who are less job involved.

Similarly, in situations where organizations have established well-known company policies regarding the careful monitoring of subordinate job behavior, it is postulated that supervisors who are highly committed to the organization will internalize this norm and observe their subordinates more closely than will less committed individuals. Where the organization's policy in this domain is not clearly specified, organizational commitment will have less of an effect on the careful monitoring of subordinate daily behavior. One example of such a situation might be a company which has not previously required supervisors to conduct a detailed appraisal of their subordinates' performance. When this is the case, it is proposed that rater job involvement levels will have a stronger comparative impact on the observation of employee behavior by supervisors.

During the rating phase of the appraisal process, organizational commitment is presumed to affect rater behavior to the extent that the

organization's interests in the assessment program are typically conveyed to supervisors as part of the instructions for conducting the appraisals. An increasing number of organizations are introducing formal training programs for instructing raters in how to use the rating system and minimize the effect of bias on their assessments of subordinates (Latham & Wexley, 1981). This type of training typically stresses the importance of minimizing rating errors and providing reliable, accurate, and valid evaluations. If these objectives are identified through training programs as important goals of the organization, assuming that the training is adequate and that the raters are able to comprehend the instruction, it is then proposed that highly committed individuals will tend to assign ratings which are more reliable and less subject to constant rating errors than the ratings given by supervisors less committed to the organization and its objectives.

Rater job involvement is predicted to influence rating behavior to the degree that supervisors perceive the assignment of psychometrically superior ratings as enhancing their self-esteem. However, in most instances, raters do not have any means by which to assess the quality of the ratings they assign nor do they receive feedback regarding how good a job they have done in appraising their subordinates. This absence of a strong linkage between performance and improved self-image suggests that the level of rater job involvement will not have as strong an impact on the assignment of ratings as it will on the observation of subordinate behavior. The effort - reward contingency is clearer in the case of observation. Also raters are

better able to personally assess how good a job they are doing in monitoring their subordinates' job performance.

The proposed model indicates that the effects of rater job involvement and organizational commitment on behavioral intentions and actual behavior are predicted to influence rater behavior in all phases of the appraisal process. However, on the basis of the relationships discussed above, job involvement is predicted to have a comparatively larger impact than organizational commitment during the observation phase of the appraisal process. Conversely, it is suggested that commitment will generally have a more prominent role than job involvement in the determination of rater behavior during the rating phase of the process.

The effects of job involvement and organizational commitment on rater behavior are predicted to be additive and compensatory. Hence, in order to examine the influence of these factors on behavior, a variable will be created by summing the scaled measures of both of these factors. Job involvement and organizational commitment will be weighted equally in forming this combined measure as there is no means to determine their relative importance weightings for different situations a-priori.

Some of the hypotheses considered are tested using either raters or ratees as the units of analysis. When raters are used, the sum of each supervisor's mean item scores on the job involvement and organizational commitment scores are used in the accompanying statistical analyses. When ratees are the basic experimental unit, their raters are divided into two groups on the basis of whether the sum of their

involvement and commitment scores is above or below the median value for the entire rater sample.

The proposed model is intended as a guide for the prediction of rater behavior in the context of the rating process. For purposes of testing the model, this behavior is measured by examining the psychometric characteristics of the performance ratings assigned. The specific characteristics considered include the prevalence of constant rating errors (leniency/severity, central tendency, range restriction, and halo) as well as reliability and the convergent and discriminant validity of the ratings. Accuracy, per se is not assessed because of the difficulty in obtaining "true" measures of performance in a field setting.

The first set of hypotheses concerns the rating error of leniency or the converse, severity. Leniency has been defined by Saal & Landy (1977) as the tendency to assign a higher or lower rating to a ratee than is warranted by that person's behavior. Sharon and Bartlett (1969) have conceptualized the term as a shift in mean ratings from the midpoint of the rating scale in either the favorable (positive leniency) or unfavorable (negative leniency or severity) directions. For ease of explanation, this rating error will be referred to solely as leniency in the remainder of this report although it is meant to refer to systematic bias in the level of ratings assigned in either the positive or negative directions.

The hypotheses concerning leniency are as follows:

Hypothesis 1A. A composite measure of rater job involvement and organizational commitment will be inversely related to the amount of leniency present in assigned ratings of subordinate job performance.

This hypothesis and the corresponding ones that follow for each of the other rating errors considered addresses the fundamental tenet of the model that job involvement and organizational commitment act concomitantly to inversely effect the extent of systematic bias in the ratings assigned. In addition to considering the effect of the combined measure of these two attitudes, the individual influence of job involvement and organizational commitment on leniency and the other three rating errors will also be assessed.

Hypothesis 1B. The relationship of the composite measure of rater job involvement and organizational commitment with the amount of leniency in assigned ratings of subordinate job performance will be stronger for ratings made on a behaviorally-based rating format than on a trait-based rating format.

Hypothesis 1B is based on the fact that certain types of behaviorally-based rating formats (e.g., checklists) are particularly susceptible to leniency errors (Cascio, 1978, p. 325; Bass; 1956; Borman & Vallon, 1974). As a larger amount of leniency is believed to be present in the ratings made on behaviorally-based scales to start, controlling for the effect of rater attitudes on the characteristics

of ratings assigned may accentuate the difference between ratings made using behaviorally-based and trait-based rating formats.

Hypothesis 1C. Rater ability will moderate the relationship between the composite measure of rater job involvement and organization commitment and the amount of leniency present in assigned ratings of subordinate job performance.

This hypothesis and the corresponding ones which follow are included to test the portion of the model which predicts that rater ability will mediate the relationship between behavioral intentions and overt behaviors. It is predicted that raters who are characterized as having higher levels of task related ability will demonstrate a stronger association between behavioral intentions and actual rating behavior and vice versa. While the model posits that role clarity and opportunity will also moderate this relationship, only supervisor ability could be reasonably assessed in the field setting of this study and thus was the only potential moderator variable considered. The influence of the other two factors remains to be examined in a future research effort.

Hypothesis 1D. A larger proportion of the variance in the measure of leniency in assigned ratings of subordinate job performance will be accounted for by rater organizational commitment than by rater job involvement.

The rationale for hypothesis 1D, and the corresponding hypotheses that follow, is based upon the belief that the factors causing the occurrence of constant rating errors primarily affect the appraisal process during the rating phase. The model suggests that organizational commitment will have a stronger influence on rater behavior during this part of the process than will job involvement.

The next rating error to be considered is central tendency. Central tendency is defined as a rater's unwillingness to assign extreme ratings in either the positive or negative direction. This causes ratings to cluster around the midpoint of the scale and restricts the range of the distribution (Guilford, 1954, p. 278). Use of the terms central tendency and restriction of range has led to some confusion in the literature as to whether they constitute separate or synonymous concepts. Saal, Downey, and Lahey (1980) have clarified the issue by stating that although central tendency implies range restriction, the converse is not necessarily true. In instances where ratings are clustered around a point in the scale continuum above or below the midpoint, restriction of range may be evident while central tendency is not. It is therefore advisable to clearly distinguish between the two concepts. In accord with this suggestion, these two types of constant errors will be considered separately.

The following hypotheses concerning central tendency will be tested in this study:

Hypothesis 2A. A composite measure of rater job involvement and organizational commitment will be inversely related to the amount of central tendency present in assigned ratings of subordinate performance.

Hypothesis 2B. The relationship between the composite measure of rater job involvement and organizational commitment and the amount of central tendency in assigned ratings of subordinate job performance will be stronger for ratings made on a trait-based rating format than on a behaviorally based rating format.

The rationale for hypothesis 2B is essentially the same as that of hypothesis 1b except that trait-based scales are purportedly more prone to the error of central tendency than behaviorally-based rating scales. This would appear to be especially true for scales in the form of behavioral checklists where the degree of inference required on the part of the rater is minimized (Cascio, 1978).

Hypothesis 2C. Rater ability will moderate the relationship between the composite measure of job involvement and organization commitment and the amount of central tendency present in assigned ratings of subordinate job performance.

Hypothesis 2D. A larger proportion of the variance in the measure of central tendency in assigned ratings

of subordinate job performance will be accounted for by rater organizational commitment than by rater job involvement.

In terms of its relevance as an index of rating quality, restriction of range is defined as the degree to which ratings discriminate among different ratees in terms of their respective levels of performance (Motowidlo & Borman, 1977). Ratings characterized by insufficient variance are of limited use in making administrative decisions or for serving as a criterion for validation purposes. Thus, restriction of range is an important and legitimate standard to consider in assessing the quality of performance ratings (Saal, et al, 1980). The following hypotheses concerning restriction of range in assigned performance ratings will be tested:

Hypothesis 3A. A composite measure of rater job involvement and organizational commitment will be inversely related to the amount of restriction of range present in assigned ratings of subordinate job performance.

Hypothesis 3B. The relationship between the composite measure of rater job involvement and organizational commitment with the amount of restriction of range in assigned ratings of subordinate job performance will be stronger for ratings made on a trait-based rating format than on a behaviorally-based rating format.

Hypothesis 3B is based on the same rationale as the corresponding prediction for central tendency described above.

Hypothesis 3C. Rater ability will moderate the relationship between the composite measure of job involvement and organizational commitment and the amount of restriction of range present in assigned ratings of subordinate job performance.

Hypothesis 3D. A larger proportion of the variance in the measure of restriction in range in assigned ratings of subordinate job performance will be accounted for by rater organizational commitment than by rater job involvement.

It is generally agreed that the most pervasive form of constant rating error is the halo effect (Cascio, 1978, p. 321). First named by Thorndike (1920), halo is perhaps the most clearly defined of the forms of rating bias considered thus far. Halo may be conceptualized as the tendency for raters to assign ratings on different performance dimensions on the basis of a global impression of the ratee. This practice results in a failure to distinguish between different levels of ratee performance in each of the various domains assessed. Ratings characterized by high levels of halo will evidence spuriously high intercorrelations between dimensions as a result of the rater's inability to adequately distinguish between different aspects of ratee performance.

The hypotheses to be tested concerning the halo error are as follows:

Hypothesis 4A. A composite measure of rater job involvement and organizational commitment will be inversely related to the amount of halo present in assigned ratings of subordinate job performance.

Hypothesis 4B. The relationship between the composite measure of rater job involvement and organizational commitment with the amount of halo in assigned ratings of subordinate job performance will be stronger for ratings made on a trait-based rating format than on a behaviorally-based rating format.

Hypothesis 4B is based on the model and the commonly noted tendency for the halo effect to be more pronounced with trait-based rating scales than with behaviorally based measures (e.g., Burnaska & Hoffman, 1974; Borman & Dunnette, 1975).

Hypothesis 4C. Rater ability will moderate the relationship between the composite measure of job involvement and organizational commitment and the amount of halo present in assigned ratings of subordinate job performance.

Hypothesis 4D. A larger proportion of the variance in the measure of halo in assigned ratings of subordinate job performance will be accounted for by

rater organizational commitment than by job involvement.

The reliability of performance ratings represents another critical criterion for assessing the quality of the assessments produced during the rating process. Reliability is an index of the stability and consistency of the measurement results and it determines the upper limit possible for the validity of the rating process (Nunnally, 1967).

The reliability of a performance rating measure, while influenced to some extent by all aspects of the rating process, is directly affected by the observational and rating proclivities of the raters. Thus, it is likely that the attitudinal dispositions of the raters may also have a significant impact on the reliability of their ratings of the performance of others. The hypothesis to be tested in this regard is as follows:

Hypothesis 5. A composite measure of rater job involvement and organizational commitment will be directly related to the reliability of ratings assigned to assess subordinate job performance.

The relationship between levels of rater job involvement and organizational commitment and the construct validity of performance ratings assigned will next be addressed. The most widely used approach for operationally assessing construct validity is the multitrait - multimethod matrix initially proposed by Campbell and

Fiske (1959). This procedure permits the empirical demonstration of the convergent and discriminant validity of the ratings. If performance dimension ratings are obtained for ratees from two or more sources (e.g., different raters or different rating formats), then the degree of agreement between the various sources on ratings of the same traits constitutes a measure of convergent validity. Similarly, to the extent that different sources of rater evaluations on different dimensions disagree, discriminant validity may be said to exist.

Many researchers in the field have used this approach to study the rating process (e.g., Holzbach, 1978; Kavanaugh, MacKinney, & Wollins, 1971; Lawler, 1967). Most of the reported studies have found support for the existence of convergent validity in the ratings examined. However, the evidence for discriminant validity has generally been disappointing. In a unique study which attempted to create a nearly ideal environment for obtaining performance ratings, Borman (1975) did find a level of discriminant validity that was considerably higher than the levels typically been reported from studies conducted in applied settings. Borman's study indicated that various aspects of the rating process are functioning to suppress discriminant validity in ratings. It is possible that some of these elements are directly related to the characteristics of the individuals responsible for providing the ratings. Factors affecting rater attitudes and motivation may very likely impact upon the validity-related characteristics of the performance ratings they assign to others. The following hypothesis is proposed to examine this premise within the context of the proposed model:

Hypothesis 6. A composite measure of rater job involvement and organizational commitment will be directly related to the levels of convergent and discriminant validity in assigned ratings of subordinate job performance.

CHAPTER IV

METHOD

Subjects

The subjects in this study were 199 sales clerks and 38 section leaders employed at three locations of a department store chain. The number of sales clerks supervised by each section leader ranged from three to ten with a mean of 4.80. Participation in the study was voluntary. Eighty-six percent of the sales clerks and ninety-five percent of the section leaders employed in the target stores between October 1978 and June 1979 were included in the sample.

The sales clerk sample was comprised of 179 females (89.9%) and 20 males (10.1%). The average tenure of the sales clerks was 2.62 years (standard deviation = 3.91). The section leader sample included 33 females (86.8%) and 5 males (13.2%) with a average tenure of 5.41 years (standard deviation = 4.96). Demographic characteristics of the sales clerk sample are shown in Table 1. Descriptive measures for the section leader sample are summarized in Table 2.

Job Performance Measures

Three performance rating scales were developed for use in assessing the job performance of sales clerks: a behavioral checklist (BCL), a trait-based, graphic rating scale (GRS), and a single item assessment of overall job performance. A copy of the rating instru-

Table 1

Demographic Characteristics of the Sales Clerk Sample

Variable	Number	Percentage
Age		
16 - 21	27	13.6
22 - 30	98	49.2
31 - 40	19	9.5
41 - 50	25	12.6
51 - 60	24	12.1
61 - 70	6	1.0
Race		
Black	33	16.6
White	152	76.4
Asian	4	2.0
Other	10	5.0
Educational Level		
Fewer than 12 Years	13	6.5
High School Graduate	158	79.4
Associate's Degree	8	4.0
College Degree	18	9.0
Graduate School	2	1.0
Employment Status		
Full Time	129	64.8
Part Time	70	35.2
Tenure		
0 to 6 Months	72	36.2
7 to 12 Months	44	22.1
13 to 24 Months	21	10.6
25 to 36 Months	13	6.5
37 to 48 Months	25	12.6
49 to 72 Months	12	6.0
72 to 120 Months	3	1.5
Over 120 Months	9	4.5
Work Location		
Store 1	67	33.7
Store 2	77	38.7
Store 3	55	27.6

Table 2

Demographic Characteristics of the Section Leader Sample

Variable	Number	Percentage
Age		
22 - 30	16	42.1
31 - 40	3	7.9
41 - 50	8	21.0
51 - 60	9	23.7
61 - 70	2	5.3
Race		
Black	3	7.9
White	30	78.9
Asian	2	5.3
Other	3	7.9
Educational Level		
Fewer than 12 Years	4	10.5
High School Graduate	25	65.8
Associate's Degree	4	15.8
College Degree	3	7.9
Employment Status		
Full Time	100	100.0
Part Time	0	0.0
Tenure		
0 to 6 Months	2	5.3
7 to 12 Months	6	15.8
13 to 24 Months	9	23.7
25 to 36 Months	10	26.3
37 to 48 Months	6	15.8
49 to 72 Months	3	7.9
Over 72 Months	2	5.3
Work Location		
Store 1	14	36.8
Store 2	12	31.6
Store 3	12	31.6

ment developed including the rating instructions for each section and the final format of the scales is shown in Appendix A.

Behavioral checklists provide the rater with lists of specific job-related behaviors. The rater's task is to indicate which statements characterize the behavior of the ratee in general. This process objectifies the rating process by casting raters in the role of reporters of behavior rather than evaluators (Cascio, 1978). Stockford and Bissell (1949) have suggested that ratings which are more descriptive than evaluative will tend to exhibit greater reliability.

The behavioral checklist was developed using an iterative process similar to Smith and Kendall's (1963) retranslation approach. The development process began with a series of six job analysis meetings conducted with combined groups incumbent sales clerks and section leaders. No supervisor - subordinate pairs were included in the same meeting. Critical incidents of positive and negative job behaviors were elicited at these meetings using Flanagan's (1954) procedure. One hundred and sixty-five behavioral statements were later derived from the critical incidents related at these meetings.

The behavioral statements were clustered into nine broad performance dimensions. This step was performed by the researcher in order to ensure that the subject matter experts focused on specific behaviors rather than on traits (Campbell, Dunnette, Avery, and Hellervik, 1973). Definitions were also developed for each of the performance dimensions.

The performance dimensions and behavioral statements were then put into a randomly ordered list. This list was presented to a group of nine section leaders, store managers, and personnel employees. These participants were first asked to individually assign each of the behavioral examples to one of the performance rating dimensions. The participants were next asked to evaluate the various behaviors in terms of their importance for effective overall job performance using a technique similar to Thurstone's scaling method of equal appearing intervals. This was done by rating each statement on the basis of a seven-point scale of importance. The mean and standard deviations of the importance ratings for each item were determined. Two criteria were set which the statements had to meet in order to be included in the behavior checklist. First, statements had to be assigned to one of the performance dimensions by at least two thirds of the participants, and second, the standard deviations of the importance ratings for the statements had to be 1.50 or less. Fifty-seven of the statements failed to meet both of these criteria and were eliminated from further consideration.

The remaining set of 108 behavioral statements were then randomly ordered and presented to a third group of ten employees in the form of a retranslation questionnaire. The incumbents in this group were asked to categorize each of the behavioral statements in one of the nine performance dimensions. Statements which were not consistently reassigned to the appropriate dimension by seven or more of these participants were not included in the scale.

The final version of the behavioral checklist consisted of nine performance dimensions and a total of 70 behavioral statements. Each of the nine dimension scales was comprised of either seven or eight statements. The measure was scored by summing the scaled importance values of the statements which were checked within each of the performance dimensions. A total score was determined by summing scores across the nine dimensions. A list of the statements included in each scale of the behavioral checklist along with mean and standard deviations of the importance ratings is included in Appendix B.

A trait-based, graphic rating scale was also developed for evaluating the job performance of the sales clerks in the sample. The rating dimensions were the same as those determined through the behavioral checklist development process. Ratees were assessed on each dimension in terms of a seven point Likert-type scale with descriptive adjectives at each scale point. The anchors used at each scale point were as follows: 1) unacceptable, 2) marginal, 3) acceptable, 4) average 5) good, 6) very good 7) outstanding - among the best ever seen. Brief definitions of each performance dimension were provided with the rating scales. Raters indicated their assessments of subordinates by circling the number on each scale corresponding to their evaluation of each individual's performance on the nine dimensions identified.

Finally, a third scale consisting of a single item was used to assess overall job performance. The question asked raters to evaluate the job performance of each ratee on the basis of an eight-point scale with descriptive anchors at each scale point. The anchors ranged from

1) performance does not meet minimum standards, to 8) the single best performance I have ever observed or could hope to observe. This measure was included for purposes of experimental comparison with the two previously described scales.

A graphic rating scale in the same format as the one described above was developed for use in evaluating the job performance of the section leaders included in the sample. The dimensions included in this scale were determined through a series of meetings with store managers having responsibility for directly supervising the section leaders. It was determined that the dimensions and corresponding definitions identified for the sales clerks were also appropriate for use in appraising the section leaders with one addition. A dimension for supervisory ability was added to the rating scale for section leaders. A copy of this scale along with the instructions for its use is included in Appendix C.

Job Involvement and Organizational Commitment

Job involvement was measured using a subset of the items from the scale developed by Lodhal and Kejner (1965). The original scale consists of 20 items for which respondents indicate their agreement or disagreement on a seven-point Likert-type scale ranging from 1 - strongly disagree, to 7 - strongly agree. Estimates of the split-half reliability of the full scale range from .72 to .89 (Lodahl & Kejner, 1965).

Based on the results of Saleh and Hosek's (1976) factor analysis of items from several different job involvement scales, ten items were

selected from the Lodahl and Kejner scale which had the highest loadings (.30 or higher) on the factor corresponding to the performance - self esteem contingency definition of job involvement. These items and their numbers from the original scale are shown in Appendix D. A total score for each respondent was determined by summing response values across these ten items. The reliability of this scale as determined by coefficient alpha and based on the section leader sample was .84.

A scale developed by Porter, Steers, Mowday, and Boulian (1974) was used to measure organizational commitment. The scale consists of fifteen item statements to which respondents indicated their degree of agreement or disagreement also on a seven-point Likert-type scale ranging from 1 - strongly disagree, to 7 - strongly agree. The scaling direction on seven of the items is reversed to reduce response set bias. The scale items were modified slightly to make them applicable to the organization where the study was conducted. A copy of the scale used with the name of the organization deleted is included in Appendix E.

A measure of overall organizational commitment was derived for each respondent by summing response values across all of the items after reversing the values of the six negatively phrased statements. The reliability of this scale as determined by coefficient alpha is reported to range from .82 to .92 (Porter, et al., 1974). The scale reliability based on the section leader sample data was .91.

The measures of rater job involvement and organizational commitment were the independent variables for this study. A composite

variable comprised of the sum of the averaged item scores from both scales was determined for each rater. Forming the composite in this way resulted in the two attitude measures being weighted equally despite the different number of items in each scale. This was done because of the absence of a theoretical or empirical basis to determine situational importance weightings. The justification for summing the two measures was based upon the proposed model which suggests that job involvement and organizational commitment act concomitantly to influence behavior in an additive and compensatory fashion.

An effort was made to perform multiple tests of the hypotheses examined in this study. In some instances this involved using both the ratees and the raters as experimental units of analysis. In some of the analyses performed, the procedures required that the rater and ratee samples be dichotomized. This division was made on the basis of whether each rater's score on either the job involvement, organizational commitment, or combined job involvement - organizational commitment measures was above or below the corresponding median value for the sample. The use of this approach, which is a common practice in the investigation of similar cognitive attributes (Schneier, 1977), resulted in approximately equal numbers of raters and ratees in the high and low groups. When raters were used as the experimental unit, their individual scores on the job involvement, organizational commitment, or composite measures served as the independent variable.

Rating Errors

Measures of the constant rating errors considered constituted the principal dependent variables in this study. Saal, Downey, and Lahey (1980) have provided compelling evidence to demonstrate that the way a measure of the various rating errors is operationalized may have an effect upon the results of the statistical analyses using that measure as a dependent variable. In view of this, multiple operational procedures for defining the constant rating errors were used when appropriate and feasible.

Leniency was first assessed by comparing mean dimensional ratings across ratees with scale midpoints (Bernardin, et al., 1976). The more the mean ratings exceeded the midpoint, the more the ratings were considered to be characterized by positive leniency. The opposite was true for negative leniency (severity). A second measure assessed leniency in terms of the skewness of the frequency distribution of ratings within each dimension (Landy, Farr, Saal, & Freytag, 1976). Negative skewness values were indicative of the degree of positive leniency in the ratings. Positive skewness indicated severity.

Central tendency was measured by determining the mean of the distances between assigned ratings and scale midpoints for each dimension. The lower this mean value, the more central tendency was considered to be present in the ratings (DeCotiis, 1977).

Restriction of range was first operationalized by determining the standard deviation of the performance ratings assigned to ratees on each scale dimension. This approach is based on the principle that a comparatively large standard deviation for ratings on a particular

dimension is an indication that the range of the ratings assigned is also large (Borman and Dunnette, 1975). A second approach examined the degree of kurtosis or peakedness characterizing the distributions of the ratings for each dimension (Landy, et al., 1976). The higher the kurtosis value for a distribution, the more the ratings were considered to exhibit restriction of range.

Halo was assessed in two different ways. First, halo was estimated by determining the standard deviation of each ratee's ratings across dimensions. The mean of these standard deviations across ratees was interpreted as inversely reflecting the degree of halo in the ratings, e.g., a high mean standard deviation indicated a low degree of halo and vice versa (Bernardin & Walter, 1977). This definition is rationalized on the basis of the belief that raters who do not sufficiently differentiate between aspects of performance in rating others will tend to show relatively little variance in they ratings they assign across performance dimensions (Borman, 1975; Ivancevich, 1979).

The second means of assessing the extent of halo in the ratings was based on an analysis of the intercorrelations of ratings on the various dimensions from the BCL and GRS (Keaveny & McGann, 1975). With this approach, the size of the correlations is considered to be inversely related to the extent of discrimination among different aspects of behavior and thus, directly related to the degree of halo present in the ratings.

Procedure

This study was conducted as part of a project to develop and validate a test battery and performance appraisal program for use in selecting and evaluating sales clerks and section leaders in the sponsoring organization. The project began with a series of meetings with incumbent sales clerks, section leaders, and store managers to analyze the sales clerk job and develop the performance appraisal measures described previously. Based on the information obtained through these meetings, a battery of measures was formulated which included both published and original aptitude tests, the job involvement and organizational commitment scales, and a biographical information inventory. The entire battery was administered to all sales clerks and section leaders in the sample. Testing was conducted during working hours in conference rooms made available at the three store locations. The battery was administered to groups of employees varying in size from three to twelve. Sales clerks and section leaders were tested in mixed groups however no supervisor - subordinate pairs were tested in the same session.

The first set of performance ratings of the sales clerks in the sample was collected approximately two months after the testing portion of the project was completed. Section leaders completed all three performance measures for each of the sales clerks under their direct supervision who had completed the battery and had worked in their department for at least three months. Ten sales clerks were eliminated from the sample because of insufficient tenure with the organization or their current supervisor. Five other employees who had

completed the battery were no longer employed by the sponsoring organization when the ratings were collected. The resulting sample consisted of 199 sales clerks who had completed the battery and for whom complete sets of performance ratings were available.

The performance ratings were made by the section leaders during group meetings. One meeting was held at each of the three stores. The meetings were scheduled at times when all of the section leaders at each location were assured of attending. The participants were informed at the beginning of the meetings that the subordinate evaluations they would be completing would be used for research purposes only and would not be made available to the sales clerks being rated or to company management.

The meetings began with a rater training program which lasted approximately two hours. The program consisted of a lecture with handouts and a group discussion covering the purpose and process of rating subordinate job performance. The information presented focused attention on the multidimensionality of work performance and the importance of basing evaluations on observable behaviors. Examples of effective and ineffective behavior in each of the performance dimensions identified during the scale development phase were reviewed. A major portion of the allotted time was spent instructing the raters about the existence and nature of constant errors in ratings. Strategies for the avoidance of these errors were discussed extensively. The concepts of reliability, accuracy, and validity as they relate to performance ratings were also addressed. Finally, the use of the various rating formats to be completed was covered in considerable

detail. The raters were given a chance to practice completing a set of rating forms by appraising the performance of an employee with whom they were familiar but did not presently supervise. The section leaders in these sessions were encouraged to ask questions and this often lead to further group discussion aimed at clarifying the appraisal process and reducing the prevalence of rating errors.

Throughout the training programs, the importance of this study to the improvement of individual performance and the continued profitability of the organization was strongly emphasized. Section leaders were told that the success of the entire project rested largely on their ability to make reliable, valid, and error-free evaluations of their subordinates' job performance. A large number of studies have demonstrated that training programs of the same basic type and duration as the one presented in this study have a substantial impact on the psychometric properties of subsequently assigned performance ratings (e.g., Bernardin, 1978; Bernardin & Walter, 1977; Latham, Wexley, & Pursell, 1975). A more extensive program might have produced stronger and longer lasting experimental effects. However, in view of research investigating the effect of very brief training programs on the reduction of rating errors (Bernardin, 1978; Borman, 1975) it appears likely that the training procedures used in this study resulted in a reduction of certain rating biases especially in the short term. Furthermore, Fay and Latham (1980) have concluded that the characteristics of various rating scale formats are more salient in terms of their impact on the prevalence of rating errors once the raters using them have been properly trained in their use.

Following the completion of the training program, section leaders were given rating booklets containing the rating forms to complete for each of their subordinates. The order in which the forms were presented was counter-balanced to minimize the possible effect of response sets. All rating forms were completed for one sales clerk before moving on to the next one.

One week after the appraisal process for the sales clerks was completed, performance evaluations of the section leaders in the sample were obtained. The section leaders were evaluated by either the manager or assistant manager of the store where they worked. These evaluations were made using the graphic rating scale designed for this purpose. The managers were given an abbreviated version of the rater training program prior to making their appraisals.

Approximately ten months after the first set of performance ratings for the sales clerks was gathered, a second set was obtained for those individuals still employed at the sponsoring organization. The rater training program was repeated to ensure its effect and because of the presence of several new section leaders. The same forms and procedures used during the first round of ratings were used in the follow-up sessions. A total of 113 of the sales clerks from the original sample were rated a second time. Fifty-five sales clerks were rated by different section leaders during the second set of appraisals.

CHAPTER V

RESULTS

Introduction

The results of the statistical analyses performed to test the hypotheses considered in this investigation are presented in this chapter. Each hypothesis will be restated prior to reporting the results of the procedures used in its examination.

Summary statistics for the three independent variables utilized in this study - section leader job involvement, organizational commitment, and the composite variable formed by summing the mean item scores for these two factors - are presented in Table 3. The mean, median, standard deviation, and range are shown for the total score and mean item score of each scale. In testing those hypotheses which required the section leader and sales clerk samples to be dichotomized, the median value for the total score of each scale was used to divide the sample.

Table 4 presents the summary statistics for the ratings given to sales clerks on each dimension of the graphic rating scale and the behavioral checklist. Scores on the dimensions of the graphic rating scale are based on a 1 to 7 point scale. Dimension scores on the behavioral checklist represent the sum of the weights of the positive items which were checked and the weights of the negative statements which were not checked by the section leaders in evaluating each employee. The total scores on each scale shown in Table 4 are based on the mean score across all nine dimensions. Dimensional scores on

Table 3
Summary Statistics for Section Leader Attitude Scale Scores

Variable	Mean	Median	Standard Deviation	Range
Job Involvement (Total Score)	46.26	45.25	7.24	32.00 to 63.00
Job Involvement (Mean Item Score)	4.67	4.53	0.72	3.20 to 6.30
Organizational Commitment (Total Score)	85.97	85.00	11.92	58.00 to 105.00
Organizational Commitment (Mean Item Score)	5.73	5.67	0.80	3.87 to 7.00
Composite (Sum of JI and OC Total Scores)	132.24	130.50	15.02	101.00 to 159.00
Composite (Sum of JI and OC Mean Item Scores)	10.34	10.12	1.17	7.93 to 12.50

n = 38

Table 4

Summary Statistics for Rating Dimension Scores on the
Graphic Rating Scale and the Behavioral Checklist for Sales Clerks

Dimension	<u>Graphic Rating Scale</u>			<u>Behavioral Checklist</u>		
	Mean	SD	Range	Mean	SD	Range
Dependability	5.02	1.26	1 to 7	4.08	1.40	0.00 to 5.50
Knowledge of Job Procedures	5.01	1.05	1 to 7	4.67	1.20	0.00 to 5.69
Sales Ability	5.00	1.18	1 to 7	3.55	1.22	0.66 to 5.55
Customer Service	5.25	1.16	1 to 7	3.70	1.15	0.61 to 5.23
Knowledge of Company Policies	4.92	1.18	1 to 7	3.67	1.27	0.00 to 5.17
Product Knowledge	4.76	1.16	2 to 7	3.94	1.79	0.00 to 5.87
Initiative	4.66	1.29	1 to 7	2.92	1.59	0.00 to 5.29
Judgement	4.49	1.06	2 to 7	3.64	1.38	0.55 to 5.72
Employee Relations	4.37	1.20	1 to 7	4.09	1.38	0.00 to 5.10
Total (Mean of 9 dimension scores)	4.92	0.92	3.2 to 7.0	3.81	0.95	1.52 to 5.30

n = 199

the graphic rating scale and the behavioral checklist are not directly comparable because of the differences in the potential ranges of the two measures. Results of the evaluations of the section leaders' job performance provided by the store managers and assistant managers on the graphic rating scale for supervisors are summarized in Table 5.

Reliability of Performance Measures

Four different procedures were used to estimate the reliability of the sales clerk performance assessment measures employed in this study each of which takes into account a different source of error variance. First, a modified alternate forms approach was used to calculate reliability coefficients by correlating corresponding dimension and total scores from the graphic rating scale and behavioral checklist. This procedure is analogous to selecting different test questions (rating scale items) from a pool to measure each performance dimension and is subject to the effects of content sampling error. The correlations for each dimension and the overall summed scores on each rating format are shown in the first column of Table 6. While a considerable degree of variation is evident in the inter-format correlations for the various dimensions, the reliability of the overall measures as determined with this first procedure was considered to be acceptable for the types of rating instruments considered.

Second, the internal consistency of each of the dimensional scales of the behavioral checklist was calculated using the Kuder - Richardson 20 formula which for dichotomous data is equivalent to Cronbach's

Table 5
Summary Statistics for Rating Dimension Scores on the
Graphic Rating Scale for Section Leaders

Dimension	<u>Graphic Rating Scale</u>		
	Mean	SD	Range
Dependability	6.16	.92	5 to 7
Knowledge of Job Procedures	5.94	.97	4 to 7
Sales Ability	6.22	1.18	4 to 7
Customer Service	5.96	1.24	3 to 7
Knowledge of Company Policies	6.12	.93	5 to 7
Product Knowledge	6.04	1.08	4 to 7
Initiative	5.71	1.14	3 to 7
Judgement	5.67	1.07	3 to 7
Employee Relations	5.40	1.53	2 to 7
Supervisory Ability	4.75	2.18	1 to 7
Total (Mean of 10 dimension scores)	5.78	1.03	5.02 to 7.00

n = 38

Table 6

Reliability Coefficients for the Sales Clerk Performance Ratings

Scale	<u>Alternate Formats</u>	<u>KR-20</u>	<u>Test - Retest^a</u>		<u>Interrater</u>	
	GRS - BCL (<u>n</u> = 199)	BCL (<u>n</u> = 199)	GRS (<u>n</u> = 58)	BCL (<u>n</u> = 58)	GRS (<u>n</u> = 55)	BCL (<u>n</u> = 55)
Dependability	.57	.72	.77	.73	.64	.56
Job Procedures	.41	.69	.71	.65	.61	.58
Sales Ability	.41	.65	.73	.56	.66	.51
Customer Service	.39	.66	.68	.61	.59	.47
Company Policy	.32	.69	.63	.55	.54	.45
Product Knowledge	.59	.67	.69	.61	.60	.58
Initiative	.60	.75	.77	.70	.65	.56
Judgment	.62	.74	.71	.70	.63	.62
Employee Relations	.57	.62	.73	.66	.65	.57
Total	.68	--	.72	.64	.62	.55

^aThe interval between ratings correlated to determine the test - retest and interrater reliability coefficients was ten months.

coefficient alpha. Internal consistency estimates were not determined for the total scores on the behavioral checklist or the graphic rating scale as these multidimensional composites were not intended to be homogeneous. The reliability coefficients determined with this procedure are shown in second column of Table 6. The coefficients shown were also judged to be of an acceptable magnitude especially considering that the items comprising each dimension checklist were discrete behavioral descriptions which were intuitively grouped into rating categories by subject matter experts.

Third, a test-retest procedure was employed to estimate coefficients of stability by correlating ratings from the first evaluation session with those from the second round of ratings collected ten months later. These reliability estimates are subject to error variance due to changes in performance over time as well as differences in the conditions under which each set of evaluations was made. A second set of ratings was available for 113 of the sales clerks in the sample. However, only 58 of the clerks were rated by the same section leader on both occasions due primarily to changes in departmental assignments. Test - retest reliability coefficients calculated using the ratings for these employees are listed in the third and fourth columns of Table 6. These coefficients are notably higher in magnitude than those determined with the previous two approaches particularly for the graphic rating scale. This most likely indicates that either job performance in the sales clerk position tends to be relatively stable or that during the second round of evaluations, section leaders recalled how they rated their subordinates the first time.

Finally, for the 55 sales clerks who were rated by different section leaders on the two occasions, interrater reliability coefficients were determined by correlating the two sets of ratings. These coefficients are shown in the fifth and sixth columns of Table 6. As expected, these values are considerably lower than the test - retest correlations. This is most likely due not only to error variance contributed by the difference in raters and performance changes over the ten month interval but also the changes in work assignments of the ratees and/or the raters.

Based on the coefficients shown in Table 6 and a consideration of the sources of error variance affecting each of the procedures described above, it was concluded that both the graphic rating scale and behavioral checklist rating formats evidenced sufficient reliability to support their use in this study. In a related vein, a generalizability study was conducted as part of the procedure for testing hypothesis 5. This analysis simultaneously considered several of the components of variation addressed by the preceding methods for assessing reliability. The results of this analysis will be discussed when the findings from the examination of hypothesis 5 are presented.

Research Design, Test Procedures, and Power

Saai et al. (1980) have proposed that a completely crossed research design in which all ratees are evaluated by all raters on multiple performance dimensions or a partial design in which blocks of raters rate some but not all ratees are the only means by which unambiguous estimates of rating errors can be made. According to this

view, all other designed are flawed as they can not fully differentiate ratee and rater effects.

Unfortunately, fully crossed and even partially crossed are very difficult to achieve outside of the laboratory and indeed are quite rare in field studies of the performance rating process. In most cases involving supervisory ratings, an employee's immediate supervisor is the only one sufficiently familiar with the individual's performance to provide a meaningful evaluation. This was true for the current study which utilized a completely nested design in which each rater evaluated a unique group of ratees. The limitations inherent in this design preclude a definitive assessment of the prevalence of each type of rating error present in the data by any one method of measurement. However, this problem may be overcome to a certain degree by employing more than one means of operationalizing the rating errors in question and utilizing both raters and ratees as the experimental units of analysis. In accord with this, the hypotheses in this study concerning rating errors were tested with multiple procedures whenever feasible and conclusions were reached by examining the results for trends.

An additional consideration which should be stressed is that the thrust of this study was not to assess the absolute degree of the various rating errors present in the data but rather to contrast the measured amounts of these phenomena in the evaluations assigned by raters differing in their attitudinal characteristics. Hence, while accuracy of the estimates of systematic error variance may be limited by the operational definitions of the effects in question, there is no

reason to believe that an assessment of the rating errors evidenced by any one group of raters will be more biased than that of any other rater group.

The model proposed in this study from which the hypotheses tested are derived predicts the effect of a combined measure of rater job involvement and organizational commitment on the psychometric characteristics of performance ratings. Accordingly, the effect of this variable on the operationalized measures of the various rating errors considered is of primary interest in reporting the study results. However, in an effort to better understand the mechanism by which this composite attitude may influence the assignment of performance ratings, the separate effects of job involvement and organizational commitment in this regard are reported for the principal hypotheses pertaining to each of the rating errors studied.

Finally, all of the hypotheses tested are stated in such a way as to predict the directionality of the results. In general, it is posited that raters evidencing higher levels of job involvement and organizational commitment will assign ratings which are psychometrically superior to those given by raters with lower levels of these particular attitudes. However, in view of the lack of previous reported research addressing this general area of study, a decision was made to employ two-tailed statistical procedures in testing the hypotheses. This approach permits the detection of significant group differences without regard to the directionality of the results although stronger effects are required to reach a predetermined level of statistical significance. Except when specified otherwise, an

alpha level of .05 was used as the basis for establishing statistical significance with all of the testing procedures reported in this chapter. Estimates of the power of the statistical procedures used in this analysis were made on the basis of a .05 significance criterion (two-tailed). In view of the limited amount of relevant published research directly relating to this study, Cohen's (1977) suggested values for medium effect sizes were used in estimating power.

Leniency

Hypothesis 1A. A combined measure of rater job involvement and organizational commitment will be inversely related to the amount of leniency present in the assigned ratings of subordinate job performance.

Hypothesis 1A was tested using two different approaches. In both cases, sales clerks served as the units of analysis. Both procedures also involved dichotomizing the sales clerk sample on the basis of whether the section leaders evaluating each clerk scored above or below the median values on the attitude measures considered. The first procedure was based upon the operational definition of leniency as the difference between the ratings assigned and the midpoint of the rating scales. The degree to which the ratings fell above or below the midpoints was considered an indication of their leniency or severity. The difference between each of the ratings assigned to each

sales clerk on the nine dimensions scale and the midpoints of those scales on both the graphic rating scale and the behavioral checklist was determined. The midpoint value for each dimension scale on the graphic rating scale was 4. The midpoints of the dimensions scales on the behavioral checklist varied because of the different number of items and item weightings for each dimension. The scale midpoints for the behavioral checklist are shown in Appendix B. Mean leniency measures were determined for each of the nine rating dimensions on the graphic rating scale and the behavioral checklist. The means for the groups of sales clerks to be contrasted on the basis of their section leaders' attitude scale scores are shown in Table 7.

The test of hypothesis 1A based on the mean difference definition of leniency was performed using a two-way analysis of variance procedure. The first factor was group membership based on whether each sales clerk's section leader scored above or below the median on the attitude scale in question (the job involvement/organizational commitment composite, job involvement, or organizational commitment). The second factor was each of the nine dimensions of the two rating formats. The leniency measure was the dependent variable. Separate analyses were performed for the graphic rating scale and behavioral checklist to simplify interpretation of the results. Summary tables for these ANOVA procedures are presented in Tables 8, 9, and 10. Power estimates for the significance tests based on the expectation of medium experimental effects were .94 for the group effect and .99 for the dimension and the interaction effects.

Table 7

Dimension and Overall Means of the Leniency Measure for the Sales Clerk Sample Dichotomized on the Basis of the Median Values of Section Leader Attitude Scale Scores

Dimension	Graphic Rating Scale					
	<u>Composite Variable</u>		<u>Job Involvement</u>		<u>Organizational Commitment</u>	
	Above Median (<u>n</u> =98)	Below Median (<u>n</u> =101)	Above Median (<u>n</u> =97)	Below Median (<u>n</u> =102)	Above Median (<u>n</u> =98)	Below Median (<u>n</u> =101)
Dependability	1.18	.87	1.01	1.03	1.10	.94
Job Procedures	1.05	.98	1.12	.91	.99	1.04
Sales Ability	1.02	.97	1.04	.95	.99	1.00
Customer Service	1.40	1.11	1.38	1.13	1.37	1.14
Job Knowledge	.96	.89	1.10	.75	1.03	.82
Product Knowledge	.77	.76	.85	.69	.69	.84
Initiative	.70	.63	.87	.52	.71	.62
Judgment	.52	.48	.62	.38	.53	.47
Empl. Relations	1.40	1.03	1.20	1.23	1.28	1.15
Overall Score	1.00	.86	1.01	.84	.97	.89

Dimension	Behavioral Checklist					
	<u>Composite Variable</u>		<u>Job Involvement</u>		<u>Organizational Commitment</u>	
	Above Median (<u>n</u> =98)	Below Median (<u>n</u> =101)	Above Median (<u>n</u> =97)	Below Median (<u>n</u> =102)	Above Median (<u>n</u> =98)	Below Median (<u>n</u> =101)
Dependability	11.60	9.68	10.08	11.12	12.56	8.72
Job Procedures	15.12	14.16	14.05	15.09	14.85	14.37
Sales Ability	6.24	6.24	6.36	6.12	6.56	5.92
Customer Service	8.72	8.72	7.80	9.56	8.72	8.80
Job Knowledge	8.12	6.86	7.38	5.96	7.35	7.07
Product Knowledge	8.96	7.20	5.64	6.72	8.48	7.60
Initiative	3.12	1.44	2.44	2.20	3.04	1.44
Judgment	6.96	5.92	5.84	6.72	6.96	5.68
Empl. Relations	11.90	9.94	10.60	11.09	12.32	9.52
Overall Score	10.44	9.00	10.44	9.36	10.53	9.00

Note: Leniency is assessed by calculating the difference between the scale midpoint and the ratings assigned.

Table 8

Analysis of Variance with Section Leader Composite Score Group and Rating Scale Dimensions as the Independent Variables and the Leniency Measure as the Dependent Variable

<u>Graphic Rating Scale</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	8.990	6.32*	.004
Dimension	8	12.082	8.50**	.036
Group x Dimension	8	.952	.67	
Residual	1764	1.422		
Total	1781	1.472		

<u>Behavioral Checklist</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	11.263	5.85*	.003
Dimension	8	44.798	23.27**	.096
Group x Dimension	8	.458	.24	
Residual	1764	1.926		
Total	1781	2.117		

Note: Leniency is assessed by subtracting the scale midpoint from the rating assigned for each dimension.

n = 199

* p < .05 ** p < .01

Table 9

Analysis of Variance with Section Leader Job Involvement Score
Group and Rating Scale Dimensions as the Independent Variables
and the Leniency Measure as the Dependent Variable

<u>Graphic Rating Scale</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	13.038	9.19**	.005
Dimension	8	12.082	8.51**	.036
Group x Dimension	8	.891	.55	
Residual	1764	1.420		
Total	1781	1.472		

<u>Behavioral Checklist</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	10.561	5.48*	.001
Dimension	8	44.798	23.24**	.096
Group x Dimension	8	1.211	.63	
Residual	1764	1.923		
Total	1781	2.117		

Note: Leniency is assessed by subtracting the scale midpoint from the rating assigned for each dimension.

n = 199

*p < .05

**p < .01

Table 10

Analysis of Variance with Section Leader Organizational
Commitment Score Group and Rating Scale Dimensions as the
Independent Variables and the Leniency Measure as the
Dependent Variable

<u>Graphic Rating Scale</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	1.735	.19	.001
Dimension	8	12.082	8.47**	.036
Group x Dimension	8	.790	.55	
Residual	1764	1.426		
Total	1781	1.472		
<u>Behavioral Checklist</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	12.861	6.70**	.004
Dimension	8	44.804	23.24**	.096
Group x Dimension	8	1.373	.72	
Residual	1764	1.920		
Total	1781	2.117		

Note: Leniency is assessed by subtracting the scale midpoint from the rating assigned for each dimension.

n = 199

* $p < .05$

** $p < .01$

The result of primary interest in these tables for testing hypothesis 1A is the F test for the Group main effect. In order for the hypothesis to be supported, there must a significant difference in the means between the two comparison groups across rating dimensions. In addition, the hypothesis posits that the means for the sales clerk group rated by the section leaders scoring above the median on the attitude measures should be lower than those of the other group, thus indicating a lower level of leniency error. The results for the composite variable shown in Table 8 indicate that there was a significant difference in ratings across dimensions between the dichotomized sales clerk groups on both the graphic rating scale and the behavioral checklist. However, examination of the means shown in Table 7 indicates that the ratings assigned by section leaders scoring above the median on the composite attitude scale were more lenient than those assigned by the section leaders with attitude scores below the median. This finding is directly opposed to the prediction made by hypothesis 1A. Similar results are evident in Table 9 which presents the summary tables for the same procedure using section leader job involvement scores as the dichotomizing factor. Finally, Table 10 indicates that when section leader scores on organizational commitment were used to split the sales clerk sample, the same result was found for the ratings made on the behavioral checklist. The ratings on the graphic rating scale did not differ significantly in terms of leniency in this instance.

As expected, all of the results for the Dimension effect in the ANOVAs performed were significant which implies that the ratings

assigned to individuals tended to differ considerably from one dimension to the next. Also, none of the interaction effects were significant. The η^2 values for the Group effect were quite low ranging from .001 to .005. This in itself is not surprising as while it was predicted that supervisor attitudes would have a significant effect on the degree of leniency in the ratings assigned, it is acknowledged that there are many other factors which will account for a much larger percentage of the variance in these ratings not the least of which is the job performance of the rates.

The second approach used to test hypothesis 1A employed a conceptualization of leniency based on the degree of skewness evident in the ratings distribution. The model predicts that the ratings assigned by section leaders evidencing high levels of the target attitudes will be less negatively skewed (i.e., less lenient) than those of section leaders indicating lower levels of these attitudes. This test of the hypothesis was again accomplished by using the sales clerks as the units of analysis and dividing the sample on the basis of the attitude scale scores of the section leaders ratings each clerk. The skewness of the ratings assigned to each group of sales clerks on each of the nine dimensions on the two rating formats were first determined. A t test procedure was then applied to determine whether the mean of the dimension skewness values differed for the contrasted sales clerk groups. The estimated power of these tests based on the expectation of a medium effect size was .16. This low power value resulted from the nature of the design in which the skewness values of the nine dimension were contrasted. However, each

value was determined on the basis of a distribution of approximately 100 scores. The results of this analysis are shown in Table 11.

Significant differences on both rating scale formats were found between the sales clerk groups split on the basis of the section leaders' composite attitude scores. Similar results were found on the behavioral checklist when section leader job involvement scores were used to divide the sales clerks and on the graphic rating scale when scores on the organizational commitment scale were used. Once again, all of the significant differences were in the opposite direction of that which was predicted by hypothesis 1A. The section leaders who had the higher scores on the attitude measures tended to assign ratings which were more negatively skewed and hence, more lenient.

Consideration was given to whether all of the assumptions required for the use of the parametric t test were met when applied to testing the difference between two sets of skewness values. Siegel (1957) suggests that the randomization test for two independent samples be applied in such cases. However, Pitman (1937, in Siegel, 1957) has shown that if the kurtosis of the combined samples is small and if the larger sample is not more than five times larger than the smaller sample, the randomization distribution of possible outcomes is closely approximated by the t distribution. The kurtosis of the combined samples for each comparison ranged from $-.202$ to $-.344$ and the samples sizes for each group were equal. It may therefore be concluded that the t test comparisons on means of the skewness values for dimension distributions are a valid means of testing hypothesis 1A.

Table 11

Comparison of Mean Dimension Skewness (Leniency) in Ratings
Assigned by Section Leaders Scoring Above and Below the Median on
the Three Attitude Measures

Rating Format	Group	<u>n</u>	Mean Kurtosis ^a	Standard Deviation	df	<u>t</u>
<u>Composite Variable (JI + OC)</u>						
Graphic Rating Scale	Above Median	98	-0.602	0.373	16	-2.27*
	Below Median	101	0.145	0.475		
Behavioral Checklist	Above Median	98	-0.477	0.382	16	-2.22*
	Below Median	101	0.004	0.527		
<u>Job Involvement</u>						
Graphic Rating Scale	Above Median	98	-0.662	0.322	16	-2.64*
	Below Median	101	0.162	0.309		
Behavioral Checklist	Above Median	98	-0.488	0.305	16	-1.57
	Below Median	101	-0.237	0.276		
<u>Organizational Commitment</u>						
Graphic Rating Scale	Above Median	97	-0.400	0.364	16	-1.83
	Below Median	102	-0.012	0.257		
Behavioral Checklist	Above Median	97	-0.619	0.321	16	-2.52*
	Below Median	102	-0.109	0.221		

^aMean skewness of the nine dimension score distributions.

* $p < .05$

In summary, hypothesis 1A as derived from the model was not supported. A significant effect in the opposite direction predicted was found for the composite attitude measure on both the behavioral checklist and the graphic rating scale; for job involvement on the graphic rating scale and possibly the behavioral checklist; and for organizational commitment on the behavioral checklist.

Hypothesis 1B. The relationship of the combined measure of rater job involvement and organizational commitment with the amount of leniency in assigned ratings of subordinate job performance will be stronger for ratings made on a behaviorally-based rating format than on a trait-based rating format.

Hypothesis 1B was tested using the same two of operational definitions of leniency that were employed in testing hypothesis 1A. The first procedure was based on assessing leniency in terms of the level of the ratings assigned (i.e., rating - scale midpoint). Pearson product-moment correlation coefficients indicating the extent of the relationship between the composite, job involvement, and organizational commitment measures for the section leaders assigning ratings and the mean of the leniency scores across dimensions on both rating scale formats were calculated. The three pairs of correlations for the two rating formats were then contrasted by means of the Fisher's z transformation of r (Hays, 1973, 662-667).

The same analysis was performed using both the sales clerks and the section leaders as the units of analysis. This was done because

all section leaders did not evaluate an equal number of sales clerks. Hence, the procedure using sales clerks as the unit of analysis could potentially be biased if the relationship between supervisor attitude scores and leniency in assigned ratings was stronger for some section leaders than others who happened to have rated a larger number of individuals. On the other hand, the analysis using section leaders as the unit of analysis, while considering the effect of this relationship for each rater equally, has the drawback of a limited sample size. The results of both analyses, as well as those from the procedures which used a second operational definition of leniency, were taken into consideration in an effort to compensate for limitations in the study's experimental design.

Table 12 presents the results of the analysis using sales clerks as the unit of analysis. In this procedure, leniency was measured by determining the difference between the ratings and the scale midpoints for each dimension of the two rating formats and calculating the mean of these differences across dimensions for each ratee. These means were then correlated with the attitude scale scores of the section leaders evaluating each sales clerk. The pattern of statistical significance for the correlation coefficients is consistent with the previously reported results for hypothesis 1A. None of the z difference scores were significant thus indicating that the strength of the relationship between section leader attitudes and the degree of leniency in the ratings assigned did not differ significantly for the graphic rating scale and the behavioral checklist. The estimated power of the tests for significance of the product-

Table 12

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Leniency Measures Across Performance Dimensions on the Graphic Rating Scale and the Behavioral Checklist With Sales Clerks as the Unit of Analysis

Attitude Measure	Graphic Rating Scale	Behavioral Checklist	<u>z</u> diff
	<u>r</u>	<u>r</u>	
Composite Score (JI + OC)	.164**	.132*	.317
Job Involvement	.199**	.087	1.129
Organizational Commitment	.067	.119*	-.515

Note: Leniency is measured by determining the difference between the ratings and the scale midpoints for each dimension and calculating the mean of these differences for each ratee across dimensions on each rating format.

n = 199

* p < .05

** p < .01

moment correlations was .99 and for the differences between pairs of correlations, .85.

The results for the same analysis using section leaders as the unit of analysis are presented in Table 13. In this case, leniency was measured by determining the mean difference between the ratings and the scale midpoints for each dimension of the two rating formats and calculating the mean of these differences across dimensions and across the sales clerks rated by each section leader. These values were then correlated with the attitude scale scores for each section leader and the significance of the differences between the coefficients for the behavioral checklist and the graphic rating scale were tested. Only one of the product-moment correlations shown in Table 13 was statistically significant (job involvement and leniency on the graphic rating scale). As was the case in the previous analysis, none of the z difference scores were significant. The estimated power of the significance tests for the product-moment correlations was .46 and .24 for the differences between pairs of correlations. The relatively low power for the analyses using the section leaders as the units of analysis underscores the importance of considering these results as indicators of a general pattern or trend and not as a sole source upon which to base conclusions.

The skewness approach for operationalizing the measurement of leniency was also used to test hypothesis 1B. For this procedure, section leaders were divided into two groups on the basis of whether they scored above or below the median on the composite attitude measure. The skewness of the ratings assigned to sales clerks on each

Table 13

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of the Leniency Measures Across Performance Dimensions on the Graphic Rating Scale and the Behavioral Checklist with Section Leaders as the Unit of Analysis

Attitude Measure	Graphic Rating Scale	Behavioral Checklist	<u>z</u> diff
	<u>r</u>	<u>r</u>	
Composite Score (JI + OC)	.177	.248	-.305
Job Involvement	.316*	.067	1.092
Organizational Commitment	-.023	.234	-1.092

Note: Leniency is measured by determining the mean difference between the ratings and the scale midpoints for each dimension and calculating the mean of these means across dimensions and across ratees for each section leader on each rating format.

n = 38

* p < .05

dimension of the two rating formats was determined for the individuals rated by section leaders in the two groups. These values were used as the data for a two-way analysis of variance procedure with section leader score group (above or below the median) as one factor and rating format (graphic rating scale or behavioral checklist) as the other factor. The test of the hypothesis involved determining whether the Group x Format interaction effect was significant and then examining cell means to identify directionality.

A summary of the results of this ANOVA procedure is shown in Table 14. The results indicate that the main effect for section leader group was significant ($p \leq .01$) again indicating that section leaders who scored higher on the composite attitude measure tended to assign more lenient ratings. Also, the effect for rating format was significant ($p \leq .01$) and an examination of the cell means indicates that the ratings on the behavioral checklist tended to be more negatively skewed (i.e., more lenient) than those given on the graphic rating scale. However, the interaction term was not significant and hence, hypothesis 1B was again not supported. The small number of observations in each cell of the 2 x 2 design for this analysis resulted in low power estimates (.17) for each of the experimental effects. This level of power and the proportions of total variance accounted for provide a further indication of the strength of the main effects examined in this analysis.

In summary, the results of the analyses performed to test hypothesis 1B did not support the belief that the relationship between section leader attitudes and the degree of leniency present in

Table 14

Analysis of Variance with Section Leader Attitude Score Groups
and Rating Format as the Independent Variables and Dimension
Skewness as the Dependent Variable

<u>Composite Measure (JI + OC)</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	0.704	7.382**	.152
Format	1	0.988	10.369**	.211
Group x Format	1	0.002	0.020	
Residual	32	0.095		
Total	35	0.136		
<u>Job Involvement</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	0.745	7.480**	.204
Format	1	0.974	9.711**	.166
Group x Format	1	0.001	0.011	
Residual	32	0.101		
Total	35	0.121		
<u>Organizational Commitment</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	0.532	6.377**	.121
Format	1	1.022	11.973**	.301
Group x Format	1	0.003	0.037	
Residual	32	0.087		
Total	35	0.159		

n = 38

**_p < .01

performance ratings assigned would be stronger on the behavioral checklist than on the graphic rating scale format. Indeed, the data indicate that while the opposite relationship proposed by hypothesis 1A is quite apparent, it is not any more prevalent on either of the two rating formats considered.

Hypothesis 1C. Rater ability will moderate the relationship between the composite measure of rater job involvement and organizational commitment and the amount of leniency in assigned ratings of subordinate job performance.

Hypothesis 1C proposes that the relationship between levels of the composite measure and the extent of leniency in performance ratings assigned will be stronger for raters with higher ability levels than for those individuals who are lower in ability. Tests of hypothesis 1A demonstrated the presence of a direct relationship between the measures of section leader work-related attitudes and the prevalence of leniency in their ratings of subordinate job performance. Although the nature of the association of these factors was diametrically opposed to what was predicted by the model, the reasoning behind the expectation that rater ability will moderate this relationship remains unchanged. Essentially, the model posits that the likelihood that a person will translate behavioral intentions into overt behavior is subject to the individual's ability to perform the behavior in question. Thus, even though it appears that the cognitive

factors under consideration lead to unanticipated behaviors, ability should still moderate the relationship between behavioral intentions and behavior.

A moderated regression approach was utilized to test hypothesis 1C. With this strategy, the presence of a moderating effect is determined by examining the statistical significance of the difference in the values of the squared multiple correlations (R^2) for two regression models: 1) a model containing only main effects terms and 2) a model which includes both main effects and their interaction terms. Stated differently, in a multiple regression formulation involving a dependent variable (y), and independent variable (x), and a potential moderator variable (z), z is said to moderate the relationship between x and y if the regression of y on x , z , and xz (a cross product or interaction term) shows a statistically significant effect for the xz term after the main effects of x and z individually have been partialled out (e.g., Cohen, 1978; Cohen & Cohen, 1975; Saunders, 1956; Zedeck, 1971). The way this is accomplished in practice is by entering x and z into a hierarchical regression model first and then entering the interaction term and assessing the significance level of the increase in the variance of y accounted for.

A recent assessment by Stone and Hollenbeck (1984) of alternative approaches to the detection of moderating effects including subgrouping and "backward" moderated regression in which the interaction term is entered first (e.g., Arnold, 1982; Blood & Mullet, 1977) has substantiated that "conventional" moderated regression, while a conservative approach, is the most suitable procedure available. This

is particularly true when the potential moderating factor is continuous in nature as is the case in the present study.

Rater ability was operationalized as the total score assigned to each section leader on the graphic rating scale used to evaluate their job performance. The moderated regression analysis was performed with both the sales clerks and section leaders again serving as the units of analysis. This was done in order to provide additional substantiation for conclusions drawn on the basis of the results of this procedure.

For the procedure utilizing the sales clerks as the unit of analysis, leniency was measured by determining the difference between the ratings assigned and the scale midpoints for each dimension and calculating the mean of these differences across dimensions for each ratee on each rating format. Separate summary tables for the regression analyses performed with ratings on the graphic rating scale and the behavioral checklist are presented in Table 15. The F statistics shown in this table and the similar tables which follow indicate the significance of the increase in the R² value with the addition of each variable in the regression model. The significant F statistic for the interaction term in the table for the graphic rating scale indicates that the measure of section leader ability did moderate the relationship between the composite attitude measure and the amount of leniency assessed in the performance ratings assigned. This was not found to be true for the behavioral checklist. Application of McNemar's (1969) correction for shrinkage to the R² value for addition of the interaction term in the analysis for the graphic rating scale

Table 15

Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Leniency from Section Leader Composite Attitude Scores With Sales Clerks as the Unit of Analysis

<u>Dependent Variable: Mean Leniency Measure Across Dimensions on the Graphic Rating Scale</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.1642	.0270	.0270	5.46*
Section Leader Performance Rating	.2000	.0400	.0131	2.66
Composite x Performance Rating	.2667	.0711	.0311	6.53**

<u>Dependent Variable: Mean Leniency Measure Across Dimensions on the Behavioral Checklist</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.1316	.0173	.0173	3.90*
Section Leader Performance Rating	.1450	.0210	.0037	0.74
Composite x Performance Rating	.1505	.0227	.0016	0.32

Note: Leniency is measured by determining the difference between the ratings and the scale midpoints for each dimension and calculating the mean of these differences across dimensions for each ratee on each rating format.

n = 199

*p < .05

**p < .01

reduced the value to .0613 which was still significant ($F = 3.26$, $p < .05$). The low correlations between section leader ability and the measures of leniency on both the graphic rating scale and the behavioral checklist, .12 and .09 respectively, preclude the problem of multicollinearity. The estimated power of the test for the significance of the increase in the variance of the dependent variable accounted for by the addition of the interaction terms for the graphic rating scale and behavioral checklist was .99.

The analysis using section leaders as the unit of analysis produced similar findings and the results are shown in Table 16. For these procedures, leniency was measured by determining the mean difference between the ratings and the scale midpoints for each dimension and calculating the mean of these means across dimensions and across rates for each section leader on each rating format. The increase in R^2 value was significant when the interaction term was added to the model for the graphic rating scale but not for the behavioral checklist. The correction for shrinkage reduced the value of F to 3.54 which is significant at $p < .05$. The correlations between section leader performance scores and the leniency measures for the graphic rating scale and behavioral checklist were again quite low - .11 and .14 respectively, and non-significant. The estimated power for the test of the increase in variance accounted for by the addition of the interaction term for both rating formats was .44.

On the basis of the analyses summarized above, it appears that supervisor ability as measured by performance ratings does moderate the relationship between section leader scores on the composite

Table 16

Analysis of Section Leader Performance Ratings as a Moderator Variable
for the Prediction of Rating Leniency from Section Leader Composite
Attitude Scores With Section Leaders as the Unit of Analysis

<u>Dependent Variable: Mean Leniency Measure Across Dimensions on the Graphic Rating Scale</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.1767	.0312	.0312	1.16
Section Leader Performance Rating	.2261	.0511	.0199	0.73
Composite x Performance Rating	.4184	.1751	.1239	5.11*

<u>Dependent Variable: Mean Leniency Measure Across Dimensions on the Behavioral Checklist</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.2475	.0613	.0613	2.34
Section Leader Performance Rating	.2710	.0735	.0122	0.46
Composite x Performance Rating	.1505	.0227	.0016	0.16

Note: Leniency is measured by determining the mean difference between the ratings and the scale midpoints for each dimension and calculating the mean of these difference across dimensions and across rates for each section leader on each scale.

n = 38

*p < .05

attitude measure and the degree of leniency in the performance ratings of subordinates assigned on the trait-based, graphic rating scale. In other words, the relationship between composite attitude scale scores and the degree of leniency in the performance ratings assigned on the graphic rating scale is stronger for section leaders rated as exhibiting high levels of ability than for those considered to have lower levels of ability. This result was not found for the ratings of sales clerks made on the behavioral checklist. Hence, hypothesis 1C was partially supported. However, although the performance measure used to operationalize section leader ability did moderate the basic relationship between section leader attitudes and leniency in assigned ratings, the direction of the basic relationship was opposite that predicted by the model.

Hypothesis 1D. A larger proportion of the variance of the measure of leniency in the assigned ratings of subordinate job performance will be accounted for by rater organizational commitment than by rater job involvement.

The operational definition of leniency based on the level of the ratings (i.e., rating - scale midpoint) was used in testing hypothesis 1D. The statistical procedure followed required first determining Pearson product-moment correlations between the measure of leniency in the ratings assigned on the graphic rating scale and behavioral checklist and section leader scores on the job involvement and organizational commitment scales. These coefficients were next converted to

z scores using Fisher's r to z transformation. The significance of the difference between the two z scores for each rating format was then tested. If one of the z scores was found to be significantly larger than the other, it could then be concluded that one attitude measure accounted for a larger proportion of the variance in the leniency measure than the other. This analysis was conducted using both sales clerks and section leaders were used as the units of analysis.

For the analysis focusing on sales clerks, leniency was measured by determining the difference between the ratings and the scale midpoints for each dimension and calculating the mean of these distances for each ratee across dimensions on each rating format. The correlation coefficients and z difference scores for this test of hypothesis 1D are shown in Table 17. The significance of the bivariate correlations in the table reflect the findings of the tests of hypothesis 1A. However, neither of the z difference values for the graphic rating scale or behavioral checklist were statistically significant. The estimated power of the tests of significance for the product moment correlations was .99. The power of the tests for the difference between coefficients was .85.

Similar results were found when section leaders were used as the units of analysis to test hypothesis 1D. The results of this procedure are presented in Table 18. In this case, leniency was assessed by determining the mean difference between the ratings and the scale midpoints for each dimension and calculating the mean of these means across dimensions and across ratees for each section leader on both

Table 17

Comparison of Correlations Between Section Leader Attitude Scale Scores
and the Mean of Leniency Measures Across Performance Dimensions
with Sales Clerks as the Unit of Analysis

Rating Format	<u>Job Involvement</u>	<u>Organizational Commitment</u>	<u>Zdiff</u>
	<u>r</u>	<u>r</u>	
Graphic Rating Scale	.199**	.067	1.336
Behavioral Checklist	.087	.119*	-.327

Note: Leniency is measured by determining the difference between the ratings and the scale midpoints for each dimension and calculating the mean of these differences for each ratee across dimensions on each rating format.

n = 199

*p < .05

**p < .01

Table 18

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Leniency Measures Across Performance Dimensions and Across Sales Clerks with Section Leaders as the Unit of Analysis

Rating Format	<u>Job Involvement</u>	<u>Organizational Commitment</u>	<u>Zdiff</u>
	<u>r</u>	<u>r</u>	
Graphic Rating Scale	.316*	-.028	1.464
Behavioral Checklist	.067	.239	-.732

Note: Leniency is measured by determining the mean difference between the ratings and the scale midpoints for each dimension and calculating the mean of these differences across dimensions and across rates for each section leader on each rating format.

n = 38

*p < .05

rating formats. The table results indicate that neither of the z difference values were statistically significant. Power estimates for the significance tests of the product-moment correlations and the tests of the differences between coefficients were .46 and .24, respectively.

The findings of the tests of hypothesis 1D indicate that neither job involvement or organizational commitment accounted for a significantly larger proportion of the variance in the measure of leniency in ratings of subordinate job performance. Thus, it must be concluded that hypothesis 1D was not supported for either the graphic rating scale or the behavioral checklist.

Table 19 presents a summary of the results of the tests of hypotheses 1A, 1B, 1C, and 1D. None of the hypotheses as derived from the model except 1C were supported. A significant relationship between the factor considered in the opposite direction predicted was found in testing hypothesis 1A. In essence, these findings indicate that 1) the tendency to assign lenient performance ratings is directly related to a combined measure of how job involved and committed to the organization a rater is, and 2) this relationship is stronger for raters having higher levels of ability as measured by evaluations of their job performance. The basic relationship between rater attitudes and leniency in ratings does not differ for assessments made on trait-based and behaviorally-based rating formats. Also, neither job involvement or organizational commitment is a significantly better predictor of the tendency to assign lenient ratings.

Table 19

Summary of Results for Tests of Leniency Hypotheses 1A - 1D

Unit of Analysis	Attitude Measure	Rating Format	Hypothesis 1A		Hypothesis 1B		Hypothesis 1C	Hypothesis 1D
			Rating Level	Skewness	Rating Level	Skewness	Rating Level	Rating Level
Sales Clerks (ratees)	Composite	GRS	S-	S-	N.S.	N.S.	S+	
		BCL	S-	S-			N.S.	
	Job Involvement	GRS	S-	S-	N.S.	N.S.		N.S.
		BCL	S-	N.S.				N.S.
	Organization Commitment	GRS	N.S.	N.S.	N.S.	N.S.		N.S.
		BCL	S-	S-				N.S.
	Composite	GRS			N.S.		S+	
		BCL					N.S.	
Section Leaders (raters)	Job Involvement	GRS			N.S.			N.S.
		BCL						N.S.
	Organization Commitment	GRS			N.S.			N.S.
		BCL						N.S.

Note: GRS = graphic rating scale. BCL = behavioral checklist. S+ indicates that the hypothesis was supported. S- indicates that a statistically significant result was found in the opposite direction predicted by the hypothesis. N.S. indicates that the hypothesis was not supported.

Central Tendency

Hypothesis 2A. A composite measure of rater job involvement and organizational commitment will be inversely related to the amount of central tendency present in the ratings of subordinate job performance.

The procedures employed to test the hypotheses for central tendency generally follow the pattern established for the corresponding hypotheses dealing with leniency discussed above. The estimated power levels of the tests used to examine hypothesis 2A through 2D are the same as those for the corresponding tests for hypotheses 1A through 1D.

In testing hypothesis 1A, central tendency was assessed on the basis of the absolute distance between each rating and the midpoint of the rating scale for each dimension on the graphic rating scale and behavioral checklist. Low values of this distance measure indicate a high degree of central tendency in the ratings, and vice versa. This approach used sales clerks as the units of analysis and required that the sample be dichotomized on the basis of the median attitude scale scores of the section leaders assigning the ratings for each clerk. Table 20 lists the mean central tendency measures for groups of sales clerks compared on the basis of their section leaders' attitude measures.

A two-way analysis of variance procedure was used to test hypothesis 2A. The analysis was performed three times with group membership based on whether the section leader rating each clerk was

Table 20

Dimension and Overall Means of the Central Tendency Measure for the Sales Clerk Sample Dichotomized on the Basis of the Median Values of Section Leader Attitude Scale Scores

Graphic Rating Scale						
Dimension	<u>Composite Variable</u>		<u>Job Involvement</u>		<u>Organizational Commitment</u>	
	Above Median	Below Median	Above Median	Below Median	Above Median	Below Median
	(<u>n</u> = 98)	(<u>n</u> = 101)	(<u>n</u> = 97)	(<u>n</u> = 102)	(<u>n</u> = 98)	(<u>n</u> = 101)
Dependability	1.26	1.33	1.27	1.32	1.25	1.34
Job Procedures	1.10	1.25	1.07	1.27	1.13	1.20
Sales Ability	1.15	1.26	1.21	1.21	1.14	1.27
Customer Service	1.37	1.46	1.45	1.38	1.39	1.44
Job Knowledge	1.06	1.34	1.03	1.37	1.19	1.20
Product Knowledge	.96	1.16	1.05	1.07	1.02	1.10
Initiative	1.14	1.14	1.08	1.21	1.11	1.17
Judgment	.84	.88	.77	.95	.80	.91
Empl. Relations	1.50	1.57	1.51	1.58	1.50	1.54
Overall Score	1.18	1.25	1.17	1.26	1.19	1.22

Behavioral Checklist						
Dimension	<u>Composite Variable</u>		<u>Job Involvement</u>		<u>Organizational Commitment</u>	
	Above Median	Below Median	Above Median	Below Median	Above Median	Below Median
	(<u>n</u> = 98)	(<u>n</u> = 101)	(<u>n</u> = 97)	(<u>n</u> = 102)	(<u>n</u> = 98)	(<u>n</u> = 101)
Dependability	17.83	18.03	18.02	17.85	17.70	18.16
Job Procedures	18.03	18.15	18.16	18.03	18.06	18.12
Sales Ability	18.65	18.66	18.64	18.67	18.61	18.69
Customer Service	17.24	17.24	17.35	17.13	17.24	17.23
Job Knowledge	14.36	14.54	14.47	14.43	14.43	14.47
Product Knowledge	19.43	19.65	19.38	19.70	19.49	19.60
Initiative	18.12	18.33	18.21	18.24	18.13	18.33
Judgment	19.13	19.29	19.27	19.16	19.13	19.29
Empl. Relations	13.60	13.88	13.78	13.71	13.54	13.94
Overall Score	17.38	17.53	17.47	17.44	17.37	17.54

Note: Central tendency is assessed by calculating the distance of the ratings assigned from the midpoint of the scale for each dimension.

above or below the median on the composite measure, job involvement, or organizational commitment scales. The rating dimensions of each scale constituted the second factor in the design and the central tendency scores on the nine dimensions were the dependent measures. These analyses were performed separately for the graphic rating scale and behavioral checklist rating formats and the results are summarized in Tables 21, 22, and 23.

The Group effect is again of primary interest for testing the hypothesis. The findings shown in Table 21 indicate that for both rating formats the Group effect was statistically significant. However, the means of the central tendency measures as presented in Table 20 again demonstrate that the section leaders who had the higher scores on the job involvement/organizational commitment composite measure exhibited a substantially higher degree of central tendency in the ratings they assigned than the section leaders who scored below the median on this attitude measure. As was the case for hypothesis 1A, this finding is directly opposite to the intended result predicted on the basis of the experimental model being tested.

Table 22 shows that a similar significant relationship for section leader group based only on job involvement scores was found for ratings made on the graphic rating scale but not on the behavioral checklist. In contrast, when the sales clerks were dichotomized on the basis of whether their section leaders' scores on the organizational commitment scale were above or below the median, a significant group effect was found for the behavioral checklist ratings but not those made on the graphic rating scale. In each instance where a

Table 21

Analysis of Variance with Section Leader Composite Score Group and Rating Scale Dimensions as the Independent Variables and Central Tendency Measure as the Dependent Variable

<u>Graphic Rating Scale</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	3.457	4.12*	.004
Dimension	8	7.337	8.27**	.040
Group x Dimension	8	.887	1.06	
Residual	1764	.839		
Total	1781	.870		

<u>Behavioral Checklist</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	10.771	5.60*	.001
Dimension	8	816.806	23.27**	.656
Group x Dimension	8	.430	.22	
Residual	1764	1.925		
Total	1781	5.583		

Note: Central tendency is assessed on the basis of the distance of the ratings from the midpoint of the scale for each dimension.

n = 199

*p < .05 **p < .01

Table 22

Analysis of Variance with Section Leader Job Involvement Score
Group and Rating Scale Dimensions as the Independent Variables
and Central Tendency Measure as the Dependent Variable

<u>Graphic Rating Scale</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	3.288	3.93*	.003
Dimension	8	7.763	9.28**	.040
Group x Dimension	8	.930	1.12	
Residual	1764	.837		
Total	1781	.870		

<u>Behavioral Checklist</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	.680	.35	.000
Dimension	8	816.806	423.92**	.656
Group x Dimension	8	1.256	.65	
Residual	1764	1.927		
Total	1781	5.583		

Note: Central tendency is assessed on the basis of the distance of the ratings from the scale midpoint for each dimension.

n = 199

*p < .05

**p < .01

Table 23

Analysis of Variance with Section Leader Organizational
Commitment Score Group and Rating Scale Dimensions as the
Independent Variables and the Central Tendency Measure
as the Dependent Variable

<u>Graphic Rating Scale</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	.375	.45	.001
Dimension	8	7.763	9.23**	.040
Group x Dimension	8	.274	.33	
Residual	1764	.841		
Total	1781	.870		

<u>Behavioral Checklist</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	12.327	6.42*	.002
Dimension	8	816.678	425.31**	.656
Group x Dimension	8	1.251	.65	
Residual	1764	1.920		
Total	1781	5.583		

Note: Central tendency is assessed on the basis of the distance of the ratings from the scale midpoint for each dimension.

n = 199

* $p < .05$

** $p < .01$

significant effect was found, the group scoring higher on the attitude measure in question tended to exhibit a greater degree of central tendency in the ratings assigned. All of the Dimension effects and none of the interaction effects were again statistically significant. The η^2 values were again relatively small and the power estimates were the same as those for the analogous analyses performed to test hypothesis 1A.

To summarize, hypothesis 2A as stated was not supported by the data. However, a significant effect in the opposite direction predicted by the model was found for on both rating formats for the composite attitude measure, on the graphic rating scale for job involvement, and on the behavioral checklist for organizational commitment.

Hypothesis 2B. The relationship of the composite measure of job involvement and organizational commitment with the amount of central tendency in assigned ratings of subordinate job performance will be stronger for ratings made on a trait-based rating format than on a behaviorally-based rating format.

Hypothesis 2B was examined by means of a procedure similar to that employed in the first test of hypothesis 2A. Central tendency was assessed in terms of the absolute distance from the ratings to the scale midpoints. This analysis was performed using both sales

clerks and section leaders as the units of analysis. In both cases product-moment correlations were first determined between the mean of the central tendency measures across dimensions on the graphic rating scale and behavioral checklist and each of the three attitude measures for the sections leaders assigning the ratings. The three pairs of correlations were then transformed to z scores and the significance of the differences between the corresponding coefficients for the two rating formats was determined.

For the procedure using sales clerks as the units of analysis, central tendency was assessed by determining the distance between the ratings and the scale midpoints for each dimension and calculating the mean of these distances for each ratee across dimensions on each rating format. The results of this analysis are presented in Table 24. The direction and magnitude of the correlations are consistent with the results of hypothesis 2A. Scores on the composite attitude scale were inversely related to the distance measure (and directly related to the assessment of central tendency) on both the graphic rating scale and behavioral checklist rating formats. The job involvement measure was significantly related to central tendency on the graphic rating scale and organizational commitment was similarly related to central tendency in ratings made on the behavioral checklist. The negative signs of these correlations indicate that raters with high scores on the three attitude measures tended to assign ratings that were relatively close to the midpoint of the scales thus indicating a comparatively high degree of central tendency. This is contrary to the prediction of the model. However, none of the z difference

Table 24

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Central Tendency Measures Across Performance Dimensions on the Graphic Rating Scale and the Behavioral Checklist with Sales Clerks as the Unit of Analysis

Attitude Measure	Graphic Rating Scale	Behavioral Checklist	<u>Z</u> diff
	<u>r</u>	<u>r</u>	
Composite Score	-.138*	-.121*	0.139
Job Involvement	-.183**	-.068	1.138
Organizational Commitment	.040	-.120*	1.593

Note: Central tendency is measured by determining the distance between the ratings and the scale midpoints for each dimension and calculating the mean of these distances for each ratee across dimensions on each rating format.

n = 199

*p < .05

**p < .01

statistics from the tests contrasting the strength of the relationship between each attitude measure and the amount of central tendency in the ratings indicated that the association was significantly stronger for one rating format than the other.

The same analysis was also performed using the section leaders as the units of analysis. In this case, central tendency in the ratings was assessed by determining the distance between the ratings and the scale midpoints for each dimension on the graphic rating scale and behavioral checklist and then calculating the mean of these distances across dimensions and across ratees for each section leader. The results of this analysis are shown in Table 25. All of the correlation coefficients were again negative thus indicating an inverse relationship between scores on the attitude variables and the amount of central tendency in the ratings assigned. However, only the correlations of job involvement with central tendency on the graphic rating scale and organizational commitment with central tendency on the behavioral checklist were statistically significant. Once again, none of the differences between the pairs of coefficients for the attitude measures were significant.

Thus, while the results of these procedures proved to be consistent with the tests of hypothesis 2A, the strength of the relationship between rater attitudes and the tendency to assign ratings close to the midpoint of the scales was no stronger for ratings made on the graphic rating scale than for ratings on the behavioral checklist. It may therefore be concluded that hypothesis 2B was not supported.

Table 25

Comparison of Correlations Between Section Leader Attitude Scale Scores and Mean of the Central Tendency Measures Across Performance Dimensions on the Graphic Rating Scale and the Behavioral Checklist with Section Leaders as the Unit of Analysis

Attitude Measure	Graphic Rating Scale	Behavioral Checklist	<u>Z</u> diff
	<u>r</u>	<u>r</u>	
Composite Score	-.167	-.222	.238
Job Involvement	-.308*	-.072	-1.033
Organizational Commitment	-.035	-.262*	.983

Note: Central tendency is measured by determining the mean distance between the ratings and the scale midpoints for each dimension and calculating the mean of these distances across dimensions and across ratees for each section leader on each rating format.

n = 38

*p < .05

Hypothesis 2C. Rater ability will moderate the relationship between the composite measure of rater job involvement and organizational commitment and the amount of central tendency in assigned ratings of subordinate job performance.

The nature of the relationship between rater scores on the composite variable and the amount of central tendency in assigned ratings was found to be contrary to the direction predicted by the model. However, as was the case for hypothesis 1C, the logic underlying hypothesis 2C is unchanged by this fact. Hypothesis 2C proposes that the relationship between raters' scores on the composite attitude measure and the degree of central tendency in the ratings assigned will be stronger for raters with high levels of ability than for those with lower ability levels. It is suggested that ability as measured by ratings of section leader job performance will moderate an individual's likelihood of being able to convert behavioral intentions into behaviors, i.e., assigning ratings with varying amounts of central tendency.

A moderated regression procedure was again used to test this proposition. This analysis was performed using both sales clerks and section leaders as the units of analysis. The same measures of central tendency used for the two tests of hypothesis 2B were employed again for these procedures. Starting with the focus on sales clerks, the results of the regression analyses for ratings on the graphic rating scale and behavioral checklist are presented in Table 26. The significant increase in R^2 with the addition of the interaction term for

Table 26

Analysis of Section Leader Performance Ratings as a Moderator Variable
for the Prediction of Rating Central Tendency from Section Leader
Composite Attitude Scores with Sales Clerks as the Unit of Analysis

<u>Dependent Variable: Mean Central Tendency Measure Across Dimensions on the Graphic Rating Scale</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite (JI + OC)	.1382	.0191	.0191	4.04*
Section Leader Performance Rating	.1952	.0381	.0190	4.02*
Composite x Performance Rating	.2794	.0781	.0399	8.44*

<u>Dependent Variable: Mean Central Tendency Measure Across Dimensions on the Behavioral Checklist</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.1207	.0146	.0173	2.89
Section Leader Performance Rating	.1273	.0162	.0016	0.32
Composite x Performance Rating	.1286	.0165	.0003	0.07

Note: Central tendency is measured by determining the distance between the ratings and the scale midpoints for each dimension and calculating the mean of these distances across dimensions for each ratee on each rating format.

n = 199

*p < .05

the graphic rating scale ratings indicates that ability did moderate the relation between the composite attitude measure and the amount of central tendency in the assigned ratings. The increase with the addition of this term for the behavioral checklist ratings was not significant. When the correction for shrinkage was applied to the \underline{R}^2 values for the addition of the interaction terms, the values for the graphic rating scale and behavioral checklist analyses were reduced to .0638 and .0129, respectively. The increase in variance accounted for as indicated by the shrunken \underline{R}^2 value for the graphic rating scale was still significant.

The results of the same analysis performed with section leaders as the units of analysis are shown in Table 27. Only the increase in \underline{R}^2 for the addition of the interaction term in the analysis for the graphic rating scale was significant ($\underline{R}^2 = .1677$, $p < .05$). Application of the shrinkage correction formula substantially reduced the \underline{R}^2 value for the addition of the graphic rating scale interaction term to .1038.

On the basis of these results, it appears that rater ability, or performance as it was operationalized in this study, did moderate the relationship between the composite attitude measure and the amount of central tendency as measured in ratings of subordinate job performance made on the graphic rating scale. Once again, since hypothesis 2C did not specifically require that the direction of the primary relationship be in accord with the model, it may be concluded that this hypothesis was supported for ratings made on the graphic rating scale but not the behavioral checklist.

Table 27

Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Central Tendency from Section Leader Composite Attitude Scores with Section Leaders as the Unit of Analysis

Dependent Variable: Mean Central Tendency Measure Across Dimensions on the Graphic Rating Scale

Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite (JI + OC)	.1674	.0280	.0280	1.04
Section Leader Performance Rating	.2166	.0469	.0189	0.69
Composite x Performance Rating	.4095	.1677	.1207	4.93*

Dependent Variable: Mean Central Tendency Measure Across Dimensions on the Behavioral Checklist

Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite (JI + OC)	.2215	.0490	.0490	1.86
Section Leader Performance Rating	.2263	.0512	.0021	0.08
Composite x Performance Rating	.2263	.0512	.0000	0.00

Note: Central tendency is measured by determining the mean distance between the ratings and the scale midpoints for each dimension and calculating the mean of these distances across dimensions and across rates for each section leader on each rating format.

n = 38

*p < .05

Hypothesis 2D. A larger proportion of the variance of the measure of central tendency in assigned ratings of subordinate job performance will be accounted for by rater organizational commitment than by rater job involvement.

The procedure for testing hypothesis 2D was essentially the same as that used to examine hypothesis 1D. The amount of central tendency in the ratings assigned was measured in the same way for this hypothesis as it was for hypotheses 2B and 2C. Table 28 shows the correlations and the z difference values for the test of hypothesis 2D using sales clerks as the units of analysis. The significance pattern for the correlation coefficients is consistent with previously presented results again illustrating the direct relationship between scores on the job involvement and organizational commitment scales and the level of central tendency in the ratings assigned. However, neither of the z difference values was statistically significant. This indicates that the strength of the relationship between the amount of central tendency in the ratings and each of the two cognitive measures considered did not differ significantly for evaluations made on either the graphic rating scale or the behavioral checklist.

The corresponding analysis using section leaders as the units yielded similar findings. The results of this procedure are summarized in Table 29. In this case, while all of the r coefficients were negative, only the correlation between job involvement scores and the amount of central tendency in the ratings made on the graphic rating scale was significant. Once again, neither of the z difference

Table 28

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Central Tendency Measures Across Performance Dimensions with Sales Clerks as the Unit of Analysis

Rating Format	<u>Job Involvement</u>	<u>Organizational Commitment</u>	<u>zdiff</u>
	<u>r</u>	<u>r</u>	
Graphic Rating Scale	-.183**	-.040	-1.445
Behavioral Checklist	-.068	-.120*	.574

Note: Central tendency is measured by determining the distance between the ratings and the scale midpoints for each dimension and calculating the mean of these distances across dimensions for each ratee on each rating format.

n = 199

*p < .05 **p < .01

Table 29

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Mean of Central Tendency Measures Across Performance Dimensions and Across Sales Clerks with Section Leaders as the Unit of Analysis

Rating Format	<u>Job Involvement</u>	<u>Organizational Commitment</u>	<u>zdiff</u>
	<u>r</u>	<u>r</u>	
Graphic Rating Scale	-.308*	-.035	-1.165
Behavioral Checklist	-.072	-.262	.793

Note: Central tendency is measured by determining the mean distance between the ratings and the scale midpoints for each dimension and calculating the mean of these distances across dimensions and across ratees for each section leader on each rating format.

n = 38

*p < .05

values was large enough to be statistically significant. Hence, on the basis of the results from the tests of hypothesis 2D, it is apparent that neither organizational commitment or job involvement account for a significantly larger proportion of the variance in the measure of central tendency in the ratings assigned on both the graphic rating scale and the behavioral checklist.

The results of the tests of hypotheses 2A, 2B, 2C, and 2D are summarized in Table 30. In contrast to the predictions of the model, the composite measure of job involvement and organizational commitment was directly related to the amount of central tendency present in the ratings assigned for both the graphic rating scale and behavioral checklist formats. This relationship was also found to be present for the job involvement measure and the ratings made on the graphic rating scale as well as for the organizational commitment measure and subordinate evaluations made using the behavioral checklist. Rater ability was again found to moderate the relationship between the composite attitude measure and the prevalence of central tendency in the graphic rating scale ratings. Overall, the relationship between rater attitudes and the amount of central tendency in ratings assigned was no different for the graphic rating scale and behavioral checklist formats. Also, organizational commitment did not account for a significantly greater proportion of variance in the measure of central tendency than did job involvement regardless of the rating format considered.

Table 30

Summary of Results for Tests of Central Tendency Hypotheses 2A - 2D

Unit of Analysis	Attitude Measure	Rating Format	Hypothesis 2A	Hypothesis 2B	Hypothesis 2C	Hypothesis 2D
			Rating-Midpt. Distance	Rating-Midpt. Distance	Rating-Midpt. Distance	Rating-Midpt. Distance
Sales Clerks (ratees)	Composite JI + OC	GRS	S-		S+	
		BCL	S-	N. S.	N. S.	
	Job Involvement	GRS	S-			N. S.
		BCL	N. S.	N. S.		N. S.
	Organization Commitment	GRS	N. S.			N. S.
		BCL	S-	N. S.		N. S.
Section Leaders (raters)	Composite JI + OC	GRS			S+	
		BCL		N. S.	N. S.	
	Job Involvement	GRS				N. S.
		BCL		N. S.		N. S.
	Organization Commitment	GRS				N. S.
		BCL		N. S.		N. S.

Note: GRS = graphic rating scale. BCL = behavioral checklist. S+ indicates that the hypothesis was supported. S- indicates that a statistically significant result was found in the opposite direction predicted by the hypothesis. N.S. indicates that the hypothesis was not supported.

Restriction of Range

Hypothesis 3A. A composite measure of rater job involvement and organizational commitment will be inversely related to the amount of restriction of range present in assigned ratings of subordinate job performance.

Hypothesis 3A was tested using two different operational definitions of restriction of range. The first of these was based on the size of the standard deviation of the ratings assigned on each dimension. The units of analysis for this procedure were the sales clerks. The sales clerks in the sample were dichotomized three times on the basis of whether the section leaders who evaluated them scored above or below the median value on each of the three attitude measures considered in the study. The standard deviations of the ratings assigned on the nine dimensions of the graphic rating scale and behavioral checklist were then determined for the clerks in each subsample. The nine matched values for both rating formats were compared for the three pairs of dichotomized sales clerk subgroups by means of t test procedures. The power of these tests was .16 because of the small number of dimensions in each scale. However, the standard deviations used in this analysis were all based on approximately 100 ratings.

The results of the t tests for differences between the means of the dimension standard deviations are presented in Table 31. Significant differences between group means of the dimension standard

Table 31

Comparison of Restriction of Range in Ratings Assigned by Section Leaders Scoring Above and Below the Median on the Three Attitude Measures

Rating Format	Group	<u>n</u>	Mean	Standard Deviation	df	<u>t</u>
<u>Independent Variable: Composite Measure</u>						
Graphic Rating Scale	Above Median	98	1.113	0.121	16	-2.27*
	Below Median	101	1.241	0.119		
Behavioral Checklist	Above Median	98	1.295	0.135	16	-2.19*
	Below Median	101	1.429	0.125		
<u>Independent Variable: Job Involvement Score</u>						
Graphic Rating Scale	Above Median	98	1.086	0.108	16	-3.44**
	Below Median	101	1.277	0.127		
Behavioral Checklist	Above Median	97	1.377	0.185	16	0.04
	Below Median	102	1.373	0.241		
<u>Independent Variable: Organizational Commitment Score</u>						
Graphic Rating Scale	Above Median	97	1.150	0.122	16	0.77
	Below Median	102	1.193	0.116		
Behavioral Checklist Scale	Above Median	98	1.280	0.147	16	-2.76*
	Below Median	101	1.458	0.126		

Note: Restriction of range is measured by determining the standard deviations of the ratings assigned on the nine dimensions of each rating format. The means reported are the average of nine standard deviations based on the number of cases listed in the table.

* $p < .05$

** $p < .01$

deviations for both the graphic rating scale and the behavioral checklist were found when the sample was split on the basis of section leader scores on the composite variable. However, once again, the direction of the difference was inverse to the direction predicted by the model. The ratings made by section leaders scoring above the median on the composite measure showed significantly more range restriction than the assessments made by raters who had lower scores. Similar significant results were also found for ratings made on the graphic rating scale when the sample was split on the basis of section leader job involvement scores and on the behavioral checklist evaluations when organizational commitment was the dichotomizing factor. The predictor-criterion relationships of these findings were also in the opposite direction of that suggested by the hypothesis 3A.

The second approach used to operationally define the degree of restriction of range in ratings was based on the kurtosis of the distributions. The prediction made from the model was that the distribution of ratings assigned by section leaders who scored lower on the attitude measures would be comparatively peaked or narrow (mesokurtic) thus indicating a high degree of range restriction. In contrast, the distribution of ratings assigned by section leaders who exhibited higher levels of these attitudes was predicted to be more widely dispersed or flatter (platykurtic) thereby indicating less restriction of range. For this analysis, the sales clerk sample was again dichotomized on the basis of the three section leader attitude measures. The kurtosis of the distributions of ratings assigned on the nine dimensions of the graphic rating scale and behavioral

checklist was determined for the various sales clerk subsamples. A t test procedure was then used to determine whether the means of the kurtosis values differed for the pairs of contrasted sales clerk groups. The estimated power for this procedure is the same as that of the analysis using the skewness measure of leniency to test hypothesis 1A. The results of this procedure are presented in Table 32.

The findings of this analysis indicate that when the sales clerk sample was divided on the basis of section leader composite scores, the distribution of ratings on both the graphic rating scale and behavioral checklist for those employees assessed by the lower scoring section leaders tended to be significantly less peaked than the distribution of the ratings of the contrasted group. This result is again contrary to the prediction of the model but it is complementary to the results of the previous test of hypothesis 3A. Similarly, when the sales clerk sample was dichotomized on the basis of section leader job involvement scores, a significant difference in the kurtosis of the dimension ratings was found for the graphic rating scale assessments only. When organizational commitment was used as the splitting factor, a significant difference found was for ratings made on the behavioral checklist. The direction of these results as well as the correspondence between the two attitude scales and the rating format on which the significant differences in range restriction were found again coincide with the results of the test of hypothesis 3A based on the standard deviation of the ratings.

Hence, it may be concluded that the composite measure of the raters' scores on the job involvement and organizational commitment

Table 32

Comparison of Mean Dimension Kurtosis (Restriction of Range) in Ratings Assigned by Section Leaders Scoring Above and Below the Median on the Three Attitude Measures

Rating Format	Group	<u>n</u>	Mean Kurtosis ^a	Standard Deviation	df	<u>t</u>
<u>Composite Variable (JI + OC)</u>						
Graphic Rating Scale	Above Median	98	-0.102	0.284	16	2.14*
	Below Median	101	-0.399	0.354		

Behavioral Checklist	Above Median	98	-0.419	0.222	16	2.11*
	Below Median	101	-0.713	0.356		

<u>Job Involvement</u>						
Graphic Rating Scale	Above Median	98	-0.134	0.299	16	2.43*
	Below Median	101	-0.471	0.379		

Behavioral Checklist	Above Median	98	-0.457	0.209	16	1.68
	Below Median	101	-0.683	0.333		

<u>Organizational Commitment</u>						
Graphic Rating Scale	Above Median	97	-0.089	0.184	16	1.85
	Below Median	102	-0.337	0.207		

Behavioral Checlist	Above Median	97	-0.375	0.222	16	2.70*
	Below Median	102	-0.727	0.399		

^aMean kurtosis of the nine dimension score distributions.

* $p < .05$

scales was inversely related to the range of the subordinate performance ratings assigned on both the graphic rating scale and the behavioral checklist. Rater job involvement scores were directly related to rating range restriction on the graphic rating scale and the same was true for the measure of organizational commitment and the degree of restriction of range for ratings made using the behavioral checklist. Once again the data indicate conclusively that contrary to the predictions of hypothesis 3A derived from the model, rater job involvement and organizational commitment are directly related to the extent of restriction of range ratings of subordinate job performance.

Hypothesis 3B. The relationship between the composite measure of rater job involvement and organizational commitment and the amount of restriction of range in the assigned ratings of subordinate job performance will be stronger for ratings made on a trait-based rating format than on a behaviorally-based rating format.

Hypothesis 3B was also tested using both operational definitions of restriction of range. Section leaders were the units of analysis for the first procedure in which restriction of range was measured by calculating the standard deviation of the ratings assigned on each dimension by each section leader and then determining the mean of these values for the two rating formats. These means were next correlated with section leader scores on the three attitude measures. Finally, the correlations for the graphic rating scale and behavioral

checklist were transformed to z scores and the significance of the difference between the values for each attitude measure was determined. The estimated power of this procedure was .24. Table 33 shows the results of this analysis. None of the differences between the pairs of z values were statistically significant. Also, none of the correlations between attitude scores and the measure of restriction of range in the ratings on either scale were significant. The signs and relative magnitude of the coefficients, however, correspond with the test results for hypothesis 3A

An analysis of variance approach was used in the second test of hypothesis 3B and range restriction was operationalized by measuring the kurtosis of the dimensional rating distributions. This procedure parallels the method used to test hypothesis 1B which relied upon the skewness definition of leniency and the power estimates are the same. The two factors in this analysis were Group - whether each rater scored above or below the median on the attitude scale, and Format - graphic rating scale or behavioral checklist. The dependent measure was the kurtosis of each of the dimension rating distributions. The significant Group effect supports the results of the tests for hypothesis 3A. The significant effects for Format and an examination of the mean kurtosis values indicate that the ratings on the behavioral checklist tended to be more widely dispersed overall than the assessments made on the graphic rating scale. Once again, the significance of the interaction term was of primary importance for testing the hypothesis. The ANOVA summary tables shown in Table 34 indicate that in none of the three cases considered was a significant

Table 33

Comparison of Correlations Between Section Leader Attitude Scale Scores and a Measure of Restriction of Range of the Ratings Assigned by Each Section Leader on the Graphic Rating Scale and the Behavioral Checklist

Attitude Measure	Graphic Rating Scale	Behavioral Checklist	<u>z</u> _{diff}
	<u>r</u>	<u>r</u>	
Composite Score (JI + OC)	-.130	-.105	-.010
Job Involvement	-.001	-.165	.699
Organizational Commitment	-.192	-.004	-.803

Note: Restriction of range is measured by calculating the standard deviation of the ratings assigned on each dimension by each section leader and determining the mean of these SDs for each rating format.

n = 38

All r and z values are non-significant at p < .05.

Table 34

Analysis of Variance with Section Leader Composite Attitude Score Group and Rating Scale as the Independent Variables and Dimension Kurtosis as the Dependent Variable

<u>Composite Measure (JI + OC)</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	0.981	4.071*	.230
Scale	1	1.169	5.937*	.020
Group x Scale	1	0.001	0.010	
Residual	32	0.197		
Total	35	0.241		

<u>Job Involvement</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	0.828	5.173*	.260
Scale	1	1.367	8.539**	.015
Group x Scale	1	0.001	0.009	
Residual	32	0.160		
Total	35	0.209		

<u>Organizational Commitment</u>				
Source	df	Mean Square	<u>F</u>	eta ²
Group	1	1.039	4.743*	.190
Scale	1	2.047	9.348**	.044
Group x Scale	1	0.001	0.005	
Residual	32	0.219		
Total	35	0.289		

n = 36

*p < .05

**p < .01

interaction between section leader group and rating format found. Therefore, on the basis of all of the tests conducted, it is concluded that hypothesis 3B was not supported.

Hypothesis 3C. Rater ability will moderate the relationship between the composite measure of job involvement and organizational commitment and the amount of restriction of range in assigned ratings of subordinate job performance.

Hypothesis 3C addresses the question of whether rater ability moderates the relationship between raters' combined levels of job involvement and organizational commitment and the extent to which the range of performance ratings assigned is restricted. As was true for leniency and central tendency, the direction of the relationship between the composite attitude measure and the prevalence of this type of error in the ratings was found to be significant but in the opposite direction predicted. It is proposed by hypothesis 3C that the relationship between the attitude measure and the amount of rating range restriction will be more pronounced for raters evidencing higher levels of ability than for those with lower ability levels. Once again, rater ability was operationalized on the basis of the evaluations made of the section leaders' job performance.

The dependent measure in this analysis was the degree of restriction in the range of the distribution of ratings made on each scale of the graphic rating scale and behavioral checklist. This operational definition of the dependent variable required that the moderated

regression analysis conducted to test hypothesis 3C be performed only section leaders as the units of analysis. Restriction of range was measured by determining the standard deviation of the ratings assigned on each dimension by each section leader and calculating the mean of these values for the graphic rating scale and the behavioral checklist. The power of this procedure was that same as for the second test of hypothesis 1C which averaged leniency across rateres for each section leader.

Summary tables showing the results of the regression analyses for the tests of hypothesis 3C are presented in Table 35. None of the F tests for the change in R^2 with the addition of each variable into the model reached the level of statistical significance. Failure of the addition of the interaction terms to result in a significant increase in the amount of variance accounted for in the dependent measure indicates that the rater ability apparently did not moderate the relationship between the composite attitude scores and the measure of restriction of range for either rating format. However, in view of the single test of this hypothesis and the relatively low power of the statistical procedure due to the limited number of section leaders in the sample, conclusions regarding hypothesis 3C should be considered tentative pending further research.

Hypothesis 3D. A larger proportion of the variance in the measurement of restriction of range in assigned ratings of subordinate job performance will be accounted for by rater organizational commitment than by rater job involvement.

Table 35

Analysis of Section Leader Performance Ratings as a Moderator Variable for the Prediction of Rating Restriction of Range from Section Leader Composite Attitude Scores with Section Leaders as the Unit of Analysis

<u>Dependent Variable: Restriction of Range Measure on the Graphic Rating Scale</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite (JI + OC)	.1299	.0169	.0169	.618
Section Leader Performance Rating	.1299	.0169	.0169	.013
Composite x Performance Rating	.2061	.0423	.0256	.909

<u>Dependent Variable: Restriction of Range Measure on the Behavioral Checklist</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite (JI + OC)	.1053	.0111	.0111	.404
Section Leader Performance Rating	.2278	.0519	.0408	1.506
Composite x Performance Rating	.2597	.0675	.0156	.568

Note: Restriction of range is measured by calculating the standard deviation of the ratings assigned on each dimension by each section leader and determining the mean of these SDs for each rating format.

n = 38

All F ratios are non-significant at p < .05.

The test of hypothesis 3D was performed using the operational measure of range restriction based on the mean of the standard deviations of the ratings on each dimension averaged across rateres for each section leader. The same procedure used for hypotheses 1D and 2D was applied again in this instance. The correlations between scores on the job involvement and organizational commitment scales and the measure of range restriction in assigned ratings are shown in Table 36. Also shown are the z difference values for the comparison of corresponding correlations on each rating format. The power of the tests for differences between pairs of coefficients was .24. None of the product-moment correlations or differences between coefficients were significant. This finding is consistent with the results of previously reported tests using section leaders as the units of analysis. These findings indicate that neither job involvement or organizational commitment scores accounted for a significantly larger proportion of the variance in the measure of restriction of range in assigned performance ratings on the graphic rating scale or the behavioral checklist. Once again, the limited power of the analysis performed and the availability of only one procedure to test hypothesis 1D necessitate that any conclusions drawn concerning this hypothesis be viewed cautiously.

The results of the tests of hypotheses 3A, 3B, 3C, and 3D are summarized in Table 37. A significant relationship was found between the composite measure of rater job involvement and organizational commitment and the amount of restriction of range in the ratings of subordinate job performance. However, the relationship was direct in

Table 36

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Restriction of Range in Ratings Assigned on the Graphic Rating Scale and the Behavioral Checklist

Rating Format	<u>Job Involvement</u>	<u>Organizational Commitment</u>	<u>zdiff</u>
	<u>r</u>	<u>r</u>	
Graphic Rating Scale	-.001	-.192	.815
Behavioral Checklist	-.165	-.004	-.686

Note: Restriction of range is measured by calculating the standard deviation of the ratings assigned on each dimension by each section leader and determining the mean of these SDs for each rating format.

n = 38.

All r and z values are non-significant at $p < .05$.

Table 37

Summary of Results for Tests of Restriction of Range Hypotheses 3A - 3D

Unit of Analysis	Attitude Measure	Rating Format	Hypothesis 1A		Hypothesis 1B		Hypothesis 1C	Hypothesis 1D
			Standard Deviation	Kurtosis	Standard Deviation	Kurtosis	Standard Deviation	Standard Deviation
Sales Clerks (raters)	Composite	GRS	S-	S-				
		BCL	S-	S-		N.S.		
	Job Involvement	GRS	S-	S-				
		BCL	N.S.	N.S.		N.S.		
	Organization Commitment	GRS	N.S.	N.S.				
		BCL	S-	S-		N.S.		
Section Leaders (raters)	Composite	GRS					N.S.	
		BCL			N.S.		N.S.	
	Job Involvement	GRS						N.S.
		BCL			N.S.			N.S.
	Organization Commitment	GRS						N.S.
		BCL			N.S.			N.S.

Note: GRS = graphic rating scale. BCL = behavioral checklist. S+ indicates that the hypothesis was supported. S- indicates that a statistically significant result was found in the opposite direction predicted by the hypothesis. N.S. indicates that the hypothesis was not supported.

nature as opposed to inverse as predicted by the theoretical model. The measure of job involvement was predictive of range restriction on the graphic rating scale and organizational commitment scores were found to be related to the tendency to exhibit this error in ratings made on the behavioral checklist. None of the other statistical tests performed in this series yielded significant results in either direction. However, conclusions drawn in this regard are must be considered tentative owing to the operational measures of range restriction which permitted only section leaders to be used as the units of analysis in tests of hypotheses 3C and 3D.

Halo

Hypothesis 4A. A composite measure of rater job involvement and organizational commitment will be inversely related to the amount of halo present in assigned ratings of subordinate job performance.

Hypothesis 4A was tested using two different approaches to the measurement halo error in the ratings assessing the job performance of the sales clerks. Sales clerks served as the units of analysis for both methods. The first procedure involved determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format. Sales clerks in the sample were then dichotomized into subsamples to be contrasted on the basis of whether the scores of their section leaders on the job involvement, organizational commitment, and composite attitude variable scales were above or

below the median values of the score distributions for each measure. A t test procedure was then applied to determine if the differences in the amount of halo in the ratings of the compared groups was statistically significant. The estimated power of each of the t tests used in this method for testing hypothesis 4A was .94.

The results of the t test analyses used to test hypothesis 4A are presented in Table 38. The measures of halo for the two rating formats are not directly comparable from the values shown in this table because of differences in the potential range of ratings on the graphic rating scale and the behavioral checklist. Considering the composite measure first, the results of the group comparisons indicated that a significant difference was present for the ratings made by section leaders scoring above and below the median value. As was the case for the three previously considered rating errors, section leaders who demonstrated higher levels of the composite variable tended to assign performance ratings characterized by a greater degree of halo than section leaders who scored lower on this measure. The difference in the mean halo measure for ratings assigned on the behavioral checklist by the two section leader groups was not significant.

Similar results were found when the section leader sample was split on the basis of job involvement scores. The mean difference in the measure of halo for the ratings made on the graphic rating scale was statistically significant but in the opposite direction suggested by the model. Once again, the t value for the difference of the mean halo measures on the behavioral checklist was not significant. For

Table 38.

Comparison of the Measurement of Halo in Ratings Assigned by Section Leaders Scoring Above and Below the Median on the Three Attitude Measures

Rating Format	Group	<u>n</u>	Mean Halo Measure	Standard Deviation	df	<u>t</u>
<u>Composite Variable (JI + OC)</u>						
Graphic Rating Scale	Above Median	98	0.729	0.336	197	-1.98*
	Below Median	101	0.824	0.348		

Behavioral Checklist	Above Median	98	1.091	0.369	197	-0.77
	Below Median	101	1.132	0.374		

<u>Job Involvement</u>						
Graphic Rating Scale	Above Median	97	0.719	0.350	197	-2.23*
	Below Median	102	0.827	0.331		

Behavioral Checklist	Above Median	97	1.107	0.367	197	-0.20
	Below Median	102	1.117	0.377		

<u>Organizational Commitment</u>						
Graphic Rating Scale	Above Median	98	0.750	0.334	197	-1.11
	Below Median	101	0.804	0.353		

Behavioral Checklist	Above Median	98	1.113	0.337	197	0.03
	Below Median	101	1.111	0.403		

Note: Halo is measured by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format.

* $p < .05$

the analyses in which the sales clerks were divided on the basis of their supervisors' organizational commitment scale scores, the measures of halo did not differ significantly on either the graphic rating scale or the behavioral checklist.

The second operational definition of halo used to test hypothesis 4A involved consideration of the intercorrelations of the various dimension scores on the graphic rating scale and the behavioral checklist. The sales clerk sample was again dichotomized on the basis of section leader composite attitude scores relative to the median on the distribution. For each subsample, a multi-trait, multimethod (MTMM) correlation matrix was developed (Campbell & Fiske, 1959) with dimensions constituting the different traits and the graphic rating scale and behavioral checklist rating formats representing alternative methods. Data from these matrices were also used in testing hypothesis 6 and will be reviewed in detail when the results of that analysis are presented.

Table 39 shows the MTMM matrix for the sales clerks rated by section leaders scoring above the median on the composite attitude measure. The matrix based on the ratings made by section leaders scoring below the median is presented in Table 40. In both matrices, the values in circles represent the validity coefficients. These are the correlations between ratings of the same dimensions made on the two different formats. The solid lines enclose the heterotrait-monomethod coefficients representing the intercorrelations of the various dimension ratings all made on the same rating format. The coefficients within the broken lines comprise the heterotrait-

Table 39
Multitrait - Multimethod Correlation Matrix for Sales Clerks Rated by Section Leaders
Scoring Above the Median on the Job Involvement - Organizational Commitment Composite Variable

	Graphic Rating Scale								Behavioral Checklist								
	DEP	JOBPR	SAL	CUST	COMPL	PROKN	INIT	JUDGE	EMPL	DEP	JOBPR	SAL	CUST	COMPL	PROKN	INIT	JUDGE
JOBPRGRS	.72	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
SALGRS	.68	.64	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
CUSTGRS	.65	.60	.73	--	--	--	--	--	--	--	--	--	--	--	--	--	--
COMPLGRS	.69	.69	.62	.65	--	--	--	--	--	--	--	--	--	--	--	--	--
PROKNGRS	.44	.42	.62	.48	.50	--	--	--	--	--	--	--	--	--	--	--	--
INITGRS	.64	.52	.51	.55	.57	.48	--	--	--	--	--	--	--	--	--	--	--
JUDGEGRS	.56	.68	.59	.56	.67	.59	.72	--	--	--	--	--	--	--	--	--	--
EMPLGRS	.54	.44	.46	.57	.52	.41	.62	.48	--	--	--	--	--	--	--	--	--
DEPBCL	.55	.31	.30	.22	.30	.23	.42	.27	.31	--	--	--	--	--	--	--	--
JOBPRBCL	.39	.42	.25	.27	.34	.14	.31	.31	.22	.41	--	--	--	--	--	--	--
SALBCL	.23	.31	.32	.17	.20	.34	.27	.39	.15	.19	.28	--	--	--	--	--	--
CUSTBCL	.33	.17	.26	.30	.18	.26	.43	.33	.49	.38	.33	.24	--	--	--	--	--
COMPLBCL	.46	.34	.32	.17	.26	.22	.43	.26	.43	.64	.50	.31	.54	--	--	--	--
PROKNBCL	.44	.39	.40	.29	.38	.51	.46	.54	.29	.46	.30	.59	.30	.46	--	--	--
INITBCL	.40	.35	.42	.32	.38	.43	.58	.56	.31	.30	.24	.49	.25	.44	.62	--	--
JUDGEBCL	.39	.38	.35	.26	.29	.43	.43	.63	.37	.34	.37	.48	.34	.44	.66	.53	--
EMPLBCL	.12	.18	.05	.07	.16	.19	.35	.31	.56	.26	.19	.17	.34	.43	.24	.29	.52

$n = 98$

Table 40

Multitrait - Multimethod Correlation Matrix for Sales Clerks Rated by Section Leaders
Scoring Below the Median on the Job Involvement - Organizational Commitment Composite Variable

	Graphic Rating Scale									Behavioral Checklist								
	DEP	JOBPR	SAL	CUST	COMPL	PROKN	INIT	JUDGE	EMPL	DEP	JOBPR	SAL	CUST	COMPL	PROKN	INIT	JUDGE	
JOBPRGRS	.49	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
SALGRS	.29	.64	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
CUSTGRS	.30	.49	.71	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--
COMPLGRS	.36	.64	.48	.43	--	--	--	--	--	--	--	--	--	--	--	--	--	--
PROKNGRS	.25	.62	.67	.55	.59	--	--	--	--	--	--	--	--	--	--	--	--	--
INITGRS	.51	.63	.52	.53	.61	.61	--	--	--	--	--	--	--	--	--	--	--	--
JUDGEGRS	.35	.59	.58	.49	.55	.65	.72	--	--	--	--	--	--	--	--	--	--	--
EMPLGRS	.34	.46	.49	.55	.37	.34	.47	.48	--	--	--	--	--	--	--	--	--	--
DEPBCL	.57	.26	.20	.15	.17	.14	.34	.12	.22	--	--	--	--	--	--	--	--	--
JOBPRBCL	.29	.40	.21	.12	.31	.18	.34	.22	.21	.46	--	--	--	--	--	--	--	--
SALBCL	.24	.43	.52	.45	.37	.61	.49	.50	.33	.19	.31	--	--	--	--	--	--	--
CUSTBCL	.47	.37	.37	.49	.35	.24	.54	.43	.48	.37	.32	.42	--	--	--	--	--	--
COMPLBCL	.48	.44	.27	.27	.36	.25	.43	.31	.35	.43	.41	.27	.61	--	--	--	--	--
PROKNBCL	.23	.55	.59	.43	.51	.66	.55	.54	.21	.24	.39	.64	.35	.36	--	--	--	--
INITBCL	.41	.46	.31	.23	.36	.42	.60	.50	.26	.43	.38	.43	.50	.55	.51	--	--	--
JUDGEBCL	.27	.51	.55	.44	.42	.57	.60	.61	.46	.27	.35	.63	.52	.47	.63	.60	--	--
EMPLBCL	.37	.33	.17	.29	.19	.15	.36	.30	.57	.23	.43	.24	.50	.54	.25	.41	.37	--

n = 101.

heteromethod triangles and indicate the intercorrelations of the nine dimension ratings made on the two different instruments. Coefficients greater than or equal to .198 are significant at $p < .05$.

This analysis focused on the product-moment correlations in the heterotrait-monomethod triangles. These values are indicative of the relative amount of halo in the ratings made on each rating format. A comparison was made between the coefficients in the corresponding triangles from the matrices in Tables 39 and 40. The procedure required converting the correlations to z scores and using a t test procedure to determine the significance of the difference between the matched pairs of coefficients. This analysis was performed separately for the graphic rating scale and behavioral checklist. The estimated power of each of these comparisons was .55. The results of this analysis are shown in the upper portion of Table 41. A significant t value was found for the difference between the means of the z scores for the graphic rating scale. The means shown in the table indicate a greater degree of halo in the ratings assigned by section leaders scoring above the median on the composite attitude measure. The difference between the correlation means for the behavioral checklist was not significant. These results reinforce the findings of the first procedure used to test hypothesis 4A.

The relative amounts of halo in the ratings made on the graphic rating scale and behavioral checklist were also contrasted for both groups of sales clerks. The results of this procedure are included in the lower portion of Table 41. The t values determined for the difference between the z scores for the matched pairs of correlations

Table 41

Comparison of the Measurement of Halo in Ratings Assigned on the Graphic Rating Scale and Behavioral Checklist by Section Leaders Scoring Above and Below the Median on the Composite Attitude Measure

<u>Comparison of Halo in Ratings Made by the Above vs. Below Groups</u>						
Rating Format	Group	<u>n</u>	Standardized Mean Halo Measure (z)	Standard Deviation	df	<u>t</u>
Graphic Rating Scale	Above Median	98	0.670	0.139	70	2.57 ³⁸
	Below Median	101	0.576	0.165		

Behavioral Checklist	Above Median	98	0.419	0.164	70	-0.81
	Below Median	101	0.450	0.154		

<u>Comparison of Halo in Ratings Made on the GRS vs. BCL</u>						
Group	Rating Format	<u>n</u>	Standardized Mean Halo Measure (<u>z</u>)	Standard Deviation	df	<u>t</u>
Above Median	GRS	98	0.670	0.139	197	6.85**
	BCL	98	0.419	0.164		

Below Median	GRS	101	0.576	0.165	70	3.30**
	BCL	101	0.450	0.154		

Note: Halo is measured by determining the mean intercorrelation of the dimensional performance ratings from the monomethod-heterotrait triangles of the multitrait-multimethod matrices (Tables 39 and 40). All correlations were converted to z scores for use in this analysis. The n values indicate the number of pairs of ratings correlated.

* $p < .05$

** $p < .01$

indicate that there was more halo in the ratings made on the graphic rating scale than the behavioral checklist regardless of the attitude scores of the raters. This finding is consistent with the literature indicating that graphic rating scales are particularly prone to the halo error while ratings made on behavioral checklists tend to be less effected by this form of systematic bias (e.g., Cascio, 1978).

To summarize, the results of the statistical analyses conducted to test hypothesis 4A indicate that the hypothesis was not supported. However, in contrast to the predictions of the model, a significant direct relationship was found between rater scores on the composite variable and the amount of halo in performance ratings assigned to subordinates. The same result was found for the measure of section leader job involvement with ratings made on the graphic rating scale. Organizational commitment scores were not significantly related to the degree of halo in performance evaluations assigned on either rating format.

Hypothesis 4B. The relationship between the composite measure of rater job involvement and organizational commitment will be stronger for assigned ratings of subordinate job performance made on a trait-based rating format than on a behaviorally-based rating format.

The procedure employed to test the significance of the difference between the three pairs of z scores in examining the previous B hypo-

theses was used again in this instance. The analysis was performed using both sales clerks and section leaders as the units of analysis. For the sales clerk analysis, halo was measured by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of the graphic rating scale and behavioral checklist. For the section leader analysis, halo was assessed by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format and then calculating the mean of these values for each section leader.

The results of the analyses using sales clerks as the experimental units are presented in Table 42. Only the correlations between the measure of halo in ratings made on the graphic rating scale and section leader scores on the composite and job involvement scales were statistically significant. This finding is consistent with the results from the tests of hypothesis 4A. The magnitude of the z difference scores in these tables failed to provide support for hypothesis 4B. Similar conclusions were drawn from the results of the analysis using section leaders as the units of analysis. The coefficients resulting from this procedure are shown in Table 43. Thus, it may be concluded that the strength of the relationship between supervisor attitudes and the prevalence of halo in assigned performance ratings does not differ significantly for trait-based and behaviorally-based rating formats.

Table 42

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Halo on the Graphic Rating Scale and the Behavioral Checklist with Sales Clerks as the Unit of Analysis

Attitude Measure	Graphic Rating Scale	Behavioral Checklist	z_{diff}
	r	r	
Composite Score	-.121*	-.095	-.257
Job Involvement	-.161*	-.102	0.584
Organizational Commitment	-.058	-.050	-0.079

Note: Halo is measured by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format.

$n = 199$

* $p < .05$

Table 43

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Halo on the Graphic Rating Scale and the Behavioral Checklist with Section Leaders as the Unit of Analysis

Attitude Scale	Graphic Rating Scale	Behavioral Checklist	<u>z</u> diff
	<u>r</u>	<u>r</u>	
Composite Score	-.155	-.022	-.569
Job Involvement	-.202	-.014	-.812
Organizational Commitment	-.045	-.020	-.105

Note: Halo is measured by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format and calculating the mean of these SDs for each section leader.

n = 38

All r and z values are non-significant at p < .05.

Hypothesis 4C. Rater ability will moderate the relationship between the composite measure of job involvement and organizational commitment and the amount of halo present in assigned ratings of subordinate job performance.

Hypothesis 4C proposes that the relationship between rater scores on the composite attitude scale and the degree of halo in the subordinate performance ratings they assign will be stronger for raters having high ability levels and vice versa. The moderated regression technique described above was employed again to test hypothesis 4C with both sales clerks and section leaders serving as the units of analysis. Halo was assessed for these procedures in the same manner as described for hypothesis 4B. The results of the analyses performed to test hypothesis 4C are shown in Tables 44 and 45.

None of the F ratios testing the significance of the increase in variance of the halo measure accounted for by the addition of the interaction terms were statistically significant. This indicates that regardless of the rating format or units of analysis considered, rater ability did not moderate the relationship between rater composite attitude scores and the prevalence of halo in subordinate performance ratings. Thus, hypothesis 4C was not supported by the data.

Table 44

Analysis of Section Leader Performance Ratings as a Moderator Variable
for the Prediction of Rating Halo from Section Leader Composite
Attitude Scores with Sales Clerks as the Unit of Analysis

<u>Dependent Variable: Halo on the Graphic Rating Scale</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.1210	.0146	.0146	2.39
Section Leader Performance Rating	.1286	.0165	.0019	0.38
Composite x Performance Rating	.1327	.0176	.0011	0.22

<u>Dependent Variable: Halo on the Behavioral Checklist</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.0954	.0091	.0091	1.81
Section Leader Performance Rating	.0968	.0094	.0003	0.05
Composite x Performance Rating	.1016	.0103	.0010	0.19

Note: Halo is measured by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format.

n = 199

All F ratios are non-significant at p < .05.

Table 45

Analysis of Section Leader Performance Ratings as a Moderator Variable
for the Prediction of Rating Halo from Section Leader Composite
Attitude Scores with Section Leaders as the Unit of Analysis

<u>Dependent Variable: Halo on the Graphic Rating Scale</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.0218	.0005	.0005	0.02
Section Leader Performance Rating	.0294	.0009	.0004	0.01
Composite x Performance Rating	.1292	.0167	.0158	0.55

<u>Dependent Variable: Halo on the Behavioral Checklist</u>				
Variable	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Composite	.1554	.0242	.0242	0.89
Section Leader Performance Rating	.1683	.0283	.0042	0.15
Composite x Performance Rating	.1858	.0345	.0062	0.22

Note: Halo is measured by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format and calculating the mean of these SDs for each section leader.

n = 38

All F ratios are non-significant at p < .05.

Hypothesis 4D. A larger proportion of the variance in the measurement of halo in assigned ratings of subordinate job performance will be accounted for by rater organizational commitment than by job involvement.

The procedure used to test hypothesis 4D is the same as that used to examine the previous hypotheses addressing the issue of whether rater scores on the organizational commitment scale would explain more of the variance in the measure of each rating error than job involvement scores as suggested by the model. This analysis was again performed utilizing both sales clerks and section leaders as the units of analysis and the operational definition of halo based on the standard deviation of scores across rating dimensions. The power of these procedures was the same as that of the earlier reported analogous tests. The results of the analyses testing hypothesis 4D are shown in Tables 46 and 47.

Only the zero-order coefficient for the correlation between job involvement scores and the measure of halo shown in Table 46 was significant. None of the product-moment correlations for the analysis focusing on section leaders were sufficiently large to be statistically significant. With regard to the tests of hypothesis 4D, none of the z difference values determined to test the significance of the difference between the various pairs of coefficients in either table were found to be significant. This indicates that neither of the two attitude scales considered was able to better predict the likelihood

Table 46

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Halo on the Graphic Rating Scale and the Behavioral Checklist with Sales Clerks as the Unit of Analysis

Rating Format	<u>Job Involvement</u>	<u>Organizational Commitment</u>	<u>z_{diff}</u>
	<u>r</u>	<u>r</u>	
Graphic Rating Scale	-.161*	-.058	1.020
Behavioral Checklist	-.102	-.050	.514

Note: Halo is measured by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format.

n = 199

*p < .05

All r and z values are non-significant at p < .05.

Table 47

Comparison of Correlations Between Section Leader Attitude Scale Scores and the Measure of Halo on the Graphic Rating Scale and the Behavioral Checklist with Section Leaders as the Unit of Analysis

Rating Format	<u>Job Involvement</u>	<u>Organizational Commitment</u>	<u>z_{diff}</u>
	<u>r</u>	<u>r</u>	
Graphic Rating Scale	-.202	-.045	-.656
Behavioral Checklist	-.014	-.020	.025

Note: Halo is measured by determining the standard deviation of the ratings assigned to each ratee on the nine dimensions of each rating format and calculating the mean of these SDs for each section leader.

n = 38

All r and z values are non-significant at p < .05.

of assigned ratings having particular levels of halo. Hence, it is concluded, that hypothesis 4D was not supported.

The results of the tests of hypotheses 4A, 4B, 4C, and 4D are summarized in Table 48. The tests of the hypothesis 4A indicated that there is evidence to support the belief that rater scores on the composite attitude scale are directly related to the tendency to assign ratings characterized by varying amounts of halo but only on the trait-based rating format. The same was true for the job involvement measure. The strength of the relationship between the attitude measures and the amount of halo in assigned ratings did not differ significantly for the two rating formats although the graphic rating scale ratings consistently evidenced more halo than ratings made on the behavioral checklist. Rater ability did not moderate the relationship between attitude scores and the measure of halo. The difference between this relationship for organizational commitment and job involvement was also found not to be statistically significant.

Reliability

Hypothesis 5. A composite measure of rater job involvement and organizational commitment will be directly related to the reliability of ratings assigned to assess subordinate job performance.

While the construct of reliability is purely theoretical in nature, it constitutes a critical characteristic of any abstract measurement. The idea that observed scores are comprised of "true"

Table 48

Summary of Results for Tests of Halo Hypotheses 4A - 4D

Unit of Analysis	Attitude Measure	Rating Format	Hypothesis 4A		Hypothesis 4B	Hypothesis 4C	Hypothesis 4D
			Standard Deviation	Mean Correlation	Standard Deviation	Standard Deviation	Standard Deviation
Sales Clerks (ratees)	Composite JI + OC	GRS	S-	S-	N.S.	N.S.	
		BCL	N.S.				
	Job Involvement	GRS	S-		N.S.	N.S.	N.S.
		BCL	N.S.				N.S.
	Organization Commitment	GRS	N.S.		N.S.	N.S.	N.S.
		BCL	N.S.				N.S.
	Composite JI + OC	GRS			N.S.		
		BCL					
Section Leaders (raters)	Job Involvement	GRS			N.S.		N.S.
		BCL					N.S.
	Organization Commitment	GRS			N.S.		N.S.
		BCL					N.S.

Note: GRS = graphic rating scale. BCL = behavioral checklist. S+ indicates that the hypothesis was supported. S- indicates that a statistically significant result was found in the opposite direction predicted by the hypothesis. N.S. indicates that the hypothesis was not supported.

scores and scores attributable to random error is a fundamental assumption of classical test theory because only systematic score variance is predictable. Hence, the reliability of a measure represents an upper bound for validity because a test can not theoretically predict any other criterion better than it can predict itself (Guion, 1965).

Hypothesis 5 proposes that differences in the work-related attitudes of supervisors assigning subordinate performance ratings is an additional source of error variance which effects the reliability of theses ratings in a systematic fashion. Specifically, the reliability of the ratings assigned by supervisors who evidence relatively high levels of the combined measure of job involvement and organizational commitment is predicted to be significantly higher than the reliability of ratings assigned by raters with lower scores on the composite attitude measure. In other words, it is suggested that the proportion of total rating score variance due to variation in true scores as opposed to the proportion due to various sources of error variance will be greater for those raters with high scores on the combined measure of job involvement and organizational commitment than for raters with low scores on this measure.

Since there is no one absolute measure of the reliability of a set of scores, several different methods of estimating the reliability of the performance ratings assigned were utilized in this study. Each procedure focuses on estimating a different potential source of systematic error variance. Those methods for which it was feasible to determine reliability coefficients for the ratings assigned by section

leaders scoring high or low on the composite attitude measure were considered in testing hypothesis 5. Four sets of coefficients for corresponding rating dimensions were contrasted using Fisher's r to z transformation. The specific prediction being made was that the reliability coefficients for the ratings assigned by section leaders scoring above the median value of the distribution of the composite measure would be higher than the coefficients for the ratings assigned by raters scoring below the median. Comparisons were made for the alternate format, internal consistency (behavioral checklist only), and test-retest reliability estimation procedures. The interrater reliability procedure was not used in this analysis because attitude scale scores were not available for several of the section leaders who rated sales clerks during the second round of appraisals. The results of the comparisons of the dimensional reliability coefficients are presented in Table 49.

The results of tests of hypothesis 5 indicate that in every instance, the reliability coefficients for the rating dimensions as well as the overall scale scores did not differ significantly for the sales clerks rated by section leaders scoring above and below the median on the composite attitude scale. Therefore, it may be concluded that section leader work-related attitudes did not significantly effect the reliability of the performance ratings assigned to subordinates. However, it should be noted that in the overwhelming majority of cases, the coefficients for the ratings assigned by section leaders scoring below the median value tended to be higher than those assigned by raters scoring above the median although these

Table 49
Comparison of Reliability Coefficients for Sales Clerk Performance Ratings
Assigned by Section Leaders Scoring Above and Below the Median on the
Job Involvement - Organizational Commitment Composite Scale

Scale	Alternate Format Reliability GRS - BCL Correlations			Internal Consistency KR-20 Behavioral Checklist		
	Above Median	Below Median	\bar{z} diff	Above Median	Below Median	\bar{z} diff
	(\bar{n} = 98)	(\bar{n} = 101)		(\bar{n} = 98)	(\bar{n} = 101)	
Dependability	.55	.58	-0.31	.67	.78	-1.63
Job Procedures	.43	.40	0.25	.67	.72	-0.67
Sales Ability	.33	.52	-1.63	.61	.69	-0.97
Customer Service	.31	.49	-1.49	.58	.72	-1.70
Company Policy	.27	.36	-0.69	.71	.68	0.40
Product Knowledge	.51	.67	-1.72	.65	.69	-0.51
Initiative	.58	.61	-0.64	.68	.80	-1.88
Judgment	.63	.61	0.22	.76	.72	-0.61
Empl Relations	.57	.57	0.00	.63	.61	0.22
Total	.63	.73	-1.30	--	--	--

Scale	Test - Retest Reliability Graphic Rating Scale			Test - Retest Reliability Behavioral Checklist		
	Above Median	Below Median	\bar{z} diff	Above Median	Below Median	\bar{z} diff
	(\bar{n} = 25)	(\bar{n} = 28)		(\bar{n} = 25)	(\bar{n} = 28)	
Dependability	.80	.86	-1.34	.72	.77	-0.78
Job Procedures	.78	.79	-0.18	.69	.74	-0.71
Sales Ability	.83	.87	-1.01	.70	.73	-0.43
Customer Service	.78	.86	-1.72	.74	.70	0.58
Company Policy	.84	.80	0.84	.72	.81	-1.52
Product Knowledge	.79	.84	-1.04	.70	.71	-0.14
Initiative	.85	.89	-1.15	.76	.85	-1.81
Judgment	.86	.83	0.73	.73	.79	-0.99
Empl Relations	.80	.87	-1.63	.74	.70	0.58
Total	.81	.85	-1.09	.72	.76	-0.61

Note: The interval between ratings correlated in determining the test - retest reliability estimates was ten months. Only sales clerks rated by the same section leaders on both occasions were included in this analysis.

differences did not reach significance when assessed with a two-tailed testing procedure. The directional trend of these differences was again in the opposite direction predicted by the model.

Classical test theory makes the assumption that for every set of specific conditions there is a single true score which represents the accurate assessment of the quantity in question under those conditions. In contrast, the theory of generalizability (Cronbach, Rajaratnam, and Gleser, 1963) makes reference to a "universe" of conditions to which we may wish to generalize from a specific set of conditions. Generalizability theory conceptualizes a true score as a "universe score" which represents the mean of all observations across all of the specified conditions in the universe. Conditions in this sense may refer to particular test items, test forms, observers, occasions, or situations of observation, etc.

According to generalizability theory, an interest in the reliability of a given measure is based on a desire to generalize from a specific set of observations to some defined class of observations. A generalizability study is specifically designed to assess a measure in terms of the relationship between observed scores and the universe scores to which we wish to generalize. This requires an explicit definition of the universe being considered. Error in measurement is conceptualized as the difference between the observed score and the universe score. The generalizability of a measure is conceived of as the ratio of the variance of the universe scores to the variance of the observed scores. The factors which contribute to the variance of observed scores will vary according to the definition of the universe.

This feature of generalizability theory permits simultaneous consideration of the components of error variance stemming from two or more conditions or facets of the universe of interest. Hence, the various traditional coefficients of reliability (equivalence, internal consistency, stability) can be subsumed under the concept of generalizability by defining the universes of interest accordingly.

A generalizability study was conducted to contrast the generalizability of the ratings assigned by the two groups of section leaders across sales clerks, rating formats, and rating occasions. The sales clerk sample was again dichotomized on the basis of the composite attitude scale scores of the section leaders. Hypothesis 5 was tested by comparing the corresponding generalizability coefficients for the two groups. The prediction made was that ratings assigned by section leaders scoring above the median on the composite measure would evidence greater generalizability than the ratings given by the supervisors who scored below the median.

Only those sales clerks who were evaluated by the same section leaders on both occasions were included in this analysis. This was done to ensure that the assessments of each clerk on both occasions were made by section leaders who were either above or below the median on the composite measure. Generalizability coefficients were determined for both samples for the situations in which each clerk's total rating score was: a) the mean score from two rating formats - graphic rating scale and behavioral checklist standardized, on two rating occasions (i.e., the mean of four scores), b) the mean score from one rating format on two occasions, c) the mean score from two rating

formats on one occasion, and d) the score assigned on a single format on a single occasion. The estimated variance components and the generalizability coefficients for the sales clerks rated by section leaders scoring above the median are shown in Table 50. The corresponding statistics for the group appraised by section leaders scoring below the median are shown in Table 51.

The significance of the difference between corresponding generalizability coefficients for the two groups was tested by means of Fisher's \underline{r} to \underline{z} transformation. None of the \underline{z} difference statistics were significant at $p < .05$. This finding coincides with the earlier reported tests of hypothesis 5 and indicates that the job involvement and organizational commitment levels of the section leaders did not systematically influence the reliability of the job performance ratings which they assigned to their subordinates. Accordingly, it is concluded that the results of the data analyses did not support hypothesis 5.

Convergent and Discriminant Validity

Hypothesis 6. A composite measure of rater job involvement and organizational commitment will be directly related to the levels of convergent and discriminant validity in assigned ratings of subordinate job performance.

As mentioned earlier, the multitrait-multimethod scheme proposed by Campbell and Fiske (1959) is generally acknowledged as one of the

Table 50
Generalizability Study for Sales Clerks Rated by Section Leaders Scoring Above the Median on
the Composite Attitude Scale

Source of Variance	df	Mean Squares	Est. Variance Component	G Study Sit. A: $n_r=2; n_f=2$	D Study Sit. B: $n_r=1; n_f=2$	D Study Sit. C: $n_r=2; n_f=1$	D Study Sit. D: $n_r=1; n_f=1$
				Exp. Observed Score Var. Comp	Exp. Observed Score Var. Comp.	Exp. Observed Score Var. Comp	Exp. Observed Score Var. Comp.
Sales Clerks (c)	24	2.362	$s_c^2 = 1.992$	$Es_c^2 = 1.992$	$Es_c^2 = 1.992$	$Es_c^2 = 1.992$	$Es_c^2 = 1.992$
Ratings (r)	1	0.148	$s_r^2 = .005$	$Es_r^2 = .003$	$Es_r^2 = .005$	$Es_r^2 = .003$	$Es_r^2 = .005$
Rating Formats (f)	1	0.013	$s_f^2 = .019$	$Es_f^2 = .010$	$Es_f^2 = .010$	$Es_f^2 = .019$	$Es_f^2 = .019$
Clks x Ratings (cr)	24	0.134	$s_{cr}^2 = .035$	$Es_{cr}^2 = .018$	$Es_{cr}^2 = .035$	$Es_{cr}^2 = .018$	$Es_{cr}^2 = .035$
Clks x Formats (cf)	24	0.335	$s_{cf}^2 = .230$	$Es_{cf}^2 = .118$	$Es_{cf}^2 = .118$	$Es_{cf}^2 = .236$	$Es_{cf}^2 = .236$
Rtgs x Formats (rf)	1	0.245	$s_{rf}^2 = .006$	$Es_{rf}^2 = .001$	$Es_{rf}^2 = .003$	$Es_{rf}^2 = .003$	$Es_{rf}^2 = .006$
Residual (crf,e)	24	0.099	$s_{res}^2 = .099$	$Es_{res}^2 = .025$	$Es_{res}^2 = .050$	$Es_{res}^2 = .050$	$Es_{res}^2 = .099$
				2.167	2.213	2.321	2.392
Situation A (two ratings, two formats) Generalizability Coefficient:				$E_p^2 = \frac{1.992}{2.167} = .919$			
Situation B (one rating, two formats) Generalizability Coefficient:				$E_p^2 = \frac{1.992}{2.213} = .900$			
Situation C (two ratings, one format) Generalizability Coefficient:				$E_p^2 = \frac{1.992}{2.321} = .858$			
Situation D (one rating, one format) Generalizability Coefficient:				$E_p^2 = \frac{1.992}{2.392} = .833$			

Table 51
Generalizability Study for Sales Clerks Rated by Section Leaders Scoring Below the Median on
the Composite Attitude Scale

Source of Variance	df	Mean Squares	Est. Variance Component	G Study Sit. A: $n_r=2; n_f=2$	D Study Sit. B: $n_r=1; n_f=2$	D Study Sit. C: $n_r=2; n_f=1$	D Study Sit. D: $n_r=1; n_f=1$
				Exp. Observed Score Var. Comp	Exp. Observed Score Var. Comp.	Exp. Observed Score Var. Comp	Exp. Observed Score Var. Comp.
Sales Clerks (c)	24	3.721	$s_c^2 = 3.188$	$Es_c^2 = 3.188$	$Es_c^2 = 3.188$	$Es_c^2 = 3.188$	$Es_c^2 = 3.188$
Ratings (r)	1	0.155	$s_r^2 = .042$	$Es_r^2 = .021$	$Es_r^2 = .042$	$Es_r^2 = .021$	$Es_r^2 = .042$
Rating Formats (f)	1	0.426	$s_f^2 = .036$	$Es_f^2 = .018$	$Es_f^2 = .018$	$Es_f^2 = .036$	$Es_f^2 = .036$
Clks x Ratings (cr)	24	0.281	$s_{cr}^2 = .131$	$Es_{cr}^2 = .066$	$Es_{cr}^2 = .131$	$Es_{cr}^2 = .066$	$Es_{cr}^2 = .131$
Clks x Formats (cf)	24	0.402	$s_{cf}^2 = .252$	$Es_{cf}^2 = .126$	$Es_{cf}^2 = .126$	$Es_{cf}^2 = .252$	$Es_{cf}^2 = .252$
Rtgs x Formats (rf)	1	0.047	$s_{rf}^2 = .037$	$Es_{rf}^2 = .009$	$Es_{rf}^2 = .018$	$Es_{rf}^2 = .018$	$Es_{rf}^2 = .037$
Residual (crf,e)	24	0.150	$s_{res}^2 = .150$	$Es_{res}^2 = .038$	$Es_{res}^2 = .075$	$Es_{res}^2 = .075$	$Es_{res}^2 = .150$
				3.466	3.598	3.656	3.836
Situation A (two ratings, two formats) Generalizability Coefficient:				$E_p^2 = \frac{3.188}{3.466} = .920$			
Situation B (one rating, two formats) Generalizability Coefficient:				$E_p^2 = \frac{3.188}{3.598} = .886$			
Situation C (two ratings, one format) Generalizability Coefficient:				$E_p^2 = \frac{3.188}{3.656} = .872$			
Situation D (one rating, one format) Generalizability Coefficient:				$E_p^2 = \frac{3.188}{3.836} = .831$			

best approaches for operationalizing the assessment of construct validity through the empirical demonstration of convergent and discriminant validity. Campbell and Fiske propose that when ratings of the same traits from different sources (i.e., the validity diagonals) are significantly different from zero, this constitutes empirical evidence of convergent validity. Claims for the existence of discriminant validity are dependent upon three different sources of support. First, the validity diagonal coefficients should exceed coefficients for which the ratings correlated are not of the same trait or from the same source. Second, the validity diagonal coefficients for correlations between variables measuring the same trait should exceed the correlations between different traits measured by the same method (source). Third, the patterns of intercorrelations among different traits should be consistent regardless of the method or source of the ratings.

While use of the criteria proposed to establish the degree of convergent and discriminant validity of a set of ratings is theoretically appropriate, in practice determining the extent to which these conditions are present is at best inferential and often extremely difficult particularly with large matrices. An alternative approach suggested by Stanley (1961) is to conceptualize the matrix as a three-way classification model and analyze the multitrait-multimethod data using analysis of variance. If this approach is applied to the data under consideration in this study and sales clerks are considered as random with traits and rating formats as fixed, the statistical model representing this situation may be represented as:

$$Y_{ijk} = u + a_i + b_j + c_k + (ab)_{ij} + (ac)_{ik} + (bc)_{jk} + e_{ijk}$$

where:

Y_{ijk} = rating of sales clerks for traits (dimensions) by sources,

a_i = effect of sales clerk $i = 1, 2, 3, \dots, n_i$,

b_j = effect of dimension $j = 1, 2, 3, \dots, 9$,

c_k = effect of source (rating format) $k = 1, 2$,

e_{ijk} = error.

For this analysis, the effects of primary interest involve only four of the sources of variance from the above model. These sources are as follows: 1) variance due to sales clerks indicating the extent of agreement or convergent validity on ratees across rating formats, 2) sales clerk by dimension variance indicating the degree of rating discrimination on traits by sales clerks or discriminant validity, 3) sales clerk by rating format variance indicating the amount of source bias or halo in the ratings, and 4) error. By considering ratees as rows in the ratee-source-trait matrix, the repeated assessments across ratees make three sources of covariance possible: 1) within each source (method) across traits, 2) within each trait across sources, and 3) across both sources and traits. Hence, by using the analysis of variance model shown above to describe multitrait-multimethod rating data, it is possible to examine halo, discriminant validity, and convergent validity which correspond to the three sources of covariation listed.

Kavanagh et al. (1971) has suggested that application of the analysis of variance technique to the study of rating data presented in a multitrait-multimethod matrix has four principle advantages over intuitive examination of the data as suggested by Campbell and Fiske: 1) the analysis of variance method provides a highly efficient procedure for summarizing and evaluating the extent of convergent and discriminant validity, particularly in instances where the number of traits and/or sources is large, 2) this method enables researchers to quantify the degree of validity present in the ratings thus making the assessment of this information less subjective, 3) method bias and the prevalence of sampling variance in the data may be estimated, and 4) the relative strengths of the effects can be directly determined.

Procedures for computing estimates of the mean squares and variance components of the analysis of variance model using average correlations from blocks of coefficients from the multitrait-multimethod matrix have been developed by Wolins (1964); Zyanski (1962); and Boruch, Larkin, Wolins, and McKinney (1970). Since this technique has not been widely used in published research reports, the computational procedures used in this analysis are summarized in Appendix F. The variance components determined in this manner are useful in making within study comparisons. They make it possible to make inferences regarding the importance of experimental effects relative to the percentage of error variance while controlling for the influence of sample size.

In order to test hypothesis 6, it is necessary to be able to make comparisons between different matrices in terms of the relative

amounts of convergent and discriminant validity. Kavanagh et al. (1961) suggest that this may be accomplished by comparing each variance component of interest relative to the size of the error variance. Dividing a variance component by the sum of component itself and the error variance component yields an intraclass correlation. The indexes derived in this manner indicate the amount of convergent and discriminant validity as well as method bias or halo in a multitrait-multimethod matrix in a format which is comparable to corresponding indexes from other matrices. These intraclass correlations are distributed approximately the same as Pearson product-moment correlations. Therefore, corresponding pairs of coefficients from different matrices can be converted with Fisher's r to z transformation and the significance of the difference between these values can then be determined. This procedure was used to test Hypothesis 6.

Hypothesis 6 was examined by dichotomizing the sales clerk sample on the basis of whether the section leader who appraised each clerk scored above or below the median of the score distribution for the combined job involvement - organizational commitment variable. Multitrait-multimethod matrices were derived from the rating data for these two groups of employees. These matrices have previously been presented in Tables 39 and 40. Using the correlations shown in these tables, the analysis of variance procedure was applied. The summarized results of this procedure along with the corresponding variance components for the effects shown are presented in Table 52. Only the statistics for effects pertinent to the testing hypothesis 6 and the assessment of halo in the ratings are reported in this table.

Table 52

Analysis of Variance of Correlations from the Multitrait-
Multimethod Matrices for Sales Clerks Rated by Section Leaders
Scoring Above or Below the Median on the Composite Attitude Scale

Sales Clerks Rated by Section Leaders Scoring Above the Median on the Composite Attitude Measure (\underline{n} = 98)			
Source	df	Mean Square	\underline{F}
Sales Clerks (SC)	97	8.127	42.02**
SC x Dimension (D)	776	.648	3.35**
SC x Rating Format (F)	97	3.328	17.23**
Error (E)	776	.193	
Sales Clerks Rated by Section Leaders Scoring Below the Median on the Composite Attitude Measure (\underline{n} = 101)			
Source	df	Mean Square	\underline{F}
Sales Clerks (SC)	100	8.150	23.01**
SC x Dimension (D)	800	.709	2.00**
SC x Rating Format (F)	100	1.348	3.81**
Error (E)	800	.354	

** \underline{p} < .01

The findings indicate that tests of the main effects and interactions for both groups were all significant.

The results of the analysis of variance procedure provide evidence of substantial agreement across rating formats on individual dimensions thus indicating the presence of convergent validity in both sets of ratings. This is evident from the relative magnitude of variance attributable to the main effect for Sales Clerks (see Table 53). The Sales Clerk x Dimension interaction provides an indication of the extent to which sales clerks were ordered differently on the various rating scales. However, this effect is approximately one quarter the size of the Sales Clerk main effect for ratings assigned by section leaders scoring above the median on the composite attitude measure. It is even less for the ratings provided by section leaders scoring below the median. These results indicate the presence of very little discriminant validity in the ratings. Finally, the magnitude of the Sales Clerk x Rating Format effect for the ratings assigned by the high scoring section leaders indicates that there is a considerable amount of method bias or halo confounding the results for this group. The influence of this factor is considerably less for the ratings provided by the lower scoring section leaders.

Thus, while the evidence for discriminant validity in the ratings is limited, there is considerable support for convergent validity although this effect is somewhat confounded by the substantial method bias or halo indicated in the ratings, particularly those provided by the raters indicating relatively high levels of job involvement and organizational commitment. In addition, the size of the variance

Table 53

Comparison of Variance Components and Indexes from the Multitrait-Multimethod Matrices for Section Leaders Scoring Above and Below the Median on the Composite Attitude Scale

<u>Variance Components</u>			
Source	Above Median	Below Median	
Sales Clerks (Convergent Validity)	.477	.499	
Sales Clerks x Dimension (Discriminant Validity)	.120	.078	
Sales Clerks x Rating Scale (Halo)	.240	.075	
Error	.296	.394	
<u>Variance Indexes</u> (Intraclass Correlations)			
Source	Above Median	Below Median	<u>z</u> _{diff}
Sales Clerks (Convergent Validity)	.720	.630	.625
Sales Clerks x Dimension (Discriminant Validity)	.297	.167	.903
Sales Clerks x Rating Scale (Halo)	.481	.161	2.22*

* $p < .05$

components for error indicates that the ratings are influenced by other unidentified sources of variance to a considerable extent.

A comparison the the variance component indexes for the two matrices was performed to determine the extent to which rater work-related attitudes may have contributed to the error variance effecting the ratings. The indexes for the ratings assigned by the two groups of section leaders are shown in the lower portion of Table 53. Comparisons of the corresponding indexes for the three effects indicated that there were no significant differences in the amounts of convergent or discriminant validity evidenced in the ratings made by the two groups of raters. Only for the method bias effect indicating the relative amount of halo present was a significant difference noted. In this instance, the section leaders scoring above the median on the composite measure showed significantly more halo error in the ratings of their subordinates than did the raters with scores below the median. This finding supports the previously reported results of the tests of hypothesis 4A. The estimated power of the test of the differences between the pairs of intraclass correlations was .55.

In summary, the procedures conducted to test hypothesis 6 provided evidence of substantial convergent validity in the performance ratings assigned to the sales clerks in the sample by both groups of raters. The evidence for discriminant validity was rather discouraging. The relatively small variance component for the Sales Clerk x Dimension effect indicates that there was little differentiation between subjects on the different rating scales within each format. This would seem to imply that the number of dimensions rated

could be reduced. In spite of the large error variance component, it does not appear that rater work-related attitudes contributed substantially to this factor. This is shown by the absence of a significant difference in the amounts of convergent and discriminant validity in the ratings supplied by section leaders scoring above and below the median of the distribution of the composite measure.

Hence, it must be concluded that hypothesis 6 was not supported. The only difference between the two groups indicated by this analysis was in terms of the extent of method bias in the ratings. Once again, the results in this regard proved contrary to the proposed model in that individuals with relatively low levels of job involvement and organizational commitment provided psychometrically superior ratings of subordinate performance in terms of the relative amount of halo error present.

In addition to the statistical analyses performed to test the hypotheses derived from the model, two supplemental analyses were performed on an exploratory basis. First, certain demographic variables for the section leaders in the sample considered to be potential antecedents of job involvement and organizational commitment were correlated with scores on the scales measuring these attitudes as well as the composite measure. The factors considered were age, sex, educational level, tenure, and ability as measured by the section leader performance rating scale. The product-moment correlation coefficients resulting from this procedure are presented in Table 54. Employee age and tenure were significantly correlated with all three of the cognitive measures in a direct fashion. In contrast,

Table 54
Correlation of Section Leader Demographic Variables with
Work-Related Attitudes

Variable	Composite (JI + OC)	Job Involvement	Organizational Commitment
Age	.35*	.33*	.39**
Sex (1 = female 2 = male)	-.09	-.13	-.18
Educational Level (years completed)	-.27*	-.21	-.24*
Tenure (years)	.29*	.27*	.33*
Ability (GRS total)	.19	.21	.15

n = 38

*p < .05

**p < .01

educational level was found to be inversely related to organizational commitment as well as the composite measure. Sex and job performance ratings were not significantly related to any of the attitude measures considered.

In an attempt to determine if there were any systematic differences in the rating strategies employed by raters with high as opposed to low levels of the composite measure of job involvement and organizational commitment, a policy capturing analysis was performed. In this procedure, ratees were again dichotomized on the basis of whether their section leaders' scores on this variable were above or below the median of the score distribution. Regression analyses were performed in which the dimension ratings from both the graphic rating scale and the behavioral checklist were separately regressed upon the single item appraisal of overall job performance. Comparisons were then made between the groups on the basis of the relative importance of the various dimensions in accounting for the variance in the overall ratings (see Zedeck & Kafry, 1977). The results for the ratings made on the graphic rating scale and behavioral checklist are presented in Tables 55 and 56, respectively.

Appraisals made on the graphic rating scale dimensions of initiative and knowledge of job procedures accounted for the greatest portion of variance in the overall performance ratings. Although the order of entry into the equation was reversed for those raters scoring above and below the median, it is apparent that both groups clearly put the most emphasis on these graphic rating scale dimensions in terms of determining an assessment of each ratee's overall job

Table 55

Relative Importance of Graphic Rating Scale Dimension Scores for Predicting Overall Performance Assessments Assigned by Section Leaders Scoring Above and Below the Median on the Composite Attitude Scale

<u>Above the Median (n = 98)</u>				
Dimension	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Initiative	.6321	.3995	.3995	63.871**
Job Procedures	.7120	.5070	.1075	20.713**
Dependability	.7209	.5198	.0128	2.469
Judgment	.7270	.5285	.0088	1.727
Customer Relations	.7303	.5334	.0048	0.955
Employee Relations	.7345	.5395	.0061	1.205
Company Policy	.7362	.5421	.0026	0.511
Product Knowledge	.7368	.5428	.0008	0.146
Sales Ability	.7369	.5430	.0002	0.034
<u>Below the Median (n = 101)</u>				
Dimension	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Job Procedures	.6384	.4076	.4076	68.109**
Initiative	.6826	.4660	.0584	10.712**
Sales Ability	.7032	.4945	.5758	5.468*
Employee Relations	.7099	.5039	.0095	1.835
Dependability	.7153	.5116	.0077	1.492
Product Knowledge	.7224	.5219	.0103	2.014
Company Policy	.7243	.5246	.0028	0.539
Customer Relations	.7256	.5264	.0018	0.357
Judgment	.7257	.5267	.0002	0.044

* $p < .05$ ** $p < .01$

Table 56

Relative Importance of Behavioral Checklist Dimension Scores for Predicting Overall Performance Assessments Assigned by Section Leaders Scoring Above and Below the Median on the Composite Attitude Scale

<u>Above the Median (n = 98)</u>				
Dimension	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Dependability	.5135	.2637	.2637	34.378**
Initiative	.6180	.3819	.1183	18.177**
Employee Relations	.6655	.4429	.0609	10.282**
Job Procedures	.6859	.4705	.0276	4.844*
Product Knowledge	.6888	.4745	.0040	0.702
Sales Ability	.6892	.4751	.0006	0.103
Company Policy	.6894	.4754	.0003	0.053
Customer Relations	.6895	.4755	.0001	0.010
Judgment	.6895	.4755	.0000	0.005
<u>Below the Median (n = 101)</u>				
Dimension	Multiple <u>R</u>	<u>R</u> ²	<u>R</u> ² Change	<u>F</u>
Judgment	.6277	.3940	.3940	64.368**
Job Procedures	.7119	.5068	.1128	22.409**
Employee Relations	.7306	.5338	.0270	5.619*
Product Knowledge	.7480	.5594	.0257	5.589*
Dependability	.7512	.5644	.0049	1.073
Customer Relations	.7530	.5670	.0026	0.564
Initiative	.7538	.5682	.0012	0.263
Company Policy	.7543	.5690	.0008	0.174
Sales Ability	.7543	.5690	.0000	0.001
* <u>p</u> < .05 ** <u>p</u> < .01				

performance. In contrast, the behavioral checklist scales which contributed most heavily to explaining variance in the single item measure for raters scoring above the composite median were dependability, initiative, employee relations, and knowledge of job procedures. For the below median group, the most salient dimensions were judgment, knowledge of job procedures, employee relations, and product knowledge. Once again, although there were some specific differences between the two groups, a fair degree of correspondence in terms of the relative importance of the various dimensions in this context is evident. These results should be considered exploratory as they would have to be cross validated before any firm conclusions regarding the similarities or differences in the strategies utilized by raters in each group could be drawn.

CHAPTER VI

DISCUSSION

Discussion of Results

The results of the tests of the theoretical model proposed are in one sense disappointing while in another, highly intriguing. Taken as a whole, the experimental findings of this study indicate a counter-intuitive phenomenon. Specifically, raters who exhibited high levels of job involvement and organizational commitment were found in general, to assign subordinate performance ratings characterized by greater degrees of leniency, central tendency, range restriction, and halo than the ratings assigned by raters evidencing lower levels of these attitudes. Although contrary to the predictions of the model, the relationship between the measured levels of these attitudes and the prevalence of the rating errors considered was notably consistent irrespective of the particular procedures used to operationalize the various effects. The search for possible explanations for this surprising turn of events requires an in-depth reconsideration of the theoretical model and the assumptions upon which it is based. However, as the possible explanations for the results of the tests of the hypotheses examined are related, it will be advantageous to first briefly summarize all of the research findings.

Hypotheses 1A, 2A, 3A, and 4A examined the central premise of the model. The results for the job involvement - organizational commitment composite uniformly showed that this measure was directly

related to the amount of leniency, central tendency, and halo in ratings assigned and inversely related to the amount of restriction of range. These relationships were evident for both of the graphic rating scale and behavioral checklist formats. The only exception was the failure of the association between the composite measure and the amount of halo evidenced on the behavioral checklist to reach a level of significance at $p < .05$. These findings are particularly impressive in light of the fact that corresponding results were found regardless of the operational definitions of the four rating errors used and the relatively weak power of some of the statistical procedures employed as necessitated by limitations inherent in the experimental design.

The results for the examination of the association between the separate measures of job involvement and organizational commitment and the prevalence of systematic bias in the ratings also showed a considerable amount of consistency. The relationship of job involvement and the four rating errors followed the same pattern found for the composite measure but only for ratings made on the graphic rating scale. None of the tests of the correspondence of job involvement scores with the occurrence of rating errors made on the behavioral checklist were significant except in the case of leniency. In contrast, rater levels of organizational commitment were found to be significantly related to the prevalence of each rating error except halo in the opposite direction suggested by the model but only for ratings made on the behavioral checklist.

None of the tests of hypotheses 1B, 2B, 3B, or 4B for the rating errors considered yielded significant test results. Hence, it may be concluded that the relationship between scores on the composite measure as well as the two component attitude scales and the degree of each rating error measured was not significantly stronger for either the graphic rating scale or the behavioral checklist.

The basic premise of hypotheses 1C, 2C, 3C, and 4C is that rater ability will moderate the relationship between the composite attitude measure and the amount of each systematic source of bias in assigned ratings of subordinate performance. This conjecture was supported by the results of the tests for leniency and central tendency but only for ratings made on the graphic rating scale. The analyses using both ratees and raters as the experimental units of analysis consistently provided evidence to reinforce the effect of ability as a moderator variable in this context. None of the tests for restriction of range or halo supported this hypothesis. While moderating effect of rater ability for leniency and central tendency strengthened the relationship between rater attitudes and these rating errors, these relationships were still in the opposite direction suggested by the model.

The results from the investigations of hypotheses 1D, 2D, 3D, and 4D were again consistently disappointing in that no significant differences between the proportions of variance in the four rating errors accounted for by the individual measures of rater job involvement and organizational commitment were found. This implies that if in fact these attitudes do influence the behavior of appraising the

job performance of subordinates, they do not have a differential effect on the portion of the process devoted to the assignment of ratings.

Finally, the tests of hypothesis 5 indicate that the level of a rater's combined measure of job involvement and organizational commitment is not systematically related to the reliability of the ratings assigned by that individual. A similar conclusion may be drawn for hypothesis 6 regarding the lack of significant relationships between the composite attitude measure and the degree of convergent and discriminant validity evidenced in the multidimensional assessments of subordinate job performance.

In reporting the results of any experimental investigation, it is always gratified when the data support the researcher's theoretical predictions. However, the process of theory building and subsequent testing is an iterative process where as much may be learned from unsupported hypotheses as from others which are confirmed.

The search for explanations of why the results of the experimental procedures performed in this study did not coincide with the predictions of the model may be thought of as a continuum. At one end is the assessment that the model and the assumptions upon which it rests are flawed. At the other end is the conclusion that the experimental procedures used to test the model were inadequate to provide a true evaluation of the model's explanatory power. The truth is most likely a combination of these factors. This conceptual approach will provide a structure for the discussion which follows. That the results of this study generally did not support the proposed

hypotheses indicates that the model must be reconsidered. However in view of the body of published research which supports the premises upon which the model is based, it may be advantageous to first consider the characteristics of the experimental conditions in this study which might have contributed to the unexpected results.

A number of plausible explanations may be suggested to account for the contrary experimental results reported in this study regarding the inverse relationship between the composite measure of job involvement and organizational commitment and the prevalence of systematic sources of bias in assigned performance ratings. First, as alluded to in the review of the literature, Fishbein (1967) has specified that behavioral intentions may only be expected to predict behaviors when the specificity level of both factors are equivalent and the appropriate category of behavioral criterion is considered. In the proposed model, the attitude of job involvement as it applies to the act of rating the job performance of others is clearly an attitude towards a behavior. As such, the theory would expect it to be predictive of a repeated observation criteria or the same behavior directed at different targets in varied contexts and at different times. Under typical organizational circumstances where the rating of subordinate performance is a commonplace part of a supervisor's job, this activity could clearly be seen as meeting the requirements for being considered a repeated observation criteria. However, in the present study, the supervisors had never before been required to complete formal evaluations of their sales clerks. Under such circumstances, the act of completing rating forms for each subordinate one time only could be

construed as a single act criterion defined as a highly specific behavior directed toward a defined target object at a particular place and time. Research conducted to test Fishbein's model suggests that an attitude toward a behavior may or may not predict a single act criterion. Inconsistencies of this nature have been suggested as a likely reason for the inability of attitudes to predict subsequent behaviors in many research investigations and this may have been the case in the present study.

A second potential problem with the model has to do with the conceptualization of organizational commitment as an index of the subjective norm. According to Wiener's (1982) model, organizational commitment represents the internalization of normative pressures to exert effort on the part of the organization and accept the goals and values of the organization as one's own. An important premise of the procedures followed to test the theoretical model was that through participation in the development of the rating instruments and indoctrination during the training program, section leaders would come to accept that providing accurate and error-free ratings was critical to the success of the project and an extremely important objective of the organization. No effective manipulation check was employed to determine the degree to which they actually accepted this concept. Since the organization had never previously supported a formal evaluation program, it is questionable as to whether the raters really saw this goal as salient to the organization's well being. Furthermore, the idea of possibly rating some employees as poor performers or even merely not as good as others may have been antithetical to the

existing organizational value of only employing "good" workers. If this normative belief was strongly ingrained, it would not be surprising that highly committed raters would tend to assign ratings that tended to be relatively lenient and intercorrelated as well as restricted to a range near the midpoint of the scales and away from the lower end.

The use of the job involvement scale as an index of how likely a section leader was to accurately observe subordinate job performance and then make unbiased ratings in an effort to enhance inner feelings of self-esteem must also be questioned. While this premise is reasonable in organizations where conducting meaningful performance appraisals is an acknowledged part of a supervisor's job, this was not the case for the company in this study. The raters were informed that the appraisal program was being conducted for research purposes only and they were given no indication that it would become a regular part of their jobs. Hence, despite the fact that the importance of the assigning appropriate ratings stressed throughout the project, they might not have taken it seriously enough for performing the task well to have much of a positive influence on their self-image. More so, it is likely that having all "good" employees who were rated highly (or at least not poorly) would tend to be a more direct reflection of their abilities as supervisors and hence more aggrandizing to their self-esteem. It is possible that the existence of such a conflicting performance - self esteem contingency might have obfuscated the relationship between providing psychometrically superior ratings and enhanced feelings of self-worth assumed by the model.

Finally the effect of the rater training must be considered. An substantial number of recent studies have focused upon the relative efficacies of different approaches to rater training (e.g., Fay & Latham, 1982; Lee, 1985). The comparative effectiveness of various forms of rater training programs for reducing systematic rating errors and increasing accuracy is still subject to considerable debate. Fay et al. (1982) suggest that differences in the conclusions drawn regarding the usefulness of similar rater training programs may be due to variations in the motivation levels of the trainees, the effectiveness of the trainers, and the length of the training programs. Regarding the latter of these factors, longer training programs generally seem to be more successful in eliminating rating errors. For example, Wexley et al. (1975) required two hours of training to effectively eliminate one rating error while Latham et al. (1975) and Pursell, Dossett, & Latham (1980) provided raters with six to eight hours of training to minimize four rating errors. In other studies where training was not found to be effective, raters received only two or three hours of training to reduce three to four rating errors (Warmke & Billings, 1979; Borman, 1979b). Rating errors apparently are the result of externally mediated factors or deeply ingrained behavioral dispositions even for people who have never before completed an appraisal form. In view of the results of studies of various training programs cited, it is highly questionable as to whether the training provided in this study was adequate to effectively reduce all of the rating errors addressed. This is an important factor in testing the hypotheses as the model suggests that

1) training is important for establishing the subjective norm of assigning error-free ratings, and 2) training will influence the moderating effects of rater ability as well as role clarity by providing raters with the information they need to reduce systematic bias in their subordinate evaluations.

Each of the factors discussed thus far poses a possible explanation for why the experimental results did not support the predictions of the model. However, the findings of the study indicate that a consistent direct relationship between rater attitudes and the prevalence of rating errors was present in the data and this requires some explanation. Before addressing this topic, a comment is in order regarding the magnitude of the significant experimental effects found and the impact of the different approaches used operationalize the measurement of rating errors.

First, in most of the instances where the statistical analyses applied yielded significant results, the relative size of these effects was rather small as evidenced by the η^2 and R^2 values reported. The question arises as to whether these effects are too small to be meaningful. The answer to this question is probably not. First of all, in studying the potential influence of rater attitudes on the characteristics of the ratings they assign, it is assumed that other factors such as rater performance and rating context factors as well as other characteristics of the ratees and raters themselves are going to account for a major portion of the variance of any statistic derived from the ratings. Hence, rater attitudes were not expected to have a proportionately large effect on the psychometric qualities of

assigned ratings. Second, it may be that the relationships pursued in this investigation are only of this small magnitude. This in itself may be a mixed blessing as rater work-related attitudes, or indeed any extraneous factor outside of ratee job performance, should ideally have as little influence on the nature of the ratings as possible. Finally, the process of progressing from theoretical models to subject manipulation and actual measurement is invariably accompanied by the introduction of a certain amount of "noise". This commonly takes the form of unreliable measures and a lack of correspondence between the construct and how it is operationalized. These factors can have a severe attenuating effect on even strong relationships between constructs. Hence, for these reasons, particularly in areas where little or no previous research has been reported, the potential importance of even relatively weak but significant experimental effects deserves consideration.

Second, in contrast to the conclusion expressed by Saal et al. (1980) regarding the different results yielded by various procedures for operationally measuring rating errors, where multiple measures were used in this study, the results were almost uniformly consistent. This finding raises the question of whether the variance in the measurement of each error with different operational definitions as reported in the studies cited by Saal et al. was due to actual differences in what was being assessed or merely the result of statistical artifacts.

Some of the possible explanations for the significant direct correspondence of rater job involvement and organizational commitment

with the presence of the rating errors considered relate to the factors discussed above. For example, if the established corporate subjective norm was that only "good" employees work for the company and that they should be rated accordingly, it would follow that highly committed supervisors would accept this tenet to a greater degree than others who are less committed to the organization. This in turn might explain why section leaders who expressed a relatively low degree of a commitment to the company were less inclined to only assign lenient ratings within a relatively restricted and centrally located range. Their observations of subordinate job performance might have also been influenced to a lesser degree by the impression that all of their employees are "good" employees.

Similarly with respect to job involvement, section leaders who indicated comparatively low levels of job involvement might have been less inclined to derive feelings of personal self worth on the basis of the job performance of their subordinates. Accordingly, these individuals would have less to lose by rating their sales clerks more objectively which in turn could have resulted in a reduction of systematic bias.

While these are tenable explanations for the results found, it is important to also consider the possibility that the attitudinal factors focused upon might only be correlates of other rater characteristics which are more directly responsible for the relative presence or absence of the rating errors studied. The results from the supplemental correlational analysis performed indicate that as reported in other studies related to this issue (e.g., Rabinowitz &

Hall, 1977; Steers, 1977), age and tenure correlated positively with the attitude measures considered. Educational level was inversely related to organizational commitment at a significant level. This coincides with the commonly reported finding (e.g., Bateman & Strasser, 1984) that organizational commitment is inversely related to an individual's perception of the availability of alternative sources of employment in other organizations. It stands to reason the more highly educated individuals would perceive a greater variety of employment alternatives. Also, section leader performance ratings were not found to correlate significantly with any of the attitudinal measures. On the basis of these results, a profile emerges of the more highly job involved and organizationally committed section leaders as being older, more tenured, and less educated than their less involved and committed counterparts. Evidence from reported studies addressing the question of whether these factors are related to the quality of assigned ratings of subordinate job performance is limited and inconclusive.

The possibility that the performance ratings assigned accurately reflected the performance characteristics of the sales clerks evaluated must also be considered. The organization in which this study was conducted typically experiences 30% annual turnover among sales personnel. It is possible that the more tenured supervisors evidencing high levels of job involvement and organizational commitment might have been able over time to attract and retain a superior work group within their departments. These selected sales clerks might actually have been better employees whose performance across the

various evaluated domains was generally consistent. If this were true, the job involvement and organizational commitment levels of these section leaders would have been only secondary factors in explaining the characteristics of the assessments of subordinate job performance. An empirical test of this proposition would require information regarding the employment history of the sales clerks in the sample and an objective measure of their job performance for use in assessing rater accuracy. Data of this nature was not accessible for the present study. Hence it remains a question for future investigation as to whether the established antecedents of the attitudes under study are directly or indirectly related to the prevalence of systematic rating errors in assigned ratings.

The policy capturing analysis performed indicated that there were only minor differences in the relative importance raters with high or low levels of the composite attitude measure assigned to the various performance dimensions in arriving at an assessment of ratee overall job performance. Hence, it appears that in general, levels of the work-related attitudes considered are not related to any systematic differences in the mental processes raters employed in the particular context of determining the relative importance of the various dimensions upon which subordinates were evaluated.

It is difficult to make dimensional comparisons across scales on the basis of the data from this study although it would appear that corresponding dimensions on the graphic rating scale and behavioral checklist had different conceptual meanings to the raters. Comparisons of characteristics of ratings on the graphic rating scale and

behavioral checklist was a secondary consideration of this study. The results for tests using the composite variable as the independent measure did not produce any notable differences between scales. However, when job involvement and organizational commitment were considered separately, a consistent pattern of differential findings did emerge. Specifically, job involvement scores were directly related to the measured amount of each rating error in ratings made on the graphic rating scale only. In contrast, organizational commitment scores predicted the degree of leniency, central tendency, and range restriction but only for those assessments made on the behavioral checklist. Job involvement also was significantly correlated with leniency in the behavioral checklist ratings.

This unexpected finding is difficult to explain in terms of the theoretical model although one possible rationale is apparent. Ratings made on the graphic rating scale are in a sense more direct than those made on the behavioral checklist where the items have been assigned differential weights of which the raters were not aware and dimension scores are based on the summation of the checked items. For this reason it could conceivably be easier to consistently assign ratings which are purposely biased in terms of the four errors considered on the graphic rating scale. Thus, if a rater was intentionally trying to rate subordinates highly on all dimensions while not using too many extreme ratings, it follows that this would be easier to accomplish on the graphic rating scale than the behavioral checklist.

On the other hand, the behavioral checklist consists of a series of specific statements which in essence present the expected

behavioral norms of the organization. It is possible that section leaders who are very committed to the company are more inclined to perceive their subordinates as having a similar attitude than section leaders who are less committed. If this were true, it would then follow that the highly committed raters would be more likely to appraise their subordinates as also accepting the norms of the organization in terms of their behaviors on the job. This could account for the tendency for highly committed raters to assign ratings which tended to be comparatively lenient and restricted in range on the behavioral checklist. These explanations are admittedly speculative and indicate that this general finding clearly requires further research.

The relative amounts of each rating error evidenced in the evaluations made on the graphic rating scale and behavioral checklist were not directly contrasted in this study with the exception of halo. The graphic rating scale was found to be consistently more prone to this source of bias. However, the related issue examined by hypotheses 1B, 2B, 3B, and 4B also pertains to potential differences between the two rating formats. The results from the tests of these hypotheses show that none of the attitudinal factors considered predicted the occurrence of rating errors to a significantly greater degree on one rating format over the other. It would appear that despite the fact that one format may traditionally have been more prone to certain errors than the other, the influence of rater attitudes on the occurrence of these sources of systematic bias does not appear to be strong enough to accentuate these differences.

In testing hypotheses 3A, 3B, 3C, and 3D, the measure of rater ability provided by the section leader performance rating scale was found to moderate the relationship between the composite attitude measure and the extent of leniency and central tendency in ratings assigned on the graphic rating scale. This effect was not noted for restriction of range and halo. However, the test of this hypothesis for range restriction was based on a single measure and utilized a statistical procedure of relatively low power due to the nature of this measure and the limited size of the section leader sample. Apparently, whether the work-related attitudes themselves or their antecedents are causing different degrees of leniency and central tendency to be present in assigned ratings, the general ability of the raters is serving to enhance the strength of this relationship. This premise corresponds with the predictions of Fishbein's model. It is suspected that the reason why this relationship did not hold for the behavioral checklist as well was again that the extent to which raters were able to purposely bias ratings, particularly in terms of leniency and central tendency, was under their volitional control to a greater extent with the trait-based, graphic rating scale. It is also possible that while the performance measure used as the moderator in this study provided a general index of section leader ability, a more direct measure of the degree to which raters were capable of providing error-free ratings would have been a more potent moderator of the basic relationship between the independent and dependent variables examined.

The last set of rating bias hypotheses tested -- 4A, 4B, 4C, and 4D, considered the relative ability of the job involvement and organizational commitment to account for the variance in measures of the four types of rating errors studied. The model suggests that organizational commitment will be a better predictor than job involvement of the extent of each error in assigned ratings as it is believed to have a more direct effect on rater behavior during the evaluation phase of the appraisal process. Tests of these hypotheses indicated that the proportion of variance explained by each attitude did not differ significantly for any of the rating errors considered.

The question addressed by these hypotheses concerns the relative ability of the two attitude measures to predict the occurrence of systematic bias in ratings of subordinate job performance. The model suggests that the two variables function concomitantly to form behavioral intentions which in turn cause behavior. According to Wiener's model, the relative weight given to each factor in predicting behavioral intentions is a function of the attributes of the situation, the subject, and the behavior itself. The data indicate that job involvement and organizational commitment when combined into a composite measure are directly related in to the prevalence of rating errors. When considered separately, the results for each attitude vary according to the particular rating format considered. The measures of rater job involvement and organizational commitment were not significantly correlated in this study ($r = .19$). This value is considerably lower than other similar correlations reported in the literature (e.g., $r = .37$, Mauer, 1968; $r = .30$ and $.54$, Mowday

et al., 1979) and may possibly be attributable to the subset of items selected from the Lodahl and Kejner scale to measure job involvement. In any event, these correlations indicate that the two measures are not redundant and are in fact measuring distinct constructs which differentially affect behavior. However, the level of development and sophistication of the theoretical model is insufficient at this point to determine which of these factors if either, will have a greater influence on the specific behaviors comprising the total appraisal process. It is apparent that contrary to the predictions made, neither attitude had a significantly greater effect on the behavior of assigning ratings as assessed in terms of the extent of the systematic rating errors detected.

Hypotheses 5 and 6 which examined the effects of rater attitudes on the reliability and convergent and discriminant validity of assigned ratings were basically exploratory in nature. It is questionable as to whether the act of assigning reliable or valid ratings can be considered a specific behavior in the same sense as assigning ratings which are excessively lenient, correlated or characterized by a limited range. In other words, it is questionable whether determining the reliability or validity of the ratings one assigns to others is under the same degree of volitional control as influencing the extent of the four systematic rating errors discussed above. Studies of training programs specifically intended to increase rater accuracy have shown that this element can be improved through training and practice (e.g., Bernardin & Pence, 1980, Latham et al., 1975). It follows that improved accuracy would result in more

reliable and valid ratings. The training program used in this study did not stress accuracy because there was no way of assessing this rating characteristic in the field setting. Hence, raters did not receive any specific training that would directly influence their ratee evaluations in this regard or make them accept the objective of assigning reliable and construct valid ratings as part of the organizational subjective norm. In light of these considerations, it is not surprising that the composite measure of work related attitudes was not related to the degree of these parameters present in the subordinate ratings assigned.

Limitations of the Study

Many of the major limitations of this study have already been alluded to in the previous discussion. The major shortcoming of the experimental procedure used was the failure to employ a fully crossed design in which all raters evaluated all ratees. Without such a design it is not possible to unambiguously estimate the all of the various rater and ratee effects (Saal et al, 1980). This problem which is very common in field studies of performance appraisal was dealt with by utilizing multiple tests of the each hypothesis whenever possible in which the procedures for operationalizing the dependent measures as well as the units of analysis were varied. This means of compensation was only partially successful as it was not possible to fully differentiate true scores from systematic error variance. The limited number of subjects available for some of these analyses,

particularly those that utilized section leaders as the experimental units, resulted in some of the statistical procedures having rather low power.

Another limitation related to the field setting of the study pertains to the lack of a measure of rating accuracy. Recent trends in research on the performance appraisal process and rater training have put a greater emphasis on the importance of accuracy in ratings than on the reduction of rating errors (e.g., Fay & Latham, 1982). The availability of an objective performance criterion against which to contrast supervisor appraisals would greatly enhance future field studies by providing a means for assessing the impact of various components of the process on the accuracy of the ratings assigned.

The inclusion of a manipulation check would have also improved this research effort. As mentioned above, it is questionable whether participation in scale development and rater training was sufficient to convey that assigning error-free ratings was part of the organization's subjective norm. A measure of the degree to which section leaders accepted this tenet might have served as a useful covariant in the analyses reported.

At the time the data for this study were collected, behavioral observation scales - BOS (Latham & Wexley, 1977, 1981) had just recently been introduced and had not yet attained their current level of popularity. These scales are developed in essentially the same fashion as the behavioral checklist used in this study but employ a Likert (1932) rather than a Thurstone (1929) scaling technique. Behavioral observation scales require raters to decide how frequently

a ratee performs each of the listed behaviors and indicate this on a five-point scale ranging from always to never. Scores are then determined by summing the numerical ratings for all of the behavioral statements categorized within a particular performance dimension. These scales have not been shown to be consistently more resistant to systematic rating errors than other formats (Bernardin, 1977). However, they do have the advantage of providing raters with greater latitude than behavioral checklists in determining the degree to which a particular statement is characteristic of an employee's general performance on the job. By combining the Likert-type response format of BOS with the importance weightings of the behavioral statements comprising behavioral checklist, it is possible that the variance of the performance ratings achieved would be significantly enhanced over that of the behavioral checklist used in this study.

Finally, throughout the course of this study, it was stressed to the participants that all test results and performance ratings would be used for research purposes only. Numerous studies have shown that ratings provided under conditions of administrative use rather than research use have fundamentally different characteristics (e.g., Borrenson 1967, Heron, 1956). The generalizability of the findings of the current study is limited by the fact that the organization studied did not have a previous history of conducting formal employee appraisals nor did they necessarily intend to install such a program. It would be advantageous in future efforts to utilize ratings from a corporate setting where performance appraisals are conducted regularly and data can be gathered in a more realistic context.

Directions for Future Research

It has been said that a worthwhile research study raises more questions than it answers. This project was intended as an initial attempt to examine the effect of two work-related rater attitudes on the characteristics of assigned ratings of subordinate job performance. The scope of future investigations in this area would benefit by incorporating other contingency relationships operating in the organizational environment to motivate various forms of behavior into the theoretical model. This is particularly important in light of the potential for conflicts between separate components of the organizational subjective norm and the relative influence of the various performance/reward relationships operating upon behavior. If the nature and impact of these factors can be systematically assessed, it is possible that the seemingly contrary behaviors observed may yet be explained within the context of the model.

Future research efforts focusing on the effect of attitudes on work behavior should be extended to include other related attitudes such as job satisfaction and feelings toward the rating process itself, as well as the effect of the interaction of rater and ratee attitudes. Additional research is needed to determine whether such attitudes themselves have a significant effect on rater behavior or whether it is really the antecedent factors precipitating the formation of these attitude which are the true determinants.

Consideration should also be given to the longitudinal effects of rater attitudes on rating behavior. It is not unlikely that time would

have a significant effect on the extent to which the section leaders in this study recognize and accept the organizational norm of providing accurate and error-free performance ratings. Similarly, changes in a supervisor's level of job involvement may be reflected in the accuracy of observation and recall of subordinate job performance. Longitudinal studies of this type would also provide the data necessary to make inferences regarding causality in the study of the relationship between attitudes, intentions, and behaviors.

Other research issues which should be addressed include the relative degree of influence and interaction of rater attitudes, ability, motivation, and demographic factors on the characteristics of evaluations assigned to others. Also, the degree to which the relationships noted in this study are generalizable to rating situations other than the appraisal of subordinate work performance should be considered. Finally, this research effort should be replicated with the improvements noted above in a different organizational setting to determine whether the variable relationships noted are unique to the present setting or representative of a general phenomenon.

Conclusions and Implications of the Study

The call for a moratorium on rating form studies has apparently been effective in redirecting the focus of research efforts in the area of performance appraisal. Numerous recent publications have addressed topics related to the overall rating process as well as an understanding of the cognitive processes involved in evaluating the behavior of others. This study may be considered to fall under the

latter of these two general categories in that it represents an attempt to further understand what happens inside the "black box" of the performance appraisal process -- the rater. The theoretical model proposed represents an attempt to integrate the apparently overlooked cognitive factor of rater attitudes into the growing body of research pertaining to this general topic.

The model essentially attempts to explain how a person's attitudes towards his/her job and work environment effect a specific set of important behaviors - the observation and evaluation of subordinate job performance. The findings of the study did not support the predictions of the model although in some instances they may be explained within the same theoretical context. It is suggested that the model be expanded to take into account the relative impact of various potentially conflicting elements in the organizational subjective norm as well as the effect of similarly contrary performance -- self-esteem contingencies. If the most potent of these elements and relationships can be identified, then the measurement of job involvement and organizational commitment may then lead to a more accurate prediction of subsequent behavior within the context of the model.

Upon initial inspection, the experimental results obtained from the test of the model would seem to indicate that individuals who have the lowest levels of job involvement and organizational commitment provide the most bias-free appraisals of their coworkers' performance on the job. However, the likelihood of convincing a manager that low

levels of these attitudes would constitute an effective criteria for promoting individuals to supervisory positions is not very great.

A more promising approach to pursue is suggested by some of the alternative explanations proposed to explain the research results. As suggested, the experimental findings obtained may be the result of performance -- self-esteem contingencies and components of the perceived subjective norm which are in conflict with the goal of attaining valid and error-free ratings of job performance. Many researchers who have studied the cognitive processes involved in performance appraisal (e.g., Murphy et al, 1982; Ilgen & Favero, 1985) have begun to conclude that the real problem is not whether raters are capable of evaluating the performance of others in an unbiased fashion but rather whether they are willing to do so. In this sense, the problem is seen as one of motivation rather than ability. The objective for the organization then becomes one of changing the contingencies and the subjective norms to ensure that the goal of providing accurate and relatively error-free ratings of subordinates become sufficiently salient to raters to motivate their behavior in this direction. It may be that the only way to effectively influence the levels of job involvement and organizational commitment of supervisors and managers is through the selection process. However, the nature of the jobs in which they are involved and the organizations to which they are committed can be modified in such a way as to maximize the behaviors precipitated by high levels of these attitudes.

If this premise is accepted, then the model implies that organizations must restructure their intrinsic and extrinsic reward systems

so as to foster the desired behaviors on the part of their supervisory personnel in the areas of subordinate performance observation and appraisal. Also, the importance of performing these supervisory functions to the fullest extent of each individual's ability for the good of the organization must be incorporated into the subjective norm of the work environment.

Such fundamental changes as those contemplated will not be accomplished easily or quickly as there are many forces in the workplace which directly conflict with the objective of providing effective performance appraisals. This is the reason why in so many cases the objectives of appraisal programs are never attained. Regardless of our degree of understanding of the rating process, performance appraisal will not work in organizations unless management makes a commitment to structure the process and the environment in such a way that it can work. This means providing the motivational inducements for supervisors to make these programs function as intended. If this difficult task can be accomplished it is likely that raters who are highly involved in their jobs and strongly committed to the goals and values of their organization will indeed provide effective and psychometrically sound appraisals of subordinate job performance.

LIST OF REFERENCES

LIST OF REFERENCES

- Arnold, H. J. Moderator variables: A clarification of conceptual, analytic, Organizational Behavior and Human Performance, 1977, 18, 269-294.
- Bass, B. M. Reducing leniency in merit ratings. Personnel Psychology, 1956, 9, 359-369.
- Bayroff, A. G., Haggerty, H. R., & Rundquist, E.A. Validity of ratings as related to rating techniques and conditions. Personnel Psychology, 1954, 7, 93-114.
- Bell, F. O., Hoff, A. L., & Hoyt, K. B. A comparison of three approaches to criterion measurement. Journal of Applied Psychology, 1983, 47, 416-418.
- Bennett, G. K., & Gelink, M. Short Employment Tests Manual, (1972 Revised), New York: Psychological Corporation, 1972.
- Bernardin, H. J. Behavior expectation scales versus summated scales: A fairer comparison. Journal of Applied Psychology, 1977 62, 422-427.
- Bernardin, H. J. Effects of rates training on leniency and halo errors in student ratings of instructors. Journal of Applied Psychology, 1978, 63, 301-308.
- Bernardin, H. J., Alvares, K. M., & Cranny, C. J. A recomparison of behavioral expectation scales. Journal of Applied Psychology, 1976, 61, 564-570.
- Bernardin, H. J., & Boetcher, R. Training and cognitive complexity effects on psychometric error in rating. Paper presented at the 86th annual meeting of the American Psychological Association, Toronto, August 1978.
- Bernardin, H. J., Cardy, R. L., & Carlyle, J. J. Cognitive complexity and appraisal effectiveness; Back to the drawing board? Journal of Applied Psychology, 1982, 67, 151-160.
- Bernardin, H. J., & Pence, E. C. Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 1980, 65, 60-66.
- Bernardin, H. J., & Walter, C. S. Effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 1977, 61, 64-69.

- Bigoness, N. J. Effect of applicant's sex, race, and performance on employer's performance ratings: Some additional findings. Journal of Applied Psychology, 1976, 61, 80-84.
- Blood, M. R., & Mullet, G. M. Where have all the moderators gone? The perils of type II error. (Tech. Rep. No. 11). College of Industrial Management, Georgia Institute of Technology, 1977.
- Borman, W. C. Effects of instructions to avoid halo error on reliability and validity of performance evaluation ratings, Journal of Applied Psychology, 1975, 60, 556-560.
- Borman, W. C. Exploring upper limits of reliability and validity in job performance ratings. Journal of Applied Psychology, 1978, 63, 135-144.
- Borman, W. C. Individual differences correlates of accuracy in evaluating other's performance effectiveness. Applied Psychological Measurement, 1979, 64, 103-115. (a)
- Borman, W.C. Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 1979, 64, 410-421. (b)
- Borman, W. C., & Dunnette, M. D. Behavior based versus trait-oriented performance ratings: An empirical study. Journal of Applied Psychology, 1975, 60, 561-565.
- Borman, W. C., & Vallon, W. R. A view of what can happen when behavioral expectation scales are developed in one setting and used in another. Journal of Applied Psychology, 1974, 59, 197-201.
- Borresen, H. A. The effects of instructions and item content on three types of ratings. Educational and Psychological Measurement, 1967, 27, 855-862.
- Boruch, R. F., Larkin, J. D., Wolins, L., & MacKinney, A. C. Alternative methods of analysis: multitrait-multimethod data. Educational and Psychological Measurement, 1970, 30, 81-105.
- Brown, E. M. Influence of training, method, and relationship on the halo effect. Journal of Applied Psychology, 1968, 52, 195-199.
- Brunner, J. S., & Tagiuri, R. The perception of people. In G. Lindzey (Ed.) Handbook of Social Psychology (Vol 2.) Cambridge, Mass: Addison-Wesley, 1954.
- Buchanan, B. Building organizational commitment: The socialization of managers in work organizations. Administrative Science Quarterly, 1974, 19, 533-546.

- Campbell, D. T., & Fiske, D. W. Convergent and discriminant validation by the multitrait mutimethod matrix. Psychological Bulletin, 1959, 56, 81-105.
- Campbell, J. P., Dunnette, M. D., Arvey, R. D., & Hellervik, L. W. The development and evaluation of behaviorally based rating scales. Journal of Applied Psychology, 1973, 57, 15-22.
- Campbell, J. P., Dunnette, M. D., Lawler, E. E., III, & Weick, K. E. Jr. Managerial Behavior, Performance, and Effectiveness. New York: McGraw-Hill, 1970.
- Cascio, W. F. Applied Psychology in Personnel Management. Reston, Virginia: Reston, 1978.
- Cascio, W. F., & Valenzi, E. R. Behaviorally anchored rating scales: Effects of education and job experience of raters and ratees. Journal of Applied Psychology, 1977, 3, 278-282.
- Cascio, W. F., & Valenzi, E. R. Relations among criteria of police performance, Journal of Applied Psychology, 1978, 62, 22-28.
- Christal, R. E. JAN: A technique for analyzing group judgment. Journal of Experimental Education, 1968, 36, 24-27.
- Cleveland, J. N., & Landy, F. J. The influence of rater and ratee age on two performance judgments. Personnel Psychology, 1981, 34, 19-29.
- Cohen, J. Statistical Power Analysis for the Behavioral Sciences. New York: Academic Press, 1977.
- Cohen, J. Partialled products are interactions; Partialled powers are curve components. Psychological Bulletin, 1977, 85, 858-866.
- Cohen, J., & Cohen, P. Applied Multiple Regression/Correlation for the Behavioral Sciences. Hillsdale, NJ: Erlbaum, 1975.
- Cooper, W. H. Ubiquitous halo. Psychological Bulletin, 1981, 90, 218-244.
- Cox, J. A., & Krumboltz, J. D. Racial Bias in peer ratings of basic airmen. Sociometry, 1958, 21, 292-299.
- Cronbach, L. J. Processes affecting scores on understanding of others and assuming similarity. Psychological Bulletin, 1955, 52, 177-193.
- Cronbach, L. J., Gleser, G. C., Nanda, H., & Rajaratnam, N. The dependability of behavioral measurements: Theory of Generalizability for Scores and Profiles. New York: Wiley, 1972.

- Crooks, L. A. (Ed.) An investigation of sources of bias in the prediction of job performance: A six-year study. Proceedings of Invitational Conference on Testing Problems, Princeton, N.J.: Educational Testing Service, 1972.
- Deaux, J. E., & Taynor, J. Evaluation of male and female ability: Bias works two ways. Psychological Reports, 1973, 32, 261-262.
- DeCotiis, T. A. An analysis of the external validity and applied relevance of three rating formats. Organizational Behavior and Human Performance, 1977, 247-266.
- DeCotiis, T. A., & Petit, A. The performance appraisal process: A model and some testable propositions. Academy of Management Review, 1978, 3, 635-646.
- DeJung, J. E., & Kaplan, H. Some differential effects of race of rater and combat attitude. Journal of Applied Psychology, 1962, 46, 370-374.
- Dickinson, T. I., & Tice, T. E. A multitrait-multimethod analysis of scales developed by retranslation. Organizational Behavior and Human Performance, 1973, 9, 421-438.
- Dunnette, M. D., & Borman, W. C. Personnel selection and classification systems. Annual Review in Psychology, 1979, 30, 477-525.
- Farris, G. F. A predictive study of turnover. Personnel Psychology, 1971, 24, 311-328.
- Fay, C. H., & Lathem, G. P., Effect of training and rating scales on rating errors. Personnel Psychology, 1982, 35, 105-116.
- Feild, H. S., & Holley, W. H. Performance appraisal - An analysis of statewide practices. Public Personnel Management, 1975, 7, 145-150.
- Feldman, J. M. Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 1981, 60, 127-148.
- Fishbein, M. Attitude and prediction of behavior. In M. Fishbein (ed.) Readings in attitude theory and measurement. New York: Wiley, 1967, 477-492.
- Fishbein, M. The prediction of behavior from attitudinal variables. In C. D. Mortensen and K. K. Sereno (Eds) Advances in Communication research. New York: Harper & Row, 1973, pp. 3-31.
- Fishbein, M., & Ajzen, I. Attitudes toward objects as predictors of single and multiple behavioral criteria. Psychological Review, 1974, 81, 59-74.

- Fishbein, M., & Ajzen, I. Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research. Reading, Massachusetts: Addison-Wesley, 1975.
- Flanagan, J. C. The critical incident technique. Psychological Bulletin, 1954, 51, 327-358.
- Frank, L. L., & Hackman, J. R. Effects of interviewer-interviewee similarity on interviewer objectivity in college admissions interviews. Journal of Applied Psychology, 1975, 60, 356-360.
- Friedman, B. A., & Cornelius, E. T. Effect of rater participation in scale construction on the psychometric properties of two rating scale formats. Journal of Applied Psychology, 1976, 61, 210-216.
- Goldberg, P. A. Are women prejudiced against women? Transaction, 1968, 5, 28-30.
- Goodman, P. S., Rose, J. H., & Furcon, J. E. Comparison of motivational antecedents of the work performance of scientists and engineers. Journal of Applied Psychology, 1970, 54, 491-495.
- Guilford, J. P. Psychometric Methods. New York: McGraw-Hill, 1954.
- Guion, R. M. Personnel Testing. New York: McGraw-Hill, 1965.
- Hall, D. T., & Mansfield, R. Organizational and individual response to external stress. Administrative Science Quarterly, 1971, 16, 533-547.
- Hamner, W. C., Kim, J. S., Baird, L., & Bigoness, W. J. Race and sex as determinants of ratings by potential employers in a simulated work-sampling task. Journal of Applied Psychology, 1974, 59, 705-711.
- Hays, W. L., Statistics for the Social Sciences (Second Edition). New York: Holt, Rinehart and Wilson, 1973.
- Heron, A. The effects of real-life motivation on questionnaire response. Journal of Applied Psychology, 1956, 40, 65-68.
- Holzbach, R. L. Rater bias in performance ratings: Superior, self, and peer ratings. Journal of Applied Psychology, 1978, 63, 5, 579-588.
- Hom, P. W., Katerberg Jr., R., & Hulin, C. L. Comparative examination of three approaches to the prediction of turnover. Journal of Applied Psychology, 1979, 64, (3), 280-290.

- Ilgen, D. R., & Favero, J. L., Limits in generalization from psychological research to performance appraisal processes. Academy of Management Review, 1985, 10, 311-321.
- Ivancevich, J. M. Longitudinal study of the effects of rater training on psychometric error in rating. Journal of Applied Psychology, 1979, 64, 5, 502-508.
- Jones, E. E., & Gerard, H. B. Foundations of Social Psychology. New York: Wiley, 1967.
- Jurgensen, C. E. Item weights in employee rating scales. Journal of Applied Psychology, 1955, 39, 305-307.
- Kavanaugh, M. J. The content issue in performance appraisal: A review. Personnel Psychology, 1971, 24, 653-669.
- Kavanaugh, M. J., MacKinney, A. C., & Wollins, L. Issues in managerial performance: Multitrait-multimethod analysis of variance. Psychological Bulletin, 1971, 75, 34-49.
- Keaveny, T. J., & McGann, A. F. A comparison of behavioral expectations scales and graphic rating scales. Journal of Applied Psychology, 1975, 60, 695-703.
- Kelly, G. A. The Psychology of Personal Constructs. New York: Norton, 1955.
- King, L. M., Hunter, J. E., & Schmidt, F. L. Halo in a multidimensional forced-choice performance evaluation scale. Journal of Applied Psychology, 1980, 65, 507-516.
- Kirchner, W. K., & Reisberg, D. J. Differences between better and less effective supervisors in appraisal of subordinates. Personnel Psychology, 1962, 15, 295-302.
- Klimoski, R. J., & London, M. Role of the rater in performance appraisal. Journal of Applied Psychology, 1974, 59, 445-451.
- Klores, M. S. Rater bias in forced-distribution ratings. Personnel Psychology, 1966, 19, 411-421.
- Kretch, D., Crutchfield, R. S., & Ballachey, E. I. Individual in Society: A Textbook of Social Psychology. New York: McGraw-Hill, 1962.
- Lahey, M. A., & Saal, F. E. Evidence incompatible with a cognitive compatibility theory of rating behavior. Journal of Applied Psychology, 1981, 66, 706-715.

- Landy, F. J., Farr, J. L., Saal, F. E., & Freytag, W. R. Behaviorally anchored scales for rating the performance of police offices. Journal of Applied Psychology, 1976, 61, 750-758.
- Landy, F. J., & Farr, J. L. Performance rating. Psychological Bulletin, 1980, 87, 72-107.
- Landy, F. J., & Trumbo, D. A. The psychology of work behavior (Rev. ed.). Homewood, Ill., The Dorsey Press, 1981.
- Latham, G. P., & Wexley, K. N., Behavioral observation scales for performance appraisal purposes. Personnel Psychology, 1977, 30, 255-268.
- Latham, G. P., & Wexley, K. N. Increasing Productivity Through Performance Appraisal. Reading, Mass.: Addison-Wesley, 1981.
- Latham, G. P., Wexley, K. N., & Purcell, E. D. Training managers to minimize rating errors in the observation of behavior. Journal of Applied Psychology, 1975, 60, 550-555.
- Lawler III, E. E. The multitrait-multirater approach to measuring managerial job performance. Journal of Applied Psychology, 1967, 51, 369-381.
- Lawler III, E. E., & Hall, D. T. Relationship of job characteristics to job involvement, satisfaction, and intrinsic motivation. Journal of Applied Psychology, 1970, 54, 305-312.
- Lee, C. Increasing performance appraisal effectiveness: Matching task types, appraisal process, and rater training. Academy of Management Review, 1985, 10, 322-331.
- Lewis, N. A., & Taylor, J. A. Anxiety and extreme response preferences. Educational and Psychological Measurement, 1955, 15, 111-116.
- Likert, R. A., Technique for the measurement of attitudes. Archives of Psychology, 1932, No. 140.
- Locher, A. H., & Teel, K. S. Performance appraisal - a survey of current practices. Personnel Journal, 1977, 56, 245-247, 254.
- Lodahl, T. M. Patterns of job attitudes in two assembly technologies. Administrative Science Quarterly, 1967, 8, 482-519.
- Lodahl, T. M., & Kejner, M. The definition and measurement of job involvement. Journal of Applied Psychology, 1965, 49, 24-33.
- London, M., & Poplawski, J. R. Effects of information on stereotype development in performance appraisal and interview contexts. Journal of Applied Psychology, 1976, 61, 199-205.

- Mandell, M. M. Supervisory characteristics and ratings: A summary of recent research. Personnel Psychology, 1956, 32, 435-440.
- Mauer, J., Work as a central life interest of industrial supervisors. Academy of Management Journal, 1968, 11, 329-339.
- Motowidlo, S. J., & Borman, W. C. Behaviorally anchored scales for measuring morale in military units. Journal of Applied Psychology, 1977, 62, 177-183.
- Mowday, R. T., Porter, L. W., & Dubin, R. Unit performance, situational factors, and employee attitudes in spatially separated work units. Organizational Behavior and Human Performance, 1974, 12, 231-248.
- Mowday, R. T., Steers, R. M., & Porter, L. W. The measurement of organizational commitment. Journal of Vocational Behavior, 1979, 14, 224-247.
- Mullins, C. J., & Force, R. C. Rater accuracy as a generalized ability. Journal of Applied Psychology, 1962, 46, 191-193.
- Murphy, K. R., Garcia, M., Kerkar, S., Martin, C., & Balzar, W. K. Relationship between observational accuracy and accuracy in evaluating performance. Journal of Applied Psychology, 1982, 67, 320-325.
- Nadler, D. A., & Lawler III, E. E. Motivation: A diagnostic approach. In Hackman, J. R., Lawler, E. E., & Porter, L. W. (eds.) Perspectives on behavior in organizations. New York: McGraw-Hill, 1977.
- Norton, S. C., Gustafson, D. P., & Foster, C. E. Assessment for management potential: Scale design and development, training effects and rater/ratee sex effects. Academy of Management Journal, 1977, 20, 117-131.
- Nunnally, J. C. Psychometric Theory. New York: McGraw-Hill, 1967.
- Osborn, R., & Vicars, W. Sex stereo-types: An artifact in leader behavior and subordinate satisfaction analysis. Academy of Management Journal, 1976, 19, 439-449.
- Patchen, M. Participation, Achievement and Involvement on the Job. Englewood Cliffs, N.J.: Prentice-Hall, 1970.
- Pheterson, G. I., Kiesler, S. B., & Goldberg, P. A. Evaluation of the performance of women as a function of their sex, achievement, and personal history. Journal of Personality and Social Psychology, 1971, 19, 114-118.

- Pitman, E. J. G., Significance tests which may be applied to samples from any populations. Supplement to J. R. Statistical Sociology, 1937, 4, 119 - 130.
- Porter, L. W., & Lawler III, E. E. Managerial attitudes and performance. Homewood, Ill.: Irwin, 1968.
- Porter, L., Steers, R., Mowday, R., & Boulian, P. Organizational commitment, job satisfaction and turnover among psychiatric technicians. Journal of Applied Psychology, 1974, 59, 603-609.
- Pursell, E. D., Dossett, D. L., & Latham, G. P., Obtaining valid predictors by minimizing rating errors in the criteria. Personnel Psychology, 1980, 33, 91-96.
- Rabinowitz, S., & Hall, D. T. Organizational research on job involvement. Psychological Bulletin, 1977, 84, 265-288.
- Rose, G. L., & Stone, T. H. Why good job performance may (not) be rewarded: Sex factors and career development. Journal of Vocational Behavior, 1978, 12, 197-207.
- Rothe, H. F. The relation of merit ratings to length of service. Personnel Psychology, 1949, 2, 237-242.
- Rothaus, P., Morton, R. B., and Hanson, P. G. Performance appraisal and psychological distance. Journal of Applied Psychology, 1965, 49, 48-54.
- Saal, F. E., & Landy, F. J. The mixed standard rating scale: An evaluation. Organizational Behavior and Human Performance, 1977, 18, 3-18.
- Saal, F. E., Downey, R. G., & Lahey, M. A. Rating the ratings: Assessing the psychometric quality of rating data. Psychological Bulletin, 1980, 88, 413-428.
- Saleh, S. D., & Hosek, J. Job involvement: Concepts and measurements. Academy of Management Journal, 1976, 19, 213-224.
- Saunders, D. R., Moderator variable in prediction. Educational and Psychological Measurement, 1956, 16, 209-222.
- Sauser, W. I., & Pond, S. B. Effects of rater training and participation on cognitive complexity: An exploration of Scheier's cognitive reinterpretation. Personnel Psychology, 1981, 34, 563-577.
- Schmidt, F. L., & Hunter, J. E. Development of a general solution to the problem of validity generalization. Journal of Applied Psychology, 1977, 62, 529-540.

- Schmidt, F. L., & Johnson, R. H. Effect of race on peer ratings in an industrial situation. Journal of Applied Psychology, 1973, 57, 237-241.
- Schneider, D. E., & Bayroof, A. G. The relationship rater characteristics and the validity of ratings. Journal of Applied Psychology, 1953, 31, 278-280.
- Schneier, C. E. Operational utility and psychometric characteristics of behavioral expectation scales: A cognitive reinterpretation. Journal of Applied Psychology, 1977, 62, 541-548.
- Schwab, D. P., Heneman III, H. G., & DeCotiis, T. Behaviorally anchored rating scales: A review of the literature. Personnel Psychology, 1975, 28, 546-562.
- Siegel, S., Nonparametric Methods for the Behavioral Sciences. New York: McGraw Hill, 1956.
- Sharon, A. T., & Bartlett, C. J. Effects of instructional conditions in producing leniency on two types of rating scales. Personnel Psychology, 1969, 22, 251-263.
- Siegel, A. L., & Ruh, R. A. Job involvement, participation in decision making, personal background, and job behavior. Organizational Behavior and Human Performance, 1973, 9, 318-327.
- Smith, P. C., & Kendall, L. M. Retranslation of expectations: An approach to the construction of unambiguous anchors for rating scales. Journal of Applied Psychology, 1963, 47, 149-155.
- Smith, P. L. S. Job Involvement and Communications. London: Business Books, 1973.
- Stanley, J. C., Analysis of unreplicated three-way classifications with applications to rater bias and trait independence. Psychometrika, 1961, 26, 205-219.
- Staw, B. M. Two sides of commitment. Paper presented at the National Meeting of the Academy of Management, 1977. In R. M. Steers & L. W. Porter (eds.) Motivation and Work Behavior. (Second edition). New York: McGraw-Hill, 1979.
- Steers, R. M. Antecedents and outcomes of organizational commitment. Administrative Science Quarterly, 1977, 22, 46-56.
- Stockford, L., & Bissel, H. W., Factors involved in establishing a management rating scale. Personnel, 1949, 26, 94-116.
- Stockford, L., & Bissell, H. W. Establishing a graphic-rating scale. In W. E. Fleishman (Ed.), Studies in Personnel and Industrial Psychology. (Rev. ed.) Homewood, Ill.: Dorsey Press, 1967.

- Stone, E. F., & Hollenbeck, J. R., Some issues associated with the use of moderated regression. Organizational Behavior and Human Performance, 1984, 34, 195-213.
- Taft, R. The ability to judge people. Psychological Bulletin, 1955, 52, 1-23.
- Taylor, E. K., Parker, J. W., Martens, L., & Ford, G. L. Supervisory climate and performance ratings, an exploratory study. Personnel Psychology, 1959, 12, 453-468.
- Taynor, J., & Deaux, K. When women are more deserving than men: Equity, attribution, and perceived sex differences. Journal of Personality and Social Psychology, 1973, 28, 360-367.
- Thorndike, E. L. A constant error in psychological ratings. Journal of Applied Psychology, 1920, 4, 25-29.
- Thorndike, R. L. Personnel Selection. New York: Wiley, 1949.
- Thornton III, C. G., & Zorich, S. Training to improve observer accuracy. Journal of Applied Psychology, 1980, 65, 351-354.
- Thurstone, L. L., Theory of attitude measurement. Psychological Bulletin, 1929, 36, 224-241.
- Vroom, V. H. Projection, negation, and self-concept. Human Relations, 1959, 12, 335-344.
- Vroom, V. H. Ego involvement, job satisfaction, and job performance. Personnel Psychology, 1962, 15, 159-177.
- Vroom, V. H. Work and motivation. New York: Wiley, 1964.
- Warmke, D. L., & Billings, R. S. Comparison of training methods for improving the psychometric quality of experimental and administrative performance ratings. Journal of Applied Psychology, 1979, 64, 124-131.
- Wendelken, D. J., & Inn, A. Nonperformance influences on performance evaluations: A laboratory phenomenon? Journal of Applied Psychology, 1981, 66, 149-158.
- Wicker, A. W. An examination of the "other variables" explanation of attitude-behavior inconsistency. Journal of Personality and Social Psychology, 1971, 19, 18-30.
- Wiener, Y., Commitment in organizations: A normative view. Academy of Management Review, 1982, 7, 418-427.

- Wiener, Y., & Vardi, Y. Relationships between job, organization, and career commitments and work outcomes - An integrative approach. Organizational Behavior and Human Performance, 1980, 26, 81-96.
- Wolins, L., A Procedure for Estimating the Amount of Trait, Method, and Error Variance Attributable to a Measure. Washington, D. C., Office of Education, U. S. Department of Health, Education, and Welfare, 1964.
- Zedeck, S. Problems with the use of moderator variables. Psychological Bulletin, 1971, 76, 295-310.
- Zedeck, S., & Kafry, D. Capturing rater policies for processing evaluation data. Organizational Behavior and Human Performance, 1977, 18, 269-294.
- Zyzanski, S. J., Analysis of Variance Applied to Factors Which Do Not Have Comparable Scales. Unpublished master's thesis, Iowa State University, Ames, Iowa, 1962.

APPENDICES

APPENDIX A

Sales Clerk Performance Appraisal Form

SALES CLERK

PERFORMANCE APPRAISAL INSTRUCTIONS

Ratings of Job Behavior

Listed on each of the following pages are a series of job behaviors categorized in nine general dimensions which are important parts of the Sales Clerk job. Read each statement carefully. If you believe that the statement is generally descriptive of the job performance of the employee being evaluated over the past year (or however long you have supervised the employee), place a check on the line to the left of the statement. If the statement does not describe the employee's typical behavior on the job over this period, leave the line blank. Evaluate the employee in relation to each of the statements in all nine areas. There is no minimum or maximum number of statements which may be checked. Mark as many or as few statements as you feel apply to the employee.

Dependability - Fulfills job duties and personal commitments reliably and punctually.

- _____ 1. Arrive for work and returns from breaks and lunch on time.
- _____ 2. Works within the department schedule without requesting excessive time off.
- _____ 3. Is flexible in determining and changing work schedules.
- _____ 4. Is rarely absent from work.
- _____ 5. Does not leave the department sales floor when there is inadequate coverage or without notifying anyone.
- _____ 6. Seldom leaves the department for reasons not related to the job.
- _____ 7. Completes departmental work assignments promptly.
- _____ 8. Rarely lets personal or family commitments interfere with job responsibilities and duties.

Knowledge of Job Procedures - Understands and performs job duties according to established procedures. Carries out these duties in an efficient manner with few errors.

- _____ 1. Makes very few errors in calculating (e.g., figuring sales tax, percentages for layaways, etc.).
- _____ 2. Balances and closes out the cash register accurately.
- _____ 3. Uses the cash register correctly and efficiently in completing sales transactions.
- _____ 4. Uses the appropriate forms in carrying out special transactions.
- _____ 5. Examines the bad check and credit card lists when appropriate.
- _____ 6. Correctly completes sales and credit card forms recording all necessary information.
- _____ 7. Does not hold items for customers without proper identification.
- _____ 8. Fills out weekly time cards properly and submits them on time.

Sales Ability - Presentation of merchandise in a favorable and appropriate manner. Use of proper sales techniques to help customers make purchases.

- _____ 1. Suggests the purchase of items which coordinate with merchandise already selected.
- _____ 2. Begins sales suggestions with items from the upper end of the price range.
- _____ 3. Works well under pressure.
- _____ 4. Is capable of selling merchandise in other departments when needed.
- _____ 5. Speaks clearly with proper diction.
- _____ 6. Interacts with customers in a self-confident manner.
- _____ 7. Uses appropriate sales techniques and the proper amount of persistence to make sales.
- _____ 8. Engages in spontaneous conversation with customers related to the sale of merchandise.

Customer Service - Treats customers in a pleasant, courteous fashion and adequately serves the customers' needs.

- _____ 1. Does not carry on conversations with friends or other employees while customers wait for service.
- _____ 2. Shows customers to dressing rooms personally rather than merely telling them where the rooms are.
- _____ 3. Offers to check availability of merchandise at other stores.
- _____ 4. Greets customers promptly and cheerfully.
- _____ 5. Maintains composure and courteous manner when dealing with an irate customer.
- _____ 6. Effectively waits on more than one customer at the same time when necessary.
- _____ 7. Is courteous to customers over the telephone.
- _____ 8. Does not rush customers when it is close to closing time or when the department is crowded.

Knowledge of Company Policies - Knowledge of Company policy and procedures as related to job duties and responsibilities.

_____ 1. Seldom needs to be reminded of company policies and procedures.

_____ 2. Does not sit while on the sales floor.

_____ 3. Understands the duties and responsibilities of the sales clerk job.

_____ 4. Does not eat, drink, or chew gum while on the sales floor.

_____ 5. Finds constructive work to do in the department when no customers are present.

_____ 6. Does not spend excessive time "hugging" the cash register instead of waiting on customers.

_____ 7. Adheres to Company policy regarding private telephone conversations.

Product Knowledge - Knowledge of the qualities and features of department merchandise including coordinated accessories, instructions for special care, and manufacturers' warranties.

- _____ 1. Keeps familiar with all merchandise in the department.
- _____ 2. Is aware of expected dates for incoming merchandise.
- _____ 3. Effectively answers questions concerning merchandise in the department without assistance of the section leader.
- _____ 4. Knows the special care procedures which some items require.
- _____ 5. Is familiar with and suggests accessory merchandise which corresponds to items selected by the customer.
- _____ 6. Is able to suggest acceptable alternatives for products/brands out of stock or not carried by the store.
- _____ 7. Provides appropriate justification for the price of more expensive items.
- _____ 8. Is able to correctly determine whether merchandise returned without a sales slip was purchased from _____.

Initiative - Performance of routine and non-routine job duties without specific requests made by the section leader; willingness to work beyond ordinary job requirements; striving to attain company goals; and a willingness to take independent action.

- _____ 1. Periodically rearranges departmental merchandise displays as appropriate without having to be asked to do so.
- _____ 2. Does stock work without having to be asked by the section leader.
- _____ 3. Is creative and comes up with useful ideas to improve the department.
- _____ 4. Straightens up the department regularly without being asked.
- _____ 5. Makes suggestions for moving merchandise from one place to another in the department to increase sales.
- _____ 6. Reports to work early to prepare for the shift.
- _____ 7. Suggests that customers open charge accounts.
- _____ 8. Takes responsibility for making certain decisions when the section leader is away from the department.

Judgment - Ability to observe and assess routine and unusual situations arising on the job and determine the appropriate action to take.

- _____ 1. Can correctly determine whether or not a customer wishes to be waited on.
- _____ 2. Suggests compromises which will satisfy customers returning used merchandise.
- _____ 3. Does not rely on the section leader to make all decisions in the department.
- _____ 4. Expresses opinions to customers when appropriate.
- _____ 5. Calls for assistance from section leaders when it is appropriate to do so.
- _____ 6. Keeps aware of potential shoplifters and notifies security when appropriate.
- _____ 7. Determines the appropriate order upon which to wait on customers.
- _____ 8. Correctly determines when it is appropriate to assist a fellow sales clerk when s/he is having a problem with a customer.

Employee Relations - Ability to maintain friendly and cooperative relationships with the section leader and other employees on the job.

- _____ 1. Gets along well with fellow employees.
- _____ 2. Shows consideration for fellow employees.
- _____ 3. Works well as a team member by sharing job responsibilities.
- _____ 4. Does not grab sales from other sales clerks.
- _____ 5. Is willing to share customers with other employees when appropriate.
- _____ 6. Takes time to answer the questions of new employees.
- _____ 7. Accepts constructive criticism and suggestions when offered by the section leader or other employees without becoming defensive.

SALES CLERK

PERFORMANCE APPRAISAL INSTRUCTIONS

Ratings of Performance Dimensions

The following page contains the names and definitions of nine dimensions of job performance identified as being important for the Sales Clerk job. You are asked to evaluate the job performance of each of the Sales Clerks you supervise in each of these areas. You will do this by first reading the description of the each dimension. Next determine the level of the employee's performance in each area over the past year (or however long you have supervised the employee) on the basis of the following scale:

1 = Unacceptable

2 = Marginal

3 = Acceptable

4 = Average

5 = Good

6 = Very Good

7 = Outstanding (among the best ever
seen)

On the scale to the right of each dimension, circle the number which corresponds to the level of the employee's job performance in that area. Be sure that you evaluation the employees on each of the nine dimensions listed.

- 1 = Unacceptable
 2 = Marginal
 3 = Acceptable
 4 = Average
 5 = Good
 6 = Very Good
 7 = Outstanding (among the best ever seen)

<u>Dimension</u>	<u>Performance Rating Scale</u>
1. <u>Dependability</u> - Fulfills job duties and personal commitments reliably and punctually.	1 2 3 4 5 6 7 -----
2. <u>Knowledge of Job Procedures</u> - Understands and performs job duties according to established procedures. Carries out these duties in an efficient manner with few duties.	1 2 3 4 5 6 7 -----
3. <u>Sales Ability</u> - Presentation of merchandise in a favorable and appropriate manner. Use of proper sales techniques to help customers make purchases.	1 2 3 4 5 6 7 -----
4. <u>Customer Service</u> - Treats customers in a pleasant, courteous fashion and adequately serves the customers' needs.	1 2 3 4 5 6 7 -----
5. <u>Knowledge of Company Policies</u> - Knowledge of Company policy and procedures as related to job duties and responsibilities.	1 2 3 4 5 6 7 -----
6. <u>Product Knowledge</u> - Knowledge of the qualities and features of department merchandise including coordinated accessories, instructions for special care, and manufacturers' warranties.	1 2 3 4 5 6 7 -----
7. <u>Initiative</u> - Performance of routine and non-routine job duties without specific requests made by the section leader; willingness to work beyond ordinary job requirements; striving to attain company goals; and a willingness to take independent action.	1 2 3 4 5 6 7 -----
8. <u>Judgment</u> - Ability to observe and assess routine and unusual situations arising on the job and determine the appropriate action to take.	1 2 3 4 5 6 7 -----
9. <u>Employee Relations</u> - Ability to maintain friendly and cooperative relationships with section leaders and other employees.	1 2 3 4 5 6 7 -----

SALES CLERK

OVERALL PERFORMANCE EVALUATION

Considering all aspects of this employee's job performance, how would you rate this person in terms of overall performance on the job? (check one)

- ☐ Performance does not meet minimum standards.
- ☐ Less than satisfactory performance in many respects.
- ☐ Satisfactory performance in most respects, but not all.
- ☐ Satisfactory performance in all respects.
- ☐ Above average performance but not superior.
- ☐ Performance is superior in most respects.
- ☐ Performance is definitely superior in all respects.
- ☐ The best performance I have ever observed or could hope to observe.

APPENDIX B

**Mean and Standard Deviation of the Importance Ratings
for Items Included on the Behavioral Checklist**

Dependability - Fulfills job duties and personal commitments reliably and punctually.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Arrive for work and returns from breaks and lunch on time.	5.56	1.27
2. Works within the department schedule without requesting excessive time off.	5.94	1.10
3. Is flexible in determining and changing work schedules.	6.01	0.90
4. Is rarely absent from work.	6.56	1.02
5. Does not leave the department sales floor when there is inadequate coverage or without notifying anyone.	4.47	1.38
6. Seldom leaves the department for reasons not related to the job.	4.26	1.33
7. Completes departmental work assignments promptly.	5.16	1.42
8. Rarely lets personal for family commitments interfere with job responsibilities and duties.	6.06	0.89

Scores on the behavioral checklist dimension scales were determined by summing the weights of the checked items and dividing by the number of items in the scale.

Maximum possible score = 5.50 Scale midpoint = 2.75

Knowledge of Job Procedures - Understands and performs job duties according to established procedures. Carries out these duties in an efficient manner with few duties.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Makes very few errors in calculating (e.g., figuring sales tax, percentages for layaways, etc.).	4.89	1.11
2. Balances and closes out the cash register accurately.	5.41	1.29
3. Uses the cash register correctly and efficiently in completing sales transactions.	5.90	1.20
4. Uses the appropriate forms in carrying out special transactions.	5.65	1.35
5. Examines the bad check and credit card lists when appropriate.	6.51	0.99
6. Correctly completes sales and credit card forms recording all necessary information.	5.59	1.10
7. Does not hold items for customers without proper identification.	6.18	1.01
8. Fills out weekly time cards properly and submits them on time.	5.38	1.46

Maximum possible score = 5.69 Scale midpoint = 2.84

Sales Ability - Presentation of merchandise in a favorable and appropriate manner. Use of proper sales techniques to help customers make purchases.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Suggests the purchase of items which coordinate with merchandise already selected.	5.41	1.27
2. Begins sales suggestions with items from the upper end of the price range.	5.27	1.15
3. Works well under pressure.	6.09	0.87
4. Is capable of selling merchandise in other departments when needed.	5.31	1.14
5. Speaks clearly with proper diction.	5.44	1.31
6. Interacts with customers in a self-confident manner.	5.11	1.40
7. Uses appropriate sales techniques and the proper amount of persistence to make sales.	5.94	1.27
8. Engages in spontaneous conversation with customers related to the sale of merchandise.	5.85	1.20

Maximum possible score = 5.55 Scale midpoint = 2.78

Customer Service - Treats customers in a pleasant, courteous fashion and adequately serves the customers' needs.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Does not carry on conversations with friends or other employees while customers wait for service.	4.87	1.44
2. Shows customers to dressing rooms personally rather than merely telling them where the rooms are.	5.11	1.32
3. Offers to check availability of merchandise at other stores.	6.21	1.01
4. Greets customers promptly and cheerfully.	4.21	1.39
5. Maintains composure and courteous manner when dealing with an irate customer.	5.64	1.28
6. Effectively waits on more than one customer at the same time when necessary.	5.78	1.22
7. Is courteous to customers over the telephone.	5.49	1.37
8. Does not rush customers when it is close to closing time or when the department is crowded.	4.56	1.48

Maximum scale score = 5.23 Scale midpoint = 2.62

Knowledge of Company Policies - Knowledge of Company policy and procedures as related to job duties and responsibilities.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Seldom needs to be reminded of company policies and procedures.	5.88	1.16
2. Does not sit while on the sales floor.	5.31	1.24
3. Understands the duties and responsibilities of the sales clerk job.	5.97	0.86
4. Does not eat, drink, or chew gum while on the sales floor.	4.02	1.01
5. Finds constructive work to do in the department when no customers are present.	5.11	1.39
6. Does not spend excessive time "hugging" the cash register instead of waiting on customers.	5.21	1.14
7. Adheres to Company policy regarding private telephone conversations.	4.71	0.88

Maximum scale score = 5.17 Scale midpoint = 2.59

Product Knowledge - Knowledge of the qualities and features of department merchandise including coordinated accessories, instructions for special care, and manufacturers' warranties.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Keeps familiar with all merchandise in the department.	5.66	1.27
2. Is aware of expected dates for incoming merchandise.	6.01	1.02
3. Effectively answers questions concerning merchandise in the department without assistance of the section leader.	6.57	0.73
4. Knows the special care procedures which some items require.	6.30	1.31
5. Is familiar with and suggests accessory merchandise which corresponds to items selected by the customer.	5.74	1.18
6. Is able to suggest acceptable alternatives for products/brands out of stock or not carried by the store.	6.14	1.11
7. Provides appropriate justification for the price of more expensive items.	4.23	1.32
8. Is able to correctly determine whether merchandise returned without a sales slip was purchased from _____.	6.32	0.99

Maximum scale score = 5.87 Scale midpoint = 2.94

Initiative - Performance of routine and non-routine job duties without specific requests made by the section leader; willingness to work beyond ordinary job requirements; striving to attain company goals; and a willingness to take independent action.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Periodically rearranges departmental merchandise displays as appropriate without having to be asked to do so.	5.21	1.49
2. Does stock work without having to be asked by the section leader.	4.37	1.45
3. Is creative and comes up with useful ideas to improve the department.	5.99	1.34
4. Straightens up the department regularly without being asked.	5.08	1.37
5. Makes suggestions for moving merchandise from one place to another in the department to increase sales.	6.11	1.22
6. Reports to work early to prepare for the shift.	4.91	1.35
7. Suggests that customers open charge accounts.	5.28	1.44
8. Takes responsibility for making certain decisions when the section leader is away from the department.	5.37	1.29

Maximum scale score = 5.29 Scale midpoint = 2.65

Judgment - Ability to observe and assess routine and unusual situations arising on the job and determine the appropriate action to take.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Can correctly determine whether or not a customer wishes to be waited on.	5.88	1.09
2. Suggests compromises which will satisfy customers returning used merchandise.	5.92	1.19
3. Does not rely on the section leader to make all decisions in the department.	5.73	1.40
4. Expresses opinions to customers when appropriate.	5.78	1.24
5. Calls for assistance from section leaders when it is appropriate to do so.	4.41	1.38
6. Keeps aware of potential shoplifters and notifies security when appropriate.	5.79	1.31
7. Determines the appropriate order upon which to wait on customers.	6.14	1.18
8. Correctly determines when it is appropriate to assist a fellow sales clerk when s/he is having a problem with a customer.	6.07	1.12

Maximum scale score = 5.72 Scale midpoint = 2.86

Employee Relations - Ability to maintain friendly and cooperative relationships with the section leader and other employees on the job.

	<u>Mean</u>	<u>Standard Deviation</u>
1. Gets along well with fellow employees.	4.61	1.21
2. Shows consideration for fellow employees.	5.47	1.33
3. Works well as a team member by sharing job responsibilities.	5.62	1.47
4. Does not grab sales from other sales clerks.	4.70	1.12
5. Is willing to share customers with other employees when appropriate.	4.82	1.41
6. Takes time to answer the questions of new employees.	4.77	1.19
7. Accepts constructive criticism and suggestions when offered by the section leader or other employees without becoming defensive.	5.69	1.40

Maximum scale score = 5.09 Scale midpoint = 2.55

APPENDIX C

Section Leader Performance Appraisal Form

SECTION LEADER

PERFORMANCE APPRAISAL INSTRUCTIONS

Ratings of Performance Dimensions

The following page contains the names and definitions of ten dimensions of job performance identified as being important for the Section Leader job. You are asked to evaluate the job performance of each of the Section Leaders you supervise in each of these areas. You will do this by first reading the description of the each dimension. Next determine the level of the employee's performance in each area over the past year (or however long you have supervised the employee) on the basis of the following scale:

- 1 = Unacceptable
- 2 = Marginal
- 3 = Acceptable
- 4 = Average
- 5 = Good
- 6 = Very Good
- 7 = Outstanding (among the best ever
seen)

On the scale to the right of each dimension, circle the number which corresponds to the level of the employee's job performance in that area. Be sure that you evaluation the employees on each of the ten dimensions listed.

<u>Dimension</u>	<u>Performance Rating Scale</u>
1. <u>Dependability</u> - Fulfills job duties and personal commitments reliably and punctually.	1 2 3 4 5 6 7 -----
2. <u>Knowledge of Job Procedures</u> - Understands and performs job duties according to established procedures. Carries out these duties in an efficient manner with few duties.	1 2 3 4 5 6 7 -----
3. <u>Sales Ability</u> - Presentation of merchandise in a favorable and appropriate manner. Use of proper sales techniques to help customers make purchases.	1 2 3 4 5 6 7 -----
4. <u>Customer Service</u> - Treats customers in a pleasant, courteous fashion and adequately serves the customers' needs.	1 2 3 4 5 6 7 -----
5. <u>Knowledge of Company Policies</u> - Knowledge of Company policy and procedures as related to job duties and responsibilities.	1 2 3 4 5 6 7 -----
6. <u>Product Knowledge</u> - Knowledge of the qualities and features of department merchandise including coordinated accessories, instructions for special care, and manufacturers' warranties.	1 2 3 4 5 6 7 -----
7. <u>Initiative</u> - Performance of routine and non routine job duties without specific requests made by the section leader; willingness to work beyond ordinary job requirements; striving to attain company goals; and a willingness to take independent action.	1 2 3 4 5 6 7 -----
8. <u>Judgment</u> - Ability to observe and assess routine and unusual situations arising on the job and determine the appropriate action to take.	1 2 3 4 5 6 7 -----
9. <u>Employee Relations</u> - Ability to maintain friendly and cooperative relationships with the section leader and other employees on the job.	1 2 3 4 5 6 7 -----
10. <u>Supervisory Ability</u> - Ability to effectively supervise employees and maintain the daily operation of the department. Willingness to accept responsibility and foster a productive atmosphere within the section.	1 2 3 4 5 6 7 -----

APPENDIX D
Job Involvement Scale

Listed below are a series of statements which represent possible feelings that individuals might have about their jobs. Use the rating scale shown below to indicate your feelings about the job in which you are now working.

1. SD means that you strongly disagree with the statement.
2. MD means that you moderately disagree with the statement.
3. SLD means that you slightly disagree with the statement.
4. N means that you neither agree or disagree with the statement.
5. SLA means that you slightly agree with the statement.
6. MA means that you moderately agree with the statement.
7. SA means that you strongly agree with the statement.

Please circle the letter or letters which best describe the degree of your agreement or disagreement with each of the statements below.

- | | | | | | | | |
|--|----|----|-----|---|-----|----|----|
| *1. I'll stay overtime to finish a job even if I'm not paid for it. | SD | MD | SLD | N | SLA | MA | SA |
| 2. You can measure a person pretty well by how good a job he does. | SD | MD | SLD | N | SLA | MA | SA |
| 3. The major satisfaction in my life comes from my job. | SD | MD | SLD | N | SLA | MA | SA |
| 4. For me, mornings at work really fly by. | SD | MD | SLD | N | SLA | MA | SA |
| * 5. I usually show up for work a little bit early to get things ready. | SD | MD | SLD | N | SLA | MA | SA |
| 6. The most important things that happen to me involve my work. | SD | MD | SLD | N | SLA | MA | SA |
| * 7. Sometimes I lie awake at night thinking ahead to the next day's work. | SD | MD | SLD | N | SLA | MA | SA |
| * 8. I'm really a perfectionist about my work. | SD | MD | SLD | N | SLA | MA | SA |
| 9. I feel depressed when I fail at something connected with my job. | SD | MD | SLD | N | SLA | MA | SA |
| 10. I have other activities more important than my work. | SD | MD | SLD | N | SLA | MA | SA |
| 11. I live, eat, and breathe my job. | SD | MD | SLD | N | SLA | MA | SA |
| *12. I would probably keep working even if I didn't need the money. | SD | MD | SLD | N | SLA | MA | SA |

- | | | | | | | | |
|---|----|----|-----|---|-----|----|----|
| *13. Quite often I feel like staying home from work instead of coming in. | SD | MD | SLD | N | SLA | MA | SA |
| 14. To me, my work is only a small part of who I am. | SD | MD | SLD | N | SLA | MA | SA |
| *15. I am very much involved personally in my work. | SD | MD | SLD | N | SLA | MA | SA |
| *16. I avoid taking on extra duties and responsibilities in my work. | SD | MD | SLD | N | SLA | MA | SA |
| *17. I used to be more ambitious about my work than I am now. | SD | MD | SLD | N | SLA | MA | SA |
| 18. Most things in life are more important than work. | SD | MD | SLD | N | SLA | MA | SA |
| 19. I used to care more about my work, but now other things are more important to me. | SD | MD | SLD | N | SLA | MA | SA |
| *20. Sometimes I'd like to kick myself for the mistakes I make in my work. | SD | MD | SLD | N | SLA | MA | SA |

* Indicates that the item was included in the scale used in this study to measure the performance - self esteem contingency conceptualization of job involvement.

APPENDIX E

Organizational Commitment Scale

APPENDIX E
Organizational Commitment Scale

Listed below are a series of statements which represent possible feelings that individuals might have about the company for which they work. Use the rating scale shown below to indicate your feelings about the company for which you are now working.

1. SD means that you strongly disagree with the statement.
2. MD means that you moderately disagree with the statement.
3. SLD means that you slightly disagree with the statement.
4. N means that you neither agree or disagree with the statement.
5. SLA means that you slightly agree with the statement.
6. MA means that you moderately agree with the statement.
7. SA means that you strongly agree with the statement.

Please circle the letter or letters which best describe the degree of your agreement or disagreement with each of the statements below.

- | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|----|----|-----|---|-----|----|----|
| | SD | MD | SLD | N | SLA | MA | SA |
| 1. I am willing to put in a great deal of effort beyond that normally expected to help _____ be profitable. | | | | | | | |
| 2. I talk about _____ to my friends as a great company to work for. | | | | | | | |
| 3. I feel very little loyalty to _____. | | | | | | | |
| 4. I would accept almost any type of work assignment in order to keep working for _____. | | | | | | | |
| 5. I find that my values and _____ are very similar. | | | | | | | |
| 6. I am proud to tell others that I am a part of _____. | | | | | | | |
| 7. I could just as well be working for a different organization as long as the type of work was similar. | | | | | | | |
| 8. _____ really inspires the very best in me in the way of job performance. | | | | | | | |
| 9. It would take very little change in my present circumstances to cause me to leave _____. | | | | | | | |
| 10. I am extremely glad that I chose _____ to work for over other companies I was considering at the time I joined. | | | | | | | |

11. There's not too much to be gained by sticking with _____ indefinitely. SD MD SLD N SLA MA SA
12. I often find it difficult to agree with _____ policies. SD MD SLD N SLA MA SA
13. I really care about the fate of _____. SD MD SLD N SLA MA SA
14. For me _____ is the best of all possible companies to work for. SD MD SLD N SLA MA SA
15. Deciding to work for _____ was a definite mistake on my part. SD MD SLD N SLA MA SA

APPENDIX F

Procedures for Computations of Variance Components and Indexes from Multitrait-Multimethod Matrix Data

Direct Computation of Variance Components

<u>Source</u>	<u>Variance Component</u>
Sales Clerks (s)	$\bar{r}_b + \frac{(\bar{r}_{wd} - \bar{r}_b)}{n} + \frac{(\bar{r}_{wf} - \bar{r}_b)}{m}$
S x Dimension (d)	$\bar{r}_{wd} - \bar{r}_b$
S x Rating Format (f)	$\bar{r}_{wf} - \bar{r}_b$
Error	$1 - \bar{r}_{wd} - \bar{r}_{wf} + \bar{r}_b$

Mutlitrtrait-Multimethod Matrix Indexes (Intraclass Correlations)

<u>Source</u>	<u>Index</u>
Sales Clerks (S)	$\frac{VC_s}{VC_{sc} + VC_e}$
S x Dimension (d)	$\frac{VC_{s \times d}}{VC_{s \times d} + VC_e}$
S x Rating Format (f)	$\frac{VC_{f \times s}}{VC_{f \times s} + VC_e}$

Note: \bar{r}_b = average correlations of heteromethod-heterotrait triangles; \bar{r}_{wd} = average correlation of validity diagonals; \bar{r}_{wf} = average correlations of monomethod-heterotrait triangles; n = number of dimensions; m = number of rating formats. VC = variance component.

VITA

Steven Ronald Gordon was born on February 28, 1951 in New York City, New York. In 1968, he graduated from Elmont Memorial High School. That same year he enrolled at Cornell University in Ithaca, New York where he majored in Psychology and graduated in 1973. He received his Master of Arts degree in Clinical Psychology from the State University College of New York at New Paltz in 1975. He then enrolled for further graduate study at the University of Tennessee in the Industrial and Organizational Psychology Program. He married Leslie Anne Acomb in 1981. He was awarded a Doctor of Philosophy degree with a major in Industrial and Organizational Psychology in 1986. He currently resides in Atlanta, Georgia where he is a manager in the Human Resources Research department of BellSouth Corporation and president of Assessment Resources, a management consulting firm.