This is to certify that the

dissertation entitled

THE EFFECTS OF GROUP VARIABLES ON EVALUATIONS OF INDIVIDUALS IN ASSESSMENT CENTER GROUP EXERCISES

presented by

Jeffrey Robert Schneider

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Psychology

Major professor

Date September 18, 1992

THE EFFECTS OF GROUP VARIABLES ON EVALUATIONS OF INDIVIDUALS IN ASSESSMENT CENTER GROUP EXERCISES

by

Jeffrey Robert Schneider

A DISSERTATION

Submitted to
Michigan State University
Department of Psychology
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Psychology

1992

ABSTRACT

THE EFFECTS OF GROUP VARIABLES ON EVALUATIONS OF INDIVIDUALS IN ASSESSMENT CENTER GROUP EXERCISES

by

Jeffrey Robert Schneider

The goal of this research was to examine possible influences on the ratings of an individual assessee in assessment center group exercises. A general system of variables was presented to facilitate a better understanding of the influence of situational and behavioral variables on the ratings of an assessee in a group exercise. The system included assessees' behavior, exercise characteristics, rating systems, and assessor cognitive processing variables. A study investigating some of the components of the system is also presented.

The study examined the extent to which assessor cognitive processes mediated the influence of group composition, group performance, and interdependence among group members on ratings of an individual in a group exercise. The assessors were 167 undergraduate psychology students who observed their peers in videotaped, group exercises. Assessors rated a target assessee whose behavior was held constant across all observations in a two (group composition--high versus low non-target performers) by two (high vs. low group performance) by two (high vs. low interdependence) experimental design with an appended control group.

Assessors' attributions about the target partially mediated the relationship between group composition and ratings of the target. The target was attributed higher responsibility for the group's performance when the non-targets' performance was manipulated at a low level. The target was rated higher when more responsibility for the group's performance was attributed to him. Other hypotheses were not supported. There was no evidence of main effects of group composition or group performance and no significant interactions involving group composition and either group performance or interdependence on ratings. Group composition effects were not mediated by assessors' perceptions of the target's performance relative to the non-targets' performance. A general impression of the target as a leader did not mediate the relationship between group performance and ratings.

The findings offer promise for a better understanding of situational influences on assessment center ratings through modeling the cognitive process of rating, and some future research is suggested. Reasons for the lack of support of other hypotheses are also discussed with the intent of stimulating future research.

ACKNOWLEDGMENTS

I consider my years at Michigan State University to be four of the most formative years of my life. This dissertation represents something on which I worked very hard but represents only a part of the learning that occurred in the four years. I acknowledge the help of the faculty, Neal Schmitt, Dan Ilgen, Steve Kozlowski, Kevin Ford and Mike Lindell, who treated me with respect and supported me intellectually and financially in the pursuit of this degree. Thanks to Neal especially for all his help, patience, and friendship, and thanks to my parents for their continual support.

I could not have asked for better timing than to converge on East Lansing at the same time as Stephen, Mickey, Sherry, Jose, Ron and Sandy. I have great respect for all of you for your intelligence, drive, and ability to have a good time. You are a great group of friends, making every day special.

Above all, feelings of thanks, love, and joy go to Linda, from whom I have learned the most. You have given me a sense of balance, emotion, and spirituality that provided me with the energy for this work and that gave meaning to my life in a way that I have never before experienced.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

INTRODUCTION

The use of assessment centers has grown rapidly since their predictive validity for managerial jobs was first demonstrated (Bray & Grant, 1966). Assessment centers are now used for selection, placement, early identification of management potential, and employee development in large and small, manufacturing, government, educational, military, and service organizations (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Klimoski & Brickner, 1987). Given their widespread use, assessment center results are used in important decisions regarding the careers of thousands of people annually (Thornton & Byham, 1982).

What is an Assessment Center?

Perhaps the most defining characteristic of an assessment center is the use of multiple *assessors* to observe *assessees'* behavior in multiple *exercises* and to rate assessees on multiple *dimensions* (Task Force on Assessment Center Standards, 1989). *Assessees*, the persons being evaluated in the center, participate in a number of different exercises in which their behavior is observed and rated. *Exercises* are constructed either to simulate critical elements of the task domain or to simulate whole tasks from a focal job. For example, exercises for a managerial job commonly require assessees to run a meeting, handle a stack of paperwork, or give a formal presentation. *Assessors* are the persons who observe and rate assessees' behavior. Assessors are typically professionals such as psychologists and other human resource

specialists or members of the organization such as job incumbents or their supervisors. Professional assessors work frequently and evaluate many different individuals for various jobs. Organizational members serve as assessors in their organization's assessment program on a rotating basis and may only serve once or twice a year and two to three times over their tenure (Schmitt, Schneider, & Cohen, 1990). *Dimensions* are person- or task-oriented constructs verified as job-relevant through job analysis. It is common for dimensions to reference psychological constructs such as leadership, social orientation, creativity, and intelligence. Dimensions may also reference classes of skills such as problem analysis, planning, organizing, delegation, or written communications.

The procedure in an assessment center is as follows. First, assessors observe assessees' behavior in some way including observing behavior in "live" or videotaped exercises or reading the assessees' written products from an exercise. Based on these observations and with the assistance of scoring guidelines, assessors rate assessees on the dimensions. The dimension ratings serve as the basis for making selection, promotion, classification, or training decisions.

Using Group Exercises to Evaluate Individuals

Group exercises are commonly used to evaluate individual skills and abilities in assessment centers. Thornton and Byham (1982), sampling over 500 assessment centers, found that assigned role leaderless group discussions were used in 85% of centers, second only to in-basket exercises. Non-assigned role leaderless group discussions occurred in 45% of centers in their sample, and management games were used in 35% of centers of their sample. More recent statistics on the frequency of

use of different assessment center exercises suggest that group discussions continue to be a frequently used exercise form with assigned role group discussions occurring in 44% and non-assigned role group discussions occurring in 59% of a comparable sample (Thornton, 1992). In group exercises, assessees perform in small groups to solve organizational problems, generate financial and strategic organizational plans, and to manufacture mock products. The groups are frequently made up of other assessees.

The present research examines whether evaluating individuals in assessment center groups is a wise practice. My hypothesis is that group exercises in which assessees participate and are evaluated with each other provide a noisy and variant context for the assessment of individuals. The performance of a particular assessee in a group is apt to affect and to be affected by the performances of other assessees. Ultimately, the performance achievements of the group as a whole are determined by the individual assessees and their combined efforts. Assessors attempting to rate particular individuals may be influenced by their observations of the performance of other individuals or by observations of the group's performance achievements. This is particularly concerning since no attempts are typically taken in assessment centers to control for group assignments. Depending on the group to which a particular assessee is assigned, his or her ratings may be higher or lower than would be predicted on the basis of independent measures of his or her skills and abilities. Further, assessment center ratings have been shown frequently to correlate highly within an exercise but not across exercises (e.g. Bycio, Alvares, & Hahn, 1987, Sackett & Dreher, 1982; Turnage & Muchinsky,

1982). This finding of "situational specificity" of assessment center ratings may be partly a function of using group exercises in which variation of elements in the rating context may influence the ratings.

Overview of The Present Study

The present study investigates the practice of evaluating individuals in assessment center group exercises. Particular emphasis is placed on the examination of how assessors' ratings of an assessee are influenced by the simultaneous observations of the performance of other assessees in a group exercise. The effects of knowledge of group performance and of exercise characteristics are also examined.

Group composition, group performance, and exercise characteristics--the focal variables in this research--are apt to affect ratings of individuals in assessment centers by two avenues. On one hand, group variables can directly affect how the characteristics of individuals combine and are behaviorally manifest in group exercises. On the other hand, since the method of evaluation is performance rating, group variables may influence ratings via assessors' observation and processing. In other words, ratings of assessees in groups are apt to reflect assessors' perceptions of and beliefs about group variables. Though my review acknowledges how group variables can influence ratings by both avenues, the review and present study is focused on assessor effects. In the study, assessors viewed videotaped assessees whose behavior was scripted to create different group compositions. Assessors were also exposed to different information that shaped their perceptions of the level of achievement of the group and their perceptions of task characteristics. Assessor ratings were examined for the hypothesized effects of group composition, group performance

achievements, and task characteristics.

In the review that follows, I make a case for my choice of these group variables, concluding with specific hypotheses about their effects. In building my case, I first present a categorization system summarizing a number of variables that may affect ratings in an assessment center group exercise. This system was built from previous assessment center, group, performance rating, and social cognition research. The system was intended not only to guide my choices about the variables for this study, but also to generate future research, and to guide the choices of center designers.

Purposes of This Research

I hope to make the following contributions with the present research. First, I hope to address the issue of context effects on judgment in assessment centers by importing theory and findings from social perception and judgment research, group research, and performance rating research. These literatures enabled me to bring a theoretical perspective to resolving issues in assessment centers, an area in which research has been dominated by atheoretical validity studies (Thornton, 1991).

Second, and on more practical grounds, no steps are typically taken in assessment centers to control for the assignments of individuals into group exercises. Yet as this research proposes to demonstrate, the evaluation of individuals in groups may be influenced by variables in the group to which an assessee is assigned.

Third, the research was aimed at understanding the problematic empirical findings suggesting that dimension ratings reflect the situational influences and not individual difference constructs (e.g.

Bycio et al., Sackett & Dreher, 1982; Turnage & Muchinsky, 1982). Exercise effects represent a commonly debated (e.g. Neidig & Neidig, 1984 vs. Sackett & Dreher, 1982; Dreher & Sackett 1981 vs. Norton, 1977) and puzzling (Klimoski & Brickner, 1987) problem in assessment centers. Examining the influence of context on raters' judgments may assist in the understanding of the exercise effect findings.

Fourth, the problem of assessing individuals in groups is likely to become more pervasive as more organizations shift to more team-oriented work designs while maintaining individual appraisal and compensation systems. The proposed research looks at similar kinds of issues in assessment center group exercises--a specific application by organizations in which individuals are evaluated as they work in groups.

Group Variables Affecting Ratings of Individuals

In this section, I identify a system of variables that can affect the ratings of an individual in an assessment center group exercise. I developed this system in order to illuminate the many, previously unconsidered issues involved in assessing individuals in groups. In the present research, I tested some of the components of the model so the system also provides the broader context within which the study is conducted. A general description of the variables in the system and their relationships are described in the first part of the introduction section. A more detailed discussion of the variables from the model that are examined in the study follows in the second half of the introduction.

Variables Affecting Performance in Groups

In order to better understand the possible influences on the performance and ratings of assessees in groups, I first consulted models of group performance. A number of these models of have included individual, group, and environmental level variables as determinants of group performance (Collins & Guetzkow, 1964; Hackman 1987; Hackman & Morris, 1975; McGrath, 1964). Examples of individual-level variables included characteristics of individual group members such as ability, skills, and personality characteristics (Hackman, 1987). Examples of group-level variables include group structure, cohesiveness and size (Hackman, 1987). Examples of environmental-level variables included task characteristics, reward structure, and stress. Group performance was postulated to result from the combined influence of the variables at these three levels.

Some of the components from these general models of group performance are applicable to assessment center group exercises. The characteristics of the individual group members interacting with the exercise tasks are considered here as the key determinants of assessment center group performance. Thus, the system I have developed for assessment center group exercises, depicted in Figure 1, includes exercise characteristics and individual characteristics as the key determinants of group interaction and ultimately of group performance. I also have added the influence of assessor cognitive processes to the system, acknowledging that ratings are a function of assessee performance as viewed and evaluated by assessors. Ratings, the critical output of assessment center exercises, are considered in this research to be not only a function of assessee performance but also a

product of assessor cognitive processing. Figure 1 is intended to be a heuristic for stimulating research questions about assessment center group exercises. The variables and pathways are illustrative of possible effects but are in no way intended to be a causal model. Specific hypotheses follow at the end of the introduction. In the remaining paragraphs in this section, I explain the variables depicted in Figure 1 and their effects on assessee performance and rating.

Characteristics of Individual Members

Emphasis has been placed exercise behaviors as observable indicators of an assessee's knowledge, skills, abilities, or personality characteristics. Assessees' exercise behaviors are thought to be samples from which inferences about the assessees' enduring traits or behavioral tendencies can be made (Thornton & Byham, 1982). As a result, I have included individual characteristics such as behaviors, knowledge, skills, abilities, and other characteristics. Some variables such as motivation, race, and gender have not previously been included.

Motivation has been assumed to be constant and high for all assessees because job offers or promotions are linked to performance in the center (Arvey, Strickland, Drauden, & Martin, 1990; Sackett, Zedeck, & Fogli, 1988). I have included motivation in the model because I do not think it can be assumed to be high for all assessees. Not all assessment centers are used for selection (Thornton & Byham, 1982), and centers used for research or developmental purposes may result in different patterns of assessee motivation (Arvey, et al., 1990).

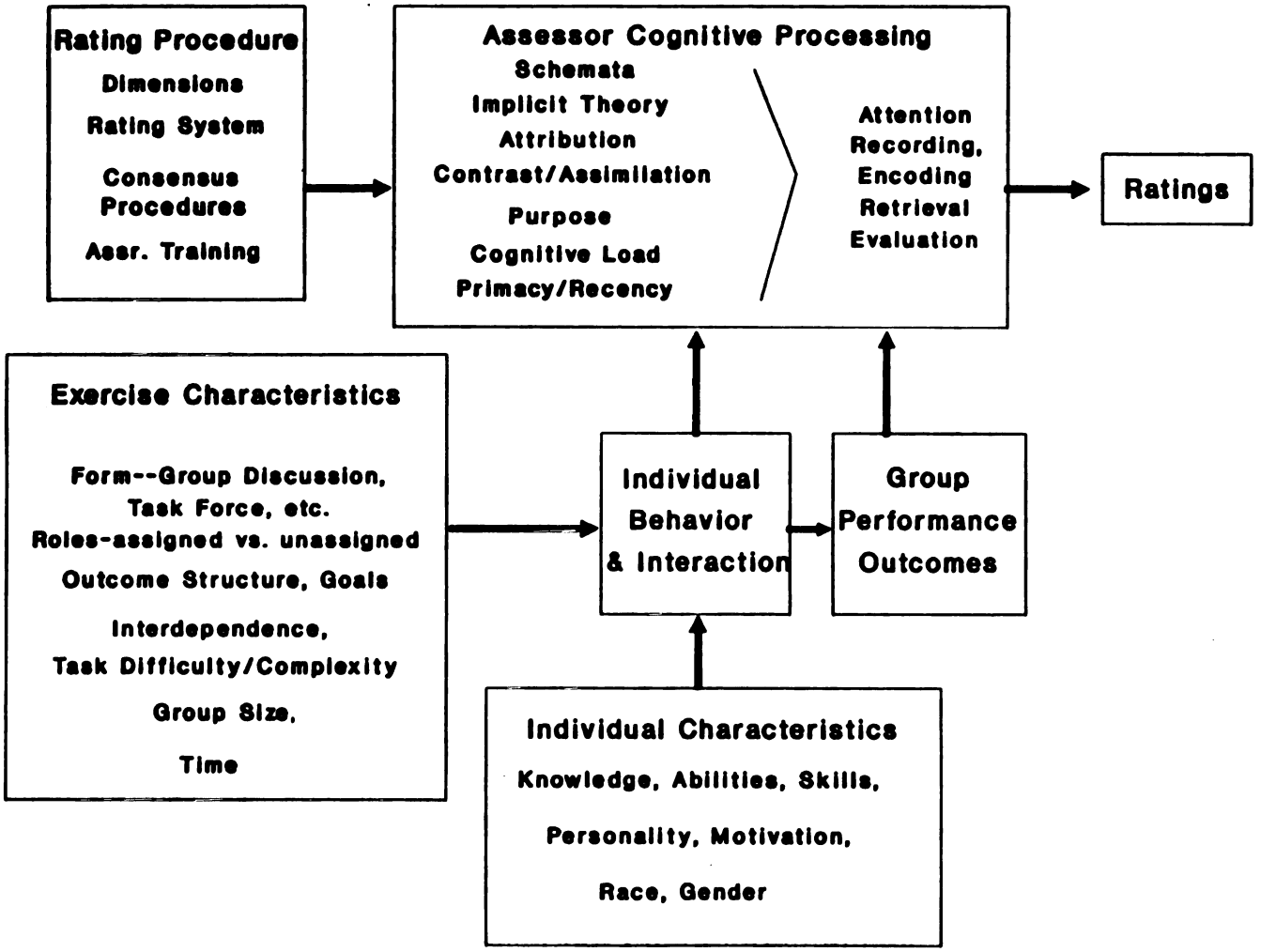

Figure 1:

A Systems View of Ratings in Assessment Center Groups

Race and gender effects have been commonly examined at the center-level for adverse impact and EEOC considerations, but race and gender effects associated with group assignments are rarely evaluated. An exception is Schmitt and Hill (1977) who found that ratings of African American women for some dimensions were correlated negatively and significantly with the number of Caucasian males in the group, and ratings of Caucasian males tended to be higher when the number of white males in a group increased.

Exercise Characteristics

The system in Figure 1 also acknowledges the influence of exercise characteristics on the performance of assessee groups. Traditionally, exercises have been portrayed as vehicles for measuring individual difference dimensions but not as determinants of or sources of influence on assessee performance. Empirical findings have suggested, however, that the association between exercises and ratings is considerable (Archambeau, 1979; Bycio, Alvares, and Hahn, 1987; Neidig, Martin, & Yates, 1979; Robertson, Gratton, & Sharpley, 1987; Sackett & Dreher, 1982; Silverman, Dalessio, Woods, & Johnson, 1986; Turnage & Muchinsky, 1982). Thus, the inclusion of exercise characteristics as an influence on performance and on ratings is consistent with these empirical findings.

Little guidance was available for determining the kinds of variables to include in the system as exercise characteristics. With the exception of Schneider and Schmitt's (1992) form and content and Russell's (1985) managerial role, very little has been done to classify exercises according to their characteristics. The system developed in the present research uses Schneider and Schmitt's dimensions, form and

content, as a starting point for developing a list of exercise characteristics that can affect assessors' ratings.

Choices about assessment center exercises tend to be driven by job analysis results, but there are a set of exercise prototypes that have a long history of use. Schneider and Schmitt referred to these different exercise prototypes as exercise forms. Common exercise forms include the in-basket exercise, the situational interview, and the case analysis (Thornton, 1992). Commonly used group exercise prototypes or forms including the group discussion, the task force, the management game, and the manufacturing exercise (Thornton, 1992).

Once the choice of a particular exercise prototype has been made, center designers may choose to assign or not to assign roles. Thornton (1992) has noted advantages and disadvantages of assigning a role such as group leader to one assessee in a group discussion exercise. Thornton suggested that leaderless groups are rare in organizations and that assigned leader group exercises likely have greater content validity for managerial jobs than non-assigned group exercises. Negatively, assigned leader exercises are restricted by the high costs involved in running enough groups so that all assessees have a chance to play the leader role. Following these traditional exercise prototypes, I have included roles (assigned vs. non-assigned) in addition to form as variables by which assessment center exercises differ.

Center designers also have choices about how to define the relationships among the assessees in groups. Thornton & Byham (1982), for example, noted that assigned role and non-assigned role group discussion exercises have typically been designed with the

relationships among group members as cooperative or competitive. In the typical competitive exercise, assessees are given the assignment of securing the most of a scarce resource such as a grant or budgetary allocation for themselves (Thornton, 1992). In the typical cooperative exercise, assessees are given assignments that encourage them to participate with each other and arrive at solutions to problems by consensus.

Schneider and Schmitt (1992) proposed that three dimensions determined the relationships between participants in an exercise. Their three characteristics included--outcome structure (Deutsch, 1949), task interdependence (Miller & Hamblin, 1963), and goal interdependence (Mitchell & Silver, 1990). In Schneider and Schmitt's study, outcome structure was varied by rules imposed upon the assessees. In the cooperative outcome structure, assessees were told to work together to arrive at the best solution, and in the competitive outcome structure, assessees were told to get the most of a scarce resource (grant money in one exercise and personnel in another) as possible for themselves. In the high task interdependence condition, different information that was important for accurate problem solving was distributed to each assessee in a group, and in the low task interdependence condition, all assessees received the same information. In the high goal interdependence condition, assessees were given group goals, and in the low interdependence condition, assessees were given individual goals. Following Schneider and Schmitt, I have included outcome structures, interdependence, and goals as variables according to which assessment center group exercises can be distinguished.

Task difficulty/complexity, group size, and time have been added to the model even though they have not previously been considered in assessment center research or exercise design. Task difficulty/complexity has been found to be significantly associated with group performance criteria. For instance, Hackman (1968) found that groups received higher evaluations for originality in high task difficulty conditions and higher evaluations of quality and accuracy in low task difficulty conditions.

Time has been shown to influence group process and task-related behaviors. Sorenson (1971) concluded that short duration groups eliminated alternatives more quickly and efficiently in approaching a solution, but long duration groups had potential to be more creative because of more time to process alternatives in detail. Further, Gersick (1977) concluded that all groups progress through similar stages in arriving at a solution, and that individuals' perceptions of the time allotted to the task was an influential variable in pacing the group through the stages.

Increases in group size have been shown to have positive and negative effects on groups. On one hand, as group size increases, the range of skills, abilities, and knowledge that a group can bring to bear to problems likely increases and improves performance potential (Levine and Moreland, 1990). On the other hand, group members may be more limited in their opportunities for expression in larger groups and feel more opportunities for involvement in smaller groups (Hackman & Vidmar (1970). Goodstadt and Kipnis (1970) found that satisfactory performance of specific employees was less likely to be recognized by observers and rewarded as groups got larger.

In sum, exercises have been traditionally designed to resemble one or more assessment center exercise prototypes or forms. Roles, outcomes, goals, and relationships among assessees can be varied. Group size and time can affect assessee performance and their ratings, yet size and time are constant within centers, and do not vary a great deal across centers.

Assessor Cognitive Processing

The position taken in present research is that assessment center ratings are a function of social perception and cognitive processes like all social judgments. The effects of these processes cannot be overlooked when attempting to understand trends in assessment center ratings. Therefore, I have included assessor cognitive processing variables in the system. First, I have incorporated a stage model of processing into the system including observation, encoding and storage, recall, integration, and rating (e.g. DeNisi, Cafferty, & Meglino 1984; Landy & Farr, 1980). Secondly, I have incorporated a set of cognitive processes that have been shown in past research to influence social judgments. The processes include attribution, primacy/recency, contrast/assimilation. In addition, assessor information processing capacity or cognitive load and implicit theories about and prototypes for performance are important factors in the assessment center evaluation process. All of the cognitive processing variables are reviewed in the following paragraphs.

Stages of Information Processing

A number of researchers have proposed that performance ratings occur according to the following stages: (a.) observation of behavior, (b.) encoding and storage, (c.) retrieval (d.) integration, and (e.)

rating (e.g. DeNisi, et al. 1984; Landy & Farr, 1980). Stage models have generated a great deal of research, including many laboratory studies examining cognitive-process hypotheses and generalizing to the performance appraisal context (See reviews such as Ilgen, Barnes-Farrell, & McKellin, in press). Much of this research has been conducted with subjects viewing short performance segments on video tape and making ratings immediately or within a week of observation. Thus, while these studies have been framed as performance appraisal research, the rating tasks highly resemble the tasks of an assessor. Many of the findings of performance appraisal research should generalize to assessment centers so a repeat of all performance appraisal studies framed as assessment center research is not necessary. There are some differences between performance appraisal and assessment center research purposes and methods making it important to conduct some research within the conceptualization of assessment centers and to follow assessment center methods strictly. Some of the key differences are described in the following paragraphs.

First, memory plays a very different role in performance appraisals and assessment centers. Performance appraisal rating periods can extend over several months which can become quite taxing on memory and recall. Acknowledging this, researchers have suggested that raters use diaries to reduce over reliance on memory (Denisi, Robbins, & Cafferty, 1989). Memory processes are likely to operate very differently in assessment centers. First, the time interval between the end of an exercise and rating is no longer than a week when consensus procedures are used and may be a matter of a few minutes when practice dictates that performance is rated immediately after an exercise. Even in cases

where the delay is as much as a week, assessors can rely on notes taken during an exercise or other aids when integrating and evaluating assessees' performance. As such, attention likely plays a greater role than memory in determining which behaviors are considered at the time of rating.

Although the memory processes are clearly different, recording of behaviors is not sufficient enough to eliminate biases in rating. Assessors are not perfect recorders, and they may need to rely on notes to prompt memory or may rely solely on memory in cases where notes are not highly detailed. Assessors' attention to behavior and the notes taken may be subject to cognitive biases. For example, Balzer (1986) found that raters tended to record information that was most inconsistent with expectations set by an eight minute videotape viewed prior to the experimental stimulus. Thus, though memory may not be taxed as extensively in assessment centers as in the performance rating context because assessors take notes, assessment centers are by no means insulated from cognitive rater biases.

Second, Denisi et al. (1984) noted that raters' preconceived notions or impressions of the ratee can influence how raters attend to ratees' behavior. One explanation offered by DeNisi et al. and based on Wyer and Srull (1981) suggested that preconceived notions may prime frames of reference, schemata, or prototypes that guide the interpretation of subsequent information. In assessment centers, assessors are shielded from the influence of preconceived notions of an assessee because assessors' assignments are typically made so that they do not rate familiar assessees (Thornton & Byham, 1982). There is still room, however, for the influence of other assessees' performance

in creating expectations for the performance of the observed assessee (i.e. a between ratee effect; Murphy, Balzer, Lockart, & Eisenman, 1985). Thus, prototypes primed by other assessees' behavior may bias assessors' ratings even though assessors rarely have previous experience with or exposure to the assessee.

Third, the performance appraisal context often involves competing demands on raters' time and attention (DeNisi, et al., 1988), in cases where supervisors do not work in the same work space as the ratee, they may not even have opportunity to see behaviors. Assessors are afforded the opportunity to focus their attention on the exercises and particularly the assessees they are assigned to rate. This is a clear advantage offered to assessors, but it has not completely eliminated concerns about overloading assessors (Gaugler & Thornton, 1989).

A Social-Cognitive Perspective on the Evaluation of Assessee Behavior

My position is that assessors' judgments are like any other social judgments in which it is postulated that perceivers' values, needs, and expectancies influence the outcomes of perceptions in ways that cannot be entirely explained by stimulus qualities (Higgins & Bargh, 1987). Research aimed at understanding perception has resulted in some discoveries about how attention, memory, and evaluation processed influence and are influenced by sources other than observed data. The topics that are covered in this section include schema models of memory and information processing, implicit theories of personality and performance, attribution theory, contrast and assimilation effects, attribution, primacy and recency, and cognitive load. I have dealt with these topics separately in the text only for the purposes of organization. I acknowledge that there is a great deal of overlap

across the topics.

Schema Models of Memory and Information Processing

Schema theorists have suggested that encoding and storage is heavily determined by guiding schema or knowledge structure that select and modify experience to arrive at a coherent, expectation-confirming, and knowledge-consistent representation of an experience (Alba & Hasher, 1983). Schema theorists have emphasized the incompleteness of and distortions in memory and tend to reject the proposals of theorists who maintain that memories for specific events are preserved in traces which are accurate and stable over long durations (Bartlett, 1932).

Although schema-based theories have been proposed and applied frequently to explain research findings in perception and memory, it is acknowledged that there is a great deal of variance in the way schemata and schema-theories are described (Alba & Hasher, 1983). For the sake of proceeding with some understanding of schemata and how they work, a "modal" schema theory offered by Alba and Hasher (1983) will be described here.

Alba and Hasher defined schema as a collection knowledge that a person possesses about a particular domain and that allows for the encoding, storage, and retrieval of information related to that domain. They suggested that from any environmental event, only the information that is relevant and important to the activated schema will be encoded. Information is not remembered totally in its occurring form, however. The semantic content of the information is abstracted and interpreted in such a way as to be consistent with the schema. The operation of these processes is likely to result in a representation that is less than totally accurate. In some cases, at the time of recall, a

reconstruction process may occur in which a few recalled details are combined with a schema to fabricate what might have happened.

It is also important to note that research by Hastie and others' adaptations that have followed (Hastie & Kumar, 1979, Hastie, 1980, Srull, 1981; Wyer, Bodenhausen, & Srull, 1984; Wyer & Gordon, 1982) have disputed the postulate that schema driven information processing results in the retention of schema-consistent information and the discarding of schema-inconsistent information. In studies by these researchers, information that was inconsistent with a primed schema was recalled with fewer intrusions (i.e. more accurately and with less fabrication) than schema-consistent behavior. Offering an integration of the divergent findings, Higgins and Bargh (1987) have emphasized the importance of differentiating between impression testing and impression formation. They suggested that when an impression is well formed, consistent information is apt to be better recalled with inconsistent information being discarded as not being representative of the target. When the goal of the observer is to form an impression, however, there is a greater likelihood that observers will attend to inconsistent behavior that distinguishes the target from its context.

Over the past decade, there has been a large increase in the use of cognitive/schema models of information processing to guide research in human factors, performance appraisal, and motivation (Lord & Maher, 1989). For example, halo error in performance appraisal may be explained by understanding that memory of employee performance consists of general impressions of the employee organized according to pre-existing categories and not of specific, accurate representations of behavioral occurrences (Feldman, 1981; Lord, 1985; Murphy & Balzer,

1986). Explorations of the sources of halo error have also examined the extent to which raters possess implicit notions of covariation among dimensions or conceptual similarity schema that results in high intercorrelations among performance dimension ratings (Cooper, 1981; Kozlowski, Kirsch, & Chao, 1986).

Very little research has been conducted in assessment centers from a social-cognitive perspective. A lack of discriminant validity of dimension ratings (i.e. halo) has been a cause of concern in assessment center research (e.g. Bycio, et al., 1987; Sackett & Dreher, 1982), but suggestions that this may be a result of schema-based processing or implicit schemata of performance (e.g. Bycio, et al., 1987; Turnage & Muchinsky, 1987) have not been followed up with research.

A good example of the application of social-cognitive theory has occurred in leadership research. A number of leadership studies have found that raters combine behavioral information and knowledge of performance into a general impression which mediates the relationship between observed data and ratings (Rush, Phillips, & Lord, 1981; Phillips & Lord, 1981). Transporting these findings to assessment centers, schemata may guide not only the processing of assessees' behavior but also other group-related information available for assessors' observations. This and other possible influences on assessment center ratings is pursued further in the sections that follow. The next section covers specific cases of schema driven processing of information where the guiding schema is an implicit theory of performance or personality.

Implicit Theories of Personality and Performance

Much research has supported an understanding that raters' processing of social information is not only driven by observed data but also by the theories of personality and performance that they possess. For example, Cantor & Mischel (1977) found support for their hypothesis that memory structures representing personality dimensions guide the processing of behavioral observations. Lastly, studies on leadership have offered evidence that there is some consensus about the dimensionality of leadership performance across raters and that raters' "implicit theories of leadership" affects how raters process and rate behavior (Eden & Leviatan, 1975; Rush, Thomas, & Lord, 1977).

Assessors' theories of personality or performance are apt to influence their attention to and evaluation of assessees' performance. This is acknowledged implicitly in assessment centers since assessors receive extensive training in an attempt to standardize the way they view the dimensions of performance. It is thought that assessors employ this dimensional schemata of performance to categorize and rate the behavioral data they observe. Assessors and their ratings, however, are also likely to be influenced by other sources. For example, despite their training, assessors may employ idiosyncratic understandings of dimensions in the processing of information. Performance appraisal theory and research seems applicable to this issue. Ostroff (1985) proposed a similar hypothesis about performance appraisal ratings. She predicted that accuracy of appraisals would be improved as performance appraisal dimensional systems were better matched with raters' implicit categorization systems, and she found that raters' ability to dimensionalize behaviors in a manner consistent

with the rating scale was related to accuracy.

Assessors may possess other beliefs that influence how they process behavioral information. For example, leadership research has suggested that observers may infer leadership from their observations of results because of a theory that leaders cause results (Calder, 1977, Pfeffer, 1977). Research has shown that knowledge of group performance may cue raters' implicit leadership theories with the result that ratings of group leaders or members on a variety of process and outcome scales will be consistent with their knowledge of performance (Larson, 1982; Larson, Lingle & Serbo, 1984; Phillips, 1984; Phillips & Lord, 1981; Rush, Thomas & Lord 1977). Extending these findings to assessment centers, it follows that assessors' ratings of individuals in a group exercise may be influenced by the observations of the results that a group achieves. Though assessors are not given anchors for judging the effectiveness of a group, it is likely that assessors who view a number of groups during their tenure are apt to begin to develop a notion of the distribution of group performance outcomes. Their perceptions of how the group performs versus other groups may influence assessors' attributions about and evaluations of the group members. I will revisit this issue of how knowledge of group performance can influence assessors' ratings in more depth in a later section.

Attribution Theory

Attribution theory provides a framework for understanding all perception (not just perceptions of leadership) as a process of inferring causes from the observations of outcomes. Attribution theories have been applied to a variety of content domains including performance ratings. Performance appraisal researchers have suggested,

for example, that a determination of whether the person or situational factors were the cause of an event or behavior is one of the first stages in rater controlled processing of performance information (Denisi, et al., 1984; Feldman, 1981; Ilgen & Feldman, 1983). Further, attributions have been found to mediate the relationship between observed data and ratings or conclusions drawn about observations (Green & Mitchell, 1979, Phillips & Lord, 1981). A number of postulates and principles of attribution theory have emerged. These include the augmenting principle and the discounting principle, (Phillips & Lord, 1981) and actor-observer biases (McElroy & Downey, 1982; Mitchell, Larson & Green 1977). These and other attribution principles that are relevant for assessment center ratings have been included in the discussion that follows.

Kelley (1972) proposed a covariation model of attribution suggesting that persons attribute outcomes to stable causes or patterns of causes by a determination of the distinctiveness, consistency, and consensus of the outcome. In cases where evidence of an event's distinctiveness, consistency, or consensus is not available, people rely on other strategies such as the discounting or augmenting principles. The discounting principle suggests that a potential cause is discounted based on the extent to which alternative causes are available and reasonable (Fiske & Taylor, 1991). The augmenting principle describes a case where a cause is judged to be a facilitative cause of the event even in the face of other inhibitory causes which interfere with the occurrence of the event.

Phillips and Lord (1981) illustrated how these attribution principles apply to observing and rating others in their work on

leadership. They suggested that raters having knowledge of performance outcomes are apt to search for a causal agent or leader to whom they attribute responsibility for the group's success or failure. Their hypothesis was that more leadership qualities would be attributed to leaders under augmenting conditions such as when subordinates are perceived as low in ability or motivation. Fewer attributions of leadership were predicted in discounting conditions where highly able and highly motivated subordinates could be interpreted as the cause of group performance. In other words, leaders would receive higher ratings on leadership scales depending on the extent to which their subordinates are or are not perceived as plausible causes of the group's performance. Observed data was predicted to be mediated by attributions.

A number of attribution errors or biases may also be relevant. First, the fundamental attribution error is the tendency to over-attribute outcomes to others' dispositional qualities and under-attribute outcomes to situational factors when observing others (Fiske & Taylor, 1991). Thus, when success or failure occurs in a group, observers tend to look for a person and not situational factors to blame. The actor/observer bias is closely related. It suggests that persons, searching for causes of their own behavior (i.e. as actors), tend to rely on situational and not stable, dispositional explanations. When observing others, persons have a tendency to see the causes of behavior as stable and dispositional and not situational (Fiske & Taylor, 1991). The attribution of responsibility/blame bias is a tendency of observers to attribute the responsibility for failure to the stable, internal characteristics of the observed person rather than

to situational factors. Observers are more likely to attribute responsibility/blame when there is an identifiable source of action, when they believe that the person should have foreseen the outcome, and when they perceive that the person's actions were volitional and not justified by the situation (Fiske & Taylor, 1991). Further, the magnitude of the motivation to attribute blame is positively associated with the extent of negative consequences associated with the outcome (Fiske & Taylor, 1991). In an associated theory, "Just World Theory," Lerner (1970) suggested that humans have a tendency to attribute others' misfortune to internal, stable causes to preserve a belief that the world is just and that others are responsible for their own misfortunes.

Many of these theories of attribution and attribution biases are relevant to rating in an assessment center. Assessors are taught in assessor training that assessees' behaviors are indicators of their underlying traits or dispositions. This likely reinforces assessors' (as observers) attributional tendency to consider person rather than situational factors as the causes of performance. Thus, when witnessing group performance achievements, assessors are apt to search for a particular causal agent who was the source or initiator of this performance. An assessee assigned the role as leader is a particularly salient agent. In a group exercise, group members may serve as competing (augmenting or discounting) causes for the performance but assessors will likely ascribe the performance to the assessee they see as the most likely cause. Further, when assessment center groups fail or have low levels of achievement, attribution of responsibility/Just World Theory suggests that assessors are apt to search for an assessee

as the cause of failure and make inferences about assessee's dispositional characteristics on the basis of the failure.

In addition to augmenting or discounting principles of attribution theory, contrast and assimilation effects may also be an explanation why the behavior of other assessees observed at the same time may influence the ratings of an individual in a group. The potential for contrast and assimilation effects on ratings of individuals in assessment center groups is briefly discussed in the paragraphs that follow.

Contrast and Assimilation Effects

Contrast and assimilation effects have been used to describe visual perception (Helson, 1964), the perception of weight (Sherif, Taub, & Hovland, 1958), and social judgments (Holmes & Berkowitz, 1961; Hovland & Sherif, 1952). In simple terms, assimilation effects describe cases where evaluations of the person are biased toward the standard. Contrast effects describe cases where a novel stimulus is biased away from the standard. Assimilation (bias toward) or contrast (bias away) is typically judged in research by a comparison of the rating a target receives in the presence of a standard stimulus and the rating a target receives in the absence of the same standard stimulus (Maurer & Alexander, 1991). Contrast and assimilation are relevant to the process of observing and evaluating participants in assessment centers and other organizational evaluations. In fact, a body of research on selection decisions suggests that contrast and assimilation explain trends in the ratings of resumes (Hakel, Ohnesorge, & Dunnette, 1970; Landy & Bates, 1973), interviewees (Schuh, 1978; Wexley, Sanders, & Yukl, 1973; Wexley, Yukl, Kovacs, & Sanders, 1972), and assessees

(Butler, 1989; Gaugler & Rudolph, 1992).

Varied explanations have been offered for contrast and assimilation effects in social perception. Early explanations, based on trace theories of memory, suggested that incoming stimuli were contrasted with or assimilated to a distribution of similar stimuli. More recently, explanations have been framed within social-cognitive models of information processing. In particular, social cognitive explanations have emerged from research conducted with by priming subject with labels or categories and examining the influence of priming on evaluations. A number of studies have found that raters attention and memory are influenced by personality or performance categories that were primed in the minds of raters before their observation. Ambiguous stimuli, in particular, tend to be assimilated into category membership if they are similar enough to be deemed a member of the category. Ambiguous stimuli tend to be contrasted from members of the category if they clearly do not possess the qualities of category membership (Herr, Sherman, & Fazio (1983)). It has also been suggested that primed categories result in inconsistent data receiving more attention. Therefore, contrast effects may occur as a result of a disproportional amount of schema-inconsistent information receiving attention and being encoded in memory (Maurer & Alexander, 1991). Though, it is also acknowledged that the latter explanation departs from traditional explanations which posit that contrast and assimilation occur in the evaluation stage of processing. In sum, social-cognitive interpretations of contrast and assimilation effects are compatible with the other assessment center-oriented, social-cognitive postulates offered in preceding paragraphs. Viewing

assessment center ratings from a social-cognitive perspective allows for an integrated understanding of the process by which assessees' behavior and contextual influences affect assessors' ratings.

The practical implications for assessment centers are also clear. Contrast or assimilation can occur in assessment center observations "between persons"--a comparison of different persons in the same exercise or across exercises--and "within person"--a comparison of the same person's performance across settings (Murphy et al. 1985; Gaugler & Rudolph, 1992). The recommended practice in assessment centers is to have different assessors observe assessees in different exercises and to have assessors observe and rate an assessee in only one exercise (Task Force on Assessment Center Guidelines, 1989). This diminishes the potential for a "within person" contrast or assimilation effect. There is more potential for the occurrence of between persons contrast or assimilation effects. Assessors may contrast or assimilate an assessee they are currently observing with other assessees that they observed previously or contemporaneously. Previous observations that may set standards for assessors' judgments include not only assessees observed in previous exercises but also practice examples of assessees observed during training.

Primacy and Recency

Within an exercise, assessors view assessees' behavior over time. Principles of primacy and recency suggest that behaviors performed early or late in an exercise may have a disproportional effect on ratings compared to behaviors performed in the middle of an exercise. Behaviors observed early in an exercise (primacy) may simply be remembered better or may prime a performance prototype or schema that

filters attention to subsequent behaviors (Asch, 1946; Jones & Goethals, 1972). Behaviors late in an exercise (recency) may be recalled more readily and may influence the search through memory for other behavioral data. Karl and Wexley (1989) examined primacy and recency hypotheses using assessors' ratings of a group discussion exercise. They created video taped stimulus exercises in which assessees' performance was ascending, consistently average, or descending. The assessee in the ascending condition was rated higher than the consistently average or descending performer on leadership, decision making, and persuasiveness dimensions. Motivation was perceived as the cause of performance in the ascending condition more than in the consistently average condition. Ability was perceived as the cause of performance in the descending condition more than in the ascending and consistently average conditions.

Purpose

Purpose of an evaluation or appraisal is thought to influence raters motivation, particularly how raters direct their efforts (DeNisi, et al. 1984, Foti & Lord, 1987). The purpose of rating individuals in assessment center exercises is commonly to make an evaluation of an individual's job-relevant skills and abilities for the purposes of personnel decision making. In this way, the purpose of assessment center evaluations closely approximates the quantification of an individual's behavior/performance. This is not necessarily true in all organizational evaluations such as performance appraisals where ratings may be more likely to occur in response to organizational, political, and administrative purposes. Nevertheless, not all assessment centers are used strictly for selection purposes. Changing

the purpose of an assessment center to the purpose of management development, employee classification, or downsizing from the purpose of precisely quantifying the skill levels of an assessee may produce different results than a center used strictly for selection purposes. Additionally, ratings in a center in which organizational members evaluate each other may be more subject to organizational/political purposes than the ratings in a center in which external, professional assessors are used.

It is likely that differences in ratings will be found in association with different purposes. Previous research suggests that raters may be less lenient when ratings are used for developmental rather than administrative purposes (DeCotiis & Petis, 1978) and more critical when ratings determine whether the ratee will be promoted into a responsible position in the rater's department (DeNisi et al, 1984). There is further evidence that purpose of evaluation influences the kinds of information that is sought and how information is stored in memory (Hoffman, Mischel, & Mazze, 1981; Williams, Blencoe, DeNisi, & Cafferty, 1983).

Cognitive Load

Cognitive models of performance have commonly postulated that humans possess a finite pool of cognitive/attentional resources that can be devoted to task performance at a given time (Wickens, 1984). Task demands on cognitive resources increase with the difficulty of tasks. For some kinds of tasks, performers may advance with practice into an automatic processing mode where performance no longer draws on cognitive/attentional resources. For particularly complex and changing tasks, performers may never advance into an automatic processing mode,

remaining instead in a resource dependent, controlled processing mode (W. Schneider & Shiffrin, 1977; Ackerman 1987).

Views of rating as a resource dependent task have been advanced by a number of researchers (e.g. DeNisi, et al, 1984; Landy & Farr, 1980). First, it has been suggested that raters in organizational performance appraisal tasks observe and rate while there are many competing demands on their time. Rater biases such as halo error may be a result of raters employing cognitively economical processes in the face of competing demands. For example, raters may employ schema driven observation and memory processes for the sake of economy, and halo in ratings reflects raters' tendencies to store information in the form of prototypes or general impressions and not in terms of specific, dimension-related instances (Lord, 1985).

Second, it has also been suggested that rating systems with large numbers of dimensions are a source of overload in and of themselves. This hypothesis has been advanced previously in assessment centers (Gaugler & Thornton, 1989; Turnage & Muchinsky, 1982) since large numbers of dimensions have commonly been used. Gaugler and Thornton (1989) found that reducing the number of dimensions led to better accuracy in assessor ratings, presumably through reducing the load on assessors.

In sum, though assessors rarely face the number of distractions external to the task of rating that raters in an organizational performance appraisal context face, assessors do not have infinite cognitive capacity for attention, memory and evaluation. The limits of assessors' cognitive capacity may influence their ratings.

Summary

Ratings in an assessment center are a product of assessor cognitive processing. Assessors' beliefs, expectations, attributions, and theories of personality or performance can affect assessments which are intended to be evaluations of assessees' behavior alone. It is likely that all assessments are subject to the influences of these cognitive processes, but assigning assessors to evaluate individuals in groups may magnify the influence of these cognitive processes because more data is available for observation under less standardized conditions. Assessors, aware of the successes and failures of the group and guided by theories of performance and attributions, may rate particular individuals consistent with their perceptions of the group's high or low performance. The performance of other assessees in a group may influence assessors' expectations for assessee performance and a contrast or assimilation effect may occur. Both of these effects have a high likelihood of occurring in assessment centers and should be explored.

Rating Systems

A number of researchers have suggested that the influence of exercises on ratings could be reduced and the ability to measure assessees' traits could be increased by changes in the design of the components of a center. The hypothesis is that better evidence of construct validity could be found with better designs. The design of scoring guidelines, assessor training, consensus procedures, and the number of dimensions may be a source of influence on ratings.

Scoring Guidelines

The traditional rating system for assessment centers has involved taking detailed notes of assessees' behaviors, dialogue, and non-verbal behaviors (Thornton & Byham, 1982). These behavioral notes are later categorized according to dimensions and discussed in consensus procedures with other assessors. Though the first centers did not have extensive scoring guidelines and relied primarily on the expert knowledge of assessors (Bray & Grant, 1966), behavioral anchored rating scales (BARS) or behavioral observation scales (BOS) have been developed for use in some centers (Thornton & Byham, 1982). More recent innovations have included scoring checklists and computerized categorization and rating systems. Reilly, Henry, and Smither (1990) found that convergent validity increased (the average coefficient of the same dimension across different exercises increased from .24 to .43) and that discriminant validity improved slightly (the average coefficient of different dimensions within exercises decreased from .47 to .41) with checklist scoring systems in their study. In a study of In-basket scoring, Heine (1990) compared a system of checklists combined with a computer program that categorized and issued ratings based on formulas developed from expert consensus judgments and a method where single assessors categorized and rated behaviors with BARS as the sole aid. The mechanical procedure resulted in more behavioral observations, better discrimination among dimensions, and higher inter-rater reliability and agreement. The single assessor procedure resulted in higher convergent and criterion-related validity in predicting uncontaminated criterion ratings.

Assessor Training

Inadequate assessor training has been offered as an explanation for the lack of discriminant validity in ratings (Klimoski & Brickner, 1987; Sackett & Dreher, 1982; Turnage & Muchinsky, 1982). I am not aware of any assessment center studies varying training procedures in the kind of experimental design that allows conclusions about the positive or negative effects of assessor training. Similar research in performance appraisals suggests that training can be influential particularly relative to no training but that variations in the length or content of training does not result in large improvements (Landy & Farr, 1980).

Consensus procedures

To obtain final ratings for an assessee, the traditional procedure has involved a consensus discussion at the end of the center in which assessors share their behavioral reports with each other to achieve consensus on dimension ratings. More recent innovations have eliminated or modified the consensus process because research has demonstrated that ratings emerging from the consensus discussion are highly predictable from ratings made by individual assessors after exercises (Sackett & Wilson, 1982; Silzer, 1984). It has become common practice for centers to use post-exercise discussions among assessors who observed the exercise or a statistical formula for combining ratings rather than a consensus discussion (Thornton, 1992).

Silverman et al. (1986) examined how variations in the consensus procedure can influence the ratings. They hypothesized that different relationships between dimensions and exercises may be observed when ratings are made independently by assessors after each exercise (within

exercise method) and when ratings are made on each dimension after considering the data from all the exercises (within dimension method). The within-dimension method resulted in slightly better evidence of discriminant validity and better evidence of trait factors than the within-exercise method. Nevertheless, the exercise loadings in the within-dimension method were still considerably higher than the dimension loadings suggesting a relatively strong exercise effect.

Dimensions

Researchers have highlighted the importance of requiring raters to rate only those dimensions for which they have sufficient data to rate (Turnage & Muchinsky, 1982). Researchers have also questioned the tradition of using large numbers of dimensions. It is thought that large numbers of dimensions may promote halo because of significant overlap in the behaviors anchoring different dimensions or because of excessive cognitive load on assessors (Bycio et al., 1987; Sackett & Dreher, 1982; Silverman et al., 1986). Gaugler and Thornton (1989) found greater observational accuracy when three rather than six or nine dimensions were used.

Summary

There was a time in the history of assessment centers when almost all operating centers were designed after the AT&T assessment center. Centers of this kind used many dimensions, consensus rating procedures during which extensive written records of an assessees' behavior were discussed, and a relatively standard assessor training program. These standard procedures are not followed with the same consistency in the present (Moses, 1992). Alternatives in center design including the introduction of checklist/mechanical scoring procedures, the

elimination of consensus discussions, the reduction in the number of dimensions, and variations in assessor training programs are apt to influence ratings. Though some research on these issues exists, more research is needed to understand the influence of these changes.

Group Variables Affecting Individual Ratings: A Summary

Theoretically, an assessors' role is to observe the behavior of the assessee whom they are assigned to rate and to make inferences about the assessee's knowledge, skills, abilities, and other characteristics (KSAO's) from these observations. As Figure 1 shows, much more data than the behavior of assessees are available for the attention of assessors, and these data may influence assessors' ratings. Exercise characteristics and features of the design of the center are the variables postulated here as having potential to influence ratings of assessees in group exercises in addition to the behavior of assessees. Since the primary goal of most centers is to measure individual differences, the influence of these variables and the extent to which they bias ratings needs to be understood.

PRESENT STUDY

The psychological and measurement impact of variables in Figure 1 are all worthy of future research. In choosing the variables to study in the present research, I thought that examining the cognitive processes involved in assessor rating would be a significant advancement because very little rating research in assessment centers has been framed within a social/cognitive model of person perception and judgment. I also considered it important to study those variables which posed the greatest threat to the ratings of individual assessees. The composition of the group to which an assessee is assigned is one such threat. Group composition can influence how an individual is rated since assessors' judgments are susceptible to the influence of seeing other assessees perform at the same time, by means of contrast or assimilation effects or by means of assessors' attributions. Further, composition of the group influences a groups' level of achievement, and as the performance cue research suggests, the achievements of a group can influence the ratings of an individual member of the group.

Finally, group researchers have emphasized the importance of testing hypotheses about group performance with multiple tasks (Hackman & Morris, 1975). In keeping with their recommendations I will also vary assessors' perceptions of exercise characteristics in my examination of the effects of group composition and perceptions of

group performance on the evaluation of a particular assessee. No one has previously suggested that assessors possess implicit theories about tasks in the same way that they are thought to possess implicit theories of personality and performance. It is plausible, however, that assessors' beliefs about tasks may influence ratings in the same way that their beliefs and attitudes toward individuals do. Kabanoff and O' Brien (1979) suggested that task characteristics may cue up different norms for performance among group members. Extending these findings to assessors, it is possible that their knowledge about task characteristics may cue different norms for performance. Thus, knowledge of or beliefs about the task may influence how assessees are rated.

In the present study, assessors watched scripted, videotaped performance of three actors working in an assigned leader, group discussion exercise. Assessors were assigned to observe and rate a target assessee whose performance was scripted so that he demonstrated a mix of high and low anchored behaviors. This mix of behaviors established him as an ambiguous target. The target was presented on video with either two high or two low performing assessees, thus, enabling the examination of group composition effects such as contrast or assimilation effects. Perceptions of group performance was manipulated by the information assessors were given during assessor training. The information created assessors' perceptions that the group was either a high or low achieving group. The influence of varying perceptions of the group's performance on the ratings of the target was then examined.

A strength of this study is that both group composition and perceptions of group performance were examined jointly across different types of tasks. Of the exercise characteristics variables in Figure 1, task interdependence seemed most suitable for a manipulation involving assessors' perceptions of the exercise tasks. Further, in support of using task interdependence as the exercise manipulation, it is very common for assessment center administrators/designers to control the distribution of information in group exercises. Task interdependence was defined here by the distribution of information to assessees, in keeping with Schneider and Schmitt. In the present study, assessors' perceptions of interdependence were manipulated by informing them about differences in the distribution of information to assessees.

The remainder of the introduction section is devoted to examining the potential effects associated with group composition, perceptions of group performance, and perceptions of task interdependence on the evaluation of an assessee in a group. Research from a number of areas is consulted including the social perception literature on contrast and assimilation effects and the organizational behavior literature on performance cue effects. These literatures provided the basis for hypothesis development regarding the effect of group composition and group performance on assessments of individuals in a group. I also reviewed group research on task interdependence as a basis for guiding the exercise manipulations in this research. I concluded with specific hypotheses regarding how the joint influence of group composition, perceptions of group performance, and perceptions of task interdependence will be manifest in the ratings of assessees.

Group Composition

The previous section on contrast and assimilation provided an overview of contrast and assimilation issues. This section covers some of the same territory but in greater detail. Previous contrast and assimilation research in personnel selection is reviewed first. This is followed by a more theoretical examination of whether between persons, contrast and assimilation effects can occur because of contemporaneous observations of other group members in assessment center group exercises.

Contrast and Assimilation in Selection Research

A number of studies have demonstrated that contrast and assimilation effects occur in personnel selection judgments. Many of these studies have examined between ratee contrast or assimilation effects in which other ratees seen prior to the target influence the ratings of the target. Some of the research cited also provides evidence suggesting that a between ratee contrast and assimilation effect can occur because of contemporaneous observations of other ratees. All of these results are reviewed to develop the hypothesis pursued in this study, namely that a between assessee contrast and assimilation effect can occur because of contemporaneous observation of other assessees in assessment center group exercises.

Resumes and Interviewees

In a study using job applicant resumes, Hakel, et al. (1970) found modest between ratee contrast effects relative to prior evaluations. Professional interviewers and student subjects read three resumes in the study. The first two resumes set the standard or "expectation" for the subjects and were developed to portray equivalent levels of

performance (High, High (HH); Average, Average (AA); or Low, Low (LL)). A third resume was manipulated at a high or low performance level, creating six different levels of the performance manipulation (HHH, HHL, AAH, AAL, LLH, and LLL). Contrast effects were significant for both professional interviewers and students, the effects accounted for a small percentage of variance (1% to 2%) in ratings compared to ratees' qualifications. Leonard and Hakel (cited in Landy & Bates, 1973) attempted to extend these findings by using four in addition to two homogeneous applicants preceding the presentation of the final experimental applicant. Increasing the number of stimluli to set the standard did not prove to be a significant factor influencing the student interview ratings.

Rowe (1967) also examined between ratee contrast and assimilation effects. She constructed sets of descriptions of persons that were composed of different ratios of favorable (F) and unfavorable adjectives (U). All descriptions contained six adjectives. The different mixes of adjectives were: 5F-1U, 4F-2U, 3F-3U, 2F-4U, and 1F-5U. Contrast effects were greatest with the neutral target (3F-3U). Student participants, indicating whether they liked the person, demonstrated a tendency to respond more favorably to a neutral person if preceded by an unfavorable person and more unfavorably to a neutral person if preceded by a favorable person.

Landy and Bates (1973) conducted two studies using Hakel et al.'s (1970) manipulations and samples of student raters and professional recruiters to determine whether experience of subjects was a factor contributing to contrast effects. Landy and Bates had the raters read 12 resumes in total. The first nine resumes were randomly selected

good, average, or poor resumes. Resumes ten and eleven induced homogeneous performance expectation at either good, average, or poor levels. The target applicant, resume 12, was varied at good and poor levels. Study two was a replication of study one but Landy and Bates changed the criterion in the second study to a decision to offer a second interview to the candidate, in keeping with the kinds of decisions that recruiters typically make. The bulk of the variance in ratings of the target in both studies was associated with the level of performance of the target, and no evidence of a significant contrast or assimilation effect was found in either study.

Wexley, et al. (1972) noted that neither Hakel et al. (1970) nor Rowe (1967) employed procedures that allowed the rater to observe the applicant. In Wexley et al., student participants watched videotaped structured interviews in which performance was manipulated at eight levels (HHH, HHA, HHL, LLH LLA, LLL, AAH, and AAL). High applicants on the video portrayed uniformly high behaviors/qualifications; low applicants portrayed uniformly low behaviors/qualifications; and average applicants portrayed a combination of high, average, and low behaviors/qualifications. When H's were preceded by L's and L's were preceded by H's, contrast effects were significant but of the same modest magnitude as those reported by Hakel et al. A much larger contrast effect (as much as 80% of total rating variance) was demonstrated when the third interviewee was the average performer. Average performers preceded by low performers received a higher average rating than average performers preceded by high performers. Wexley et al. suggested that the average target's high or low qualities gain salience to raters as the target is contrasted with preceding low and

high applicants. Raters' ratings reflect these differences in perceived salience.

Wexley, et al. (1973) conducted a series of experiments to determine whether any of the following training interventions would reduce contrast effects: (a.) a warning about contrast effects as a source of error; (b.) an anchoring treatment in which raters were given examples of a person's behavior at the high and low ends of a nine point scale; (c.) a combination of the warning and anchoring treatment including the addition of average anchors and a discussion of the anchors; and (d.) a two hour workshop including a discussion of the job requirements, training in how leniency, halo, central tendency, contrast, and stereotyping affect ratings, and observation and rating practice. Only the workshop training program was successful in reducing the contrast effects. A comparison of this condition with the other manipulations suggests that practice in observation and rating may have contributed to the reduction in contrast effects whereas verbal warnings and more detailed scoring anchors alone were not sufficient.

Kopelman (1975) examined contrast effects in selection interviews for admission to medical school. He used videos of two good or two bad interviewees to set the performance standard. A third video of a good, average, or low performer was used as the target. The target's level of behavior accounted for the greatest variation in his/her ratings, though 11 percent of the variance in the ratings was explained by contrast with the standard set in the first two exposures.

Schuh (1978) also conducted an interview contrast effect study using MBA applicants for a management trainee position and professional

recruiters as evaluators. He used three randomly presented videos of applicants and a constant applicant in the fourth position. Regardless of the order of presentation of the first three applicants, the applicant presented first tended to get the highest rating. The rating of the fourth or target applicant was highly associated with the level of performance of the applicant seen immediately prior.

Assessees

Recent assessment center research has confirmed the existence of group composition effects in the form of contrast effects (Butler, 1989; Gaugler & Rudolph, 1992). Butler (1989) used video tapes of an average assessee performing in a group discussion with either two high performers, two low performers, or one low and one high performer. The target "average" assessee received higher ratings when she was observed in a group discussion with two below average assessees than when she was observed with two high or one low and one high performing assessee. Butler's research is the only demonstration of a between assessee contrast and assimilation effect based on contemporaneous observations of assessees.

Gaugler and Rudolph (1992) found "within assessee" and "between assessee" contrast effects using observations of high and low performance manipulations of the same and different assessees across different exercises. In the "within assessee" conditions, the low performance of a target assessee was rated lower when the same assessee's performance had previously been high as opposed when the same assessee's performance had been low. In the "between assessee" conditions, they found, in two of three exercises, that the low performing target assessee was rated significantly lower when he was

evaluated after observing two high performers in two previous exercises than when he was evaluated after observing two low performers in two previous exercises. It should be noted that between and within assessee contrast and assimilation occurred because of previous and not contemporaneous observations of assessees.

Conclusions from Contrast and Assimilation Research in Personnel Selection

The studies cited in the preceding sections have differed in findings as well as methodology. Some consistencies in the findings of contrast and assimilation effects across studies lend themselves to interpretation. The explanations here tend to focus on methodological differences, but some extensions to theoretical explanations can be made.

Ambiguity of performance target. The differences between the findings of the studies cited in the previous section can be explained by the level at which the target ratee's behavior is manipulated. Contrast effects of relatively large magnitude were found in studies using average or mixed performance targets (Butler, 1989; Kopelman, 1975; Rowe, 1967; Wexley et al., 1973; Wexley et al., 1972) and not in studies using targets whose behavior was manipulated at the extreme, high or low levels (Hakel et al, 1970; Leonard & Hakel, 1971; Landy & Bates, 1973).

This finding that ambiguous targets are most susceptible to contrast and assimilation effects has been supported in other contrast and assimilation research. Using a priming paradigm, Herr, et al. (1983) suggested that contrast and assimilation effects were based on two dimensions--ambiguity of the stimulus and extremity of the

category. Based on their research findings, they suggested that as a person encounters a stimulus, the first step involves a search to determine the appropriate category to which the stimulus belongs; categories that have been recently activated by priming will be accessed first. The subject then must decide if the stimulus sufficiently matches the features of the stimuli associated with the category. If the stimulus is ambiguous and the category is not an extreme, the stimulus will likely be assimilated into membership in the category. Contrast effects are most likely to occur when the stimulus is so unambiguously not a member of a category that it can not plausibly be assimilated into the category or when the accessed category is so extreme that even an ambiguous stimulus can not be assimilated into the category.

Experience of rater. Explanations of the findings have commonly focused on raters' familiarity with the task. Students have little knowledge and experience in making judgments about interview applicants, and it has been postulated student raters are particularly likely to be susceptible to extraneous contextual variables (Denisi & Pritchard, 1977; Smither, Reilly, Buda, 1988). The results have been equivocal. Hakel et al. (1970) and Landy and Bates (1973) found consistently small contrast effects independent of whether students or professional interviewers were used, but Wexley et al's (1972) sizeable contrast effects were found with student raters making personnel decisions. In order to eliminate a knowledge effect, attempts have been made in recent studies to use tasks that student subjects are qualified to judge such as applicants for dormitory resident assistant positions (Butler, 1989; Gaugler & Rudolph, 1992).

Strength of stimulus manipulations. It has been postulated that contrast effects are apt to be larger when a larger number of standard stimuli or when extreme stimuli are used to create the manipulation. Regarding number of standards, Leonard and Hakel (cited in Landy & Bates 1973) used four rather than two homogeneous stimuli to set the standard, and Landy and Bates used a total of twelve stimuli (two stimuli to set the standard preceded by nine random stimuli). Contrast effects were not found to be associated with the number of stimuli used to set the standard in these studies. Regarding extreme stimuli, Kopelman (1975) noted that Wexley et al. (1972) used high and low standards who were completely qualified or completely unqualified and who ultimately received nearly maximum or nearly minimum ratings in the study. Kopelman (1975) suggested that Wexley et al.'s extreme standards likely exaggerated the contrast effect when high and low performers were contrasted with average performers. Kopelman's comments are reasonable in that stimulus materials should be manipulated within a reasonable range of performance. It should be noted, however, that evidence of contrast effects were found in Schuh (1978) who used standard interviewees at different levels and in Gaugler and Rudolph's (1992) study in which a more moderate manipulation of high and low performance was used.

Type of stimuli. Wexley et al. (1972) have suggested that different effect sizes in contrast effect studies may be associated with the opportunity to observe behavior (on videotapes) versus reading written materials. This distinction holds when comparing Wexley et al (1972) with Hakel et al. and Landy and Bates, but Rowe (1967) found a contrast effect with written descriptions.

Practice during training. Based on Wexley et al. (1973), the opportunity to observe practice targets and practice making ratings may reduce contrast and assimilation effects, especially compared to other training interventions such as verbal warnings and scoring guidelines. It should also be noted, however, that there are limits to the potential effectiveness of training in reducing so called "biases" from ratings. Years of research on rater training has revealed that training can not completely eliminate biases from ratings (Landy & Farr, 1980), even though training may result in slight reductions of rater errors (Bernardin & Pence, 1980; Bernardin & Walter, 1977).

Summary. In sum, ambiguous targets seem most susceptible to the effects of contextual variables. This is threatening to assessment center ratings since it is rare that assessees perform at unambiguously high or low levels. Further, it is important that study participants be qualified to judge the people and tasks they are assigned to rate, and it is important that manipulations of standard stimuli are within reasonable ranges.

Theoretical Foundations for Understanding Contrast and Assimilation

The contrast and assimilation research cited above has confirmed that contrast and assimilation effects can occur relative to previous observations of other persons (between persons) and relative to previous observations of the target person in another context (within person). The question pursued in the present study is whether simultaneous observations of other assessees in a group exercise can bias the assessment of an assessee. With the exception of Butler (1989), none of the previously cited studies has examined "simultaneous, between persons" contrast and assimilation effects. In

the following paragraphs, I review research to determine whether it is reasonable to expect simultaneous, between persons contrast or assimilation effects.

Two explanations for the simultaneous, between persons contrast effect are considered. First, the "shifting standards hypothesis" suggests that assessors' standards may shift from standards set from their training or experience to non-target assessees that they observe in a group exercise as standards. Second, the "discrepant behavior hypothesis" is an explanation for contrast effects which suggests that behavior that stands in contrast to other behaviors in the observational context is likely most salient and therefore most likely to influence assessors' judgments. Before presenting the research regarding these two explanations, I will discuss some of the theoretical background for understanding contrast and assimilation effects.

Contrast and Assimilation

The notion of judgments of stimuli relative to a standard lies at the core of many theorists' explanations of contrast and assimilation (e.g. Berkowitz, 1960; Helson, 1948). Theorists differ, though, in their conceptualizations of the nature of these standards. In early conceptualizations, theorists explained standards in terms of subjective ranges or distributions of stimuli and their central tendency. These theorists suggested that individuals compare stimuli to be judged with a mental distribution of stimuli that they deem related to the stimuli (Helson, 1948, 1964). Individuals' sense of the central tendency of this distribution is represented by a neutral point (adaptation-level), and incoming stimuli are judged relative to this

neutral point. Individuals' distributions and neutral points are by no means fixed; both are constructed from and shift in response to experience with similar stimuli in the recent past or present (Helson, 1948, 1964; Hovland & Sherif, 1952).

More recently, theorists have proposed that standards are prototypes in memory (i.e. schemata). These prototypes or schemata are thought to serve as standards to which incoming stimuli are compared, and contrasted or assimilated (Cantor & Mischel, 1977). Several studies have shown that a schema influences how an ambiguous stimulus is judged if the schema is experimentally primed before exposure to the stimulus (e.g. Higgins, Rholes, & Jones, 1977; Srull & Wyer, 1988). Cantor and Mischel's (1977) research confirmed that traditional personality or trait concepts (e.g. introvert, extrovert) may serve as schemata employed in social information processing.

Subjective range explanations are not completely incompatible with schema explanations. Cantor and Mischel (1977) suggested the content of schemata may be abstracted prototypes that resemble the central tendency of stimuli in the distribution. If the content of schemata involve some notion of central tendency, it is possible that a schema may act in accord with subjective range propositions. Additionally, it has been acknowledged that perception and memory are both data driven and schema driven (Fiske & Neuberg, 1990; Higgins & Bargh, 1987). Higgins and Bargh (1987) suggested that the data gained in encounters with stimuli automatically activate trait constructs that are used in interpreting stimuli, particularly when the target stimulus is ambiguous. A stimulus, then, may be assimilated as a representative or member of the category that is accessed and interpreted to possess many

of the qualities associated with the category. Higgins and Bargh also suggested, however, that the data is never abandoned in this process. Accessed constructs are employed only if the incoming data is representative of the construct. Contrast effects rather than assimilation effects are more likely to occur when the stimulus is so unambiguously not a member of a category that it can not plausibly be assimilated into the category (Herr et al., 1983).

Regardless of whether standards are defined as subjective ranges or schemata, social perception and judgments are thought to occur relative to standards based on prior, recent, or present experience. In order for contrast or assimilation to operate in assessment centers because of observations of other assessees in a group exercise, assessors must shift from standards based on past experiences or assessor training to standards based on other assessees. Two explanations of how assessors may make this shift--the shifting standards hypothesis and the discrepant behavior hypothesis--are explored in the following paragraphs.

The Shifting Standards Hypothesis

Most theorists acknowledge that standards shift (Helson, 1948; Parducci, Knobel, & Thomas, 1976). A key question is whether and how rapidly do observers' standards shift from past experience with stimuli to experiences with stimuli in the present. The "shifting standards hypothesis" was advanced by Maurer and Alexander (1991). They suggested that Wexley et al.'s (1973) findings could be interpreted in terms of a transient adjustment of raters' evaluative standards in response to the behavior they observed. In other words, raters' internal representations or standards of good and poor performance

shifted temporarily in the context of the current observation. This would explain why raters rating a target would be influenced by stimuli seen immediately before the target.

In their study, Maurer and Alexander had student participants use a button pressing apparatus to respond to the behaviors they observed. The stimulus materials were teachers delivering a lecture from Murphy et al. (1985). In order to test the shifting standards model, they asked participants to respond to a rating instrument for ideal, typical, and worst possible lecturers they could imagine, immediately after observing the stimulus video. The results suggested that the level of variability in the ratings was similar for ratings of ideal, typical, or worst possible lecturers regardless of the performance they observed. In other words, raters' standards did not shift in response to the behavior they had just observed.

These results speak directly to the issue of whether assessors, who receive extensive training devoted to setting standards, may shift from the trained standards to standards based on other assessees observed simultaneously. Maurer and Alexander suggested that a shift to other ratees as standards was unlikely since raters' "ideal, typical, and worst" possible standards showed very little change across observation contexts in their study. Maurer and Alexander suggested that their results were better explained by the salience of behaviors that are discrepant from the context in which they occur. This explanation will be discussed in the next section.

The Discrepant Behavior Hypothesis

Researchers have also attempted to explain contrast effects as a result of attention and encoding processes. It is acknowledged that

attention and encoding explanations stray from the manner in which contrast and assimilation effects have been traditionally defined in psycho-physical research. Discrepant behavior explanations suggest that behavior that is at variance with expected behavior may capture observers' attention and be processed more thoroughly than behavior that is consistent with expectations (Woll & Graesser, 1982). Hastie and Kumar (1979) suggested that unexpected or incongruent acts are likely to be considered most informative about persons and are processed more deeply and remembered more richly. In addition, the relative number of unexpected or incongruent acts seems to be important. A few incongruent acts by a person will be remembered better than an equal blend of congruent and incongruent acts.

Murphy, et al. (1985) illustrated this point using a situation where a performance rating is issued to an individual who is observed over a long period of time. If the ratee has tended to be a high performing employee, a rater will likely process examples of poor behavior (prototype inconsistent) more deeply than good behavior (prototype consistent). At the time of rating, incidents of poor behavior have a higher probability of being recalled than good behaviors, thus prompting a lower rating than would be expected from a complete record of behaviors (i.e. a contrast effect).

In an assessment center group exercise, an assessee's high performance is likely to get the most attention when behavior of other assessees seen contemporaneously has tended to be low. Because of this increase in the probability of attention to this behavior (i.e. salience), it is also most likely to be stored and recalled from memory. If this scenario occurs over a series of performance

instances, the predominance of positive performance examples in memory will result in a positive rating in contrast to the low performance of others in the group.

Maurer and Alexander (1991) found support for a discrepant behavior explanation in their study of between ratee contrast effects. They hypothesized that behavior, inconsistent with behavioral and other stimuli in the observational context, should be encoded more deeply and recalled more readily, so that ratings are apt to be biased in the direction of inconsistent information. In their study, they found that 43% of the contrast effect was explained by variability in encoding, when the effects of observers' button pressing tendencies were controlled.

Summary of Group Composition Effects

Social perception research suggests that social judgments are made relative to some kind of standard that is derived from experiences with similar stimuli observed in the same or other contexts. Though assessor training is designed to instill normative standards in raters, immediate experiences, such as others observed immediately prior or concurrently with the stimulus, may also influence assessors' evaluations. Two explanations have been explored in this section. First, it is possible that assessors' evaluative standards may shift as a result of simultaneous observations of others in group exercises. Maurer and Alexander's (1991) research failed to support this explanation, however. Second, it is possible that behaviors that are most inconsistent with other behaviors observed in the same context may receive the most attention and processing in impression formation processes ((Higgins & Bargh, 1987; Maurer & Alexander, 1991). Contrast

effects are then thought to occur because context-discrepant behavior makes up the greatest portion of information in memory and provide the basis for ratings.

It is reasonable to expect contrast effects may occur in assessment centers on the basis of a context-discrepant behavior hypothesis. First, there is evidence of between persons contrast effects in assessment center group exercises (Butler, 1989), and there is evidence that a discrepant behavior explanation accounts for between persons contrast effects (Maurer & Alexander, 1991). Second, assessment centers are an impression formation process, and research suggests that context-discrepant behavior receives greater attention in impression formation (Higgins & Bargh, 1987). Finally, the prediction of a contrast effect based on a context-discrepant behavior hypothesis is also most consistent with the notion that assessors rely on memory less than raters in other applications. Assessors have notes on which to base their ratings so contrast effects may be more likely to result from contrasting behaviors receiving more attention than from effects occurring at the recall or evaluation stage.

It is also important to note that ambiguous stimuli are most susceptible to the influence of variables in the rating context (Herr et al., 1983). It is not likely very difficult to rate assessees whose behaviors are clear examples of high or low anchored behavior. Assessors may be more susceptible to the influence of contextual variables when they experience indecision about rating targets whose performances is at a mixed or ambiguous level. This problem is magnified by the fact that most assessees demonstrate a mix of high and low behaviors making them relatively ambiguous targets to judge.

Considering these issues of observation and rating of ambiguous targets, the following hypotheses are offered.

Hypothesis 1. A contrast effect was expected to be associated with observations of an ambiguous target assessee in groups of varying compositions. The ambiguous target assessee was predicted to be rated higher when observed with other assessees who were scripted to perform a majority of low anchored behaviors and lower when observed with other assessees who were scripted to perform a majority of high anchored behaviors.

Hypothesis 1A. It was predicted that perceived dissimilarity of the target's performance versus the non-targets' performance would partially mediate the relationship between the manipulated conditions of group composition and the ratings. The expected relationship is depicted in Figure 2.

This hypothesis was pursued in order to explain the relationship between group composition and ratings. Assessors were asked the extent to which they viewed the target assessee as similar to or dissimilar to the non-target assessees in terms of their performance. This analysis was thought to assist in the interpretation of the findings by confirming that the effects in the ratings were due to assessors rating the target based on a comparison with the non-targets. Other hypotheses introduced later postulated similar trends in the ratings so the measurement of dissimilarity was also thought to assist in discriminating between the different hypothesized processes that explain similar predicted trends in ratings. It should be noted that contrast and assimilation does not necessarily occur at the conscious level of processing and thus, may not be accessible from the self-

reports of assessors. It was thought, however, that assessors should be able to report on their perceptions of the relative performance levels of the target and non-targets and that assessors' ratings would reflect these perceived differences. As the differences in performance are perceived as greater, a greater degree of contrast effect is more likely to occur.

Perceptions of Group Performance

A number of research studies have found that group participant's and/or observer's evaluations about the group and its members are influenced by knowledge of group performance (Binning & Lord, 1980; Downey, Chacko, & McElroy, 1979; Guzzo, Wagner, Maguire, Herr & Hawley, 1986; Martell & Guzzo, 1991; McElroy & Downey, 1982; Staw, 1975). Typically in these studies, group participants and/or observers are given a performance cue suggesting that the group was highly successful or unsuccessful in achieving task-related outcomes compared to other groups. Ratings of process and outcome dimensions and recall of behaviors tend to be associated with the level of the cue, even though raters have observed identical stimulus materials.

These findings are relevant to assessment center evaluations. Assessors are not explicitly given a bogus performance cue as in the studies cited previously. Nevertheless, assessors, through assessor training or previous experience as assessors, are apt to develop a sense of the distribution of outcomes achieved by groups. For any group observed, then, assessors will have a sense whether the group's achievements were high or low. The observed performance achievements of the group may influence their ratings, even though assessors are instructed to rate on individual behavior.

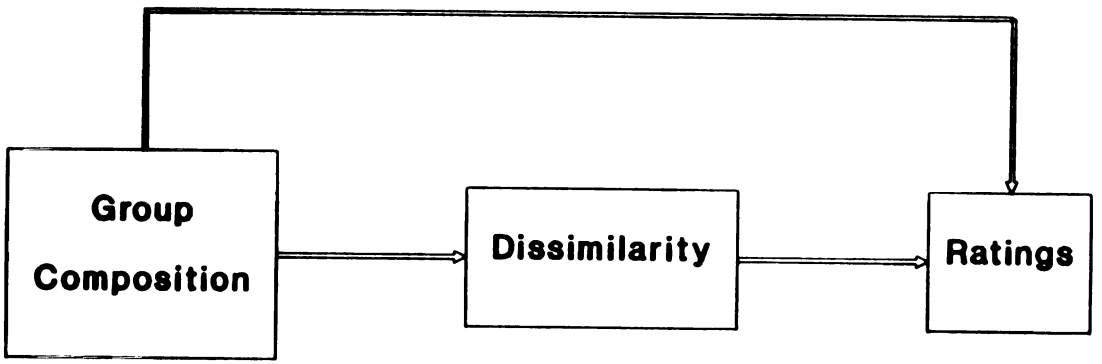

Figure 2

Hypothesis 1A: Perceptions of Dissimilarity as a Partial Mediator of the Relationship between Group Composition and Ratings.

The section that follows includes a more exhaustive review of the performance cue findings and explanations that have resulted. The aim of this review is to determine how assessors' beliefs about group performance achievements may affect ratings and to develop additional hypotheses for this study.

Performance Cue Effect

Typically in questionnaire or survey research, participants' perceptions of organizational constructs or phenomena are treated as independent variables or causes of the performance-oriented dependent variables. In contrast, Staw's (1975) finding of a performance cue effect illustrated that cross-sectional, questionnaire research in organizations may be capturing the consequences rather than the determinants of performance. Staw hypothesized that organizational members possess theories of performance that influence their responses to questionnaires. Knowledge of performance is thought to cue these theories of performance which guide attributions that respondents make about themselves, their work group, or their organization. Thus, participants may retroactively form opinions about organizational characteristics that are consistent with what they believe to be true about links between achievements and their behavioral antecedents. To illustrate his hypothesis, Staw gave participants in his participants bogus information or a performance cue suggesting that the group's performance was high or low. He found that participants rated cohesiveness, communication, motivation, and openness to change significantly higher in the high performance condition than in the low performance condition.

Familiarity

A number of studies followed in an attempt to examine the boundary conditions around Staw's findings. DeNisi and Pritchard (1977) suggested that Staw's participants did not interact with each other for a long enough period of time to accumulate a great deal of knowledge about the group and its members. This lack of information may have led raters to rely on knowledge of performance to fill in the missing details. Their hypothesis was that raters who were more familiar with ratees would have sufficient data for making judgments and would not need to rely on performance cues. Their groups worked for approximately one hour per day over a three day period, and consistent with their hypothesis, they found no association between performance cues and ratings of group characteristics.

Downey, Chacko, and McElroy (1979) tested the effects of familiarity using a set of groups who had a prior history of working together on business policy cases over 12 weeks and another set of groups who had no prior history of working together. Performance cues had a consistent and significant effect on raters' ratings of their own, other's, and the group's motivation, ability, satisfaction, and role clarity. Contrary to DeNisi and Pritchard's findings, prior history had little impact on individual perceptions of group processes such as cohesiveness, influence, communication, task conflict, openness to change, ability, and satisfaction but had a significant impact on group members' perceptions on their own and other's motivation.

Binning and Lord (1980) attempted to resolve the inconsistencies between Staw's, DeNisi and Pritchard's, and Downey et al.'s finding by using Staw's instrumentation (which had been altered by DeNisi &

Pritchard) and by using three one hour working periods as in DeNisi and Pritchard. In contrast to DeNisi and Pritchard's hypothesis that raters lacking information would be more inclined to rely on stereotypes, Binning and Lord suggested that longer periods of interaction may increase the use of stereotypes. They emphasized that social interactions as short as thirty minutes contain rich behavioral detail, and stereotype-based information processing enables raters to organize and simplify large amounts of data from long interactions.

Binning and Lord suggested the wording of some of DeNisi and Pritchard's questionnaire items may have led raters to make ratings based on the first two conditions and not the third, experimental condition. By using Staw's instrumentation, they eliminated this alternative explanation of the findings. Further, Binning and Lord suggested that DeNisi and Pritchard's raters may have discounted the feedback in condition three since they were given accurate feedback in the first two conditions and feedback that was inconsistent with observed performance in condition three. Thus, poor instrumentation and a weak performance cue manipulation that may have been discounted by raters may explain DeNisi and Pritchard's failure to find the performance cue effect.

Binning and Lord found that performance cues were significantly associated with ratings of fourteen of twenty-three items tapping characteristics such as cohesiveness, influence, communication, and ability. The effect of familiarity or length of working together was not as strong.

Actor/Observer Differences in Performance Cue Research

McElroy and Downey (1982) hypothesized that the performance cue effect may not hold for observers because of actor and observer differences in processing information. They hypothesized that observers who tend to attribute behavior to stable, internal causes may be less influenced by performance cues than actors who tend to attribute behavior to situational causes. Contrary to their hypotheses, the results suggested that observers were influenced by the performance cues as much as actors. Though their hypotheses were not supported, their results did suggest that the performance cue effect extended to observers in addition to group members.

Performance Cues in Leadership

Performance cue effects also have been examined in leadership research. Farris and Lim (1969) predated Staw in their examination of whether questionnaire responses represented antecedents or consequences of organizational outcomes; their study focused specifically on attitudes of work group members toward their leader. They examined the questionnaire responses of male graduate students working in 50 four-person groups on a case study task. Two sets of student groups were given performance feedback suggesting that their group was either high or low performing, and a third set of groups was a control condition. Participants in high performance feedback condition reported significantly greater subordinate influence in decision making, greater cohesiveness, and higher satisfaction. No relationship was found between performance feedback condition and associated leader behavior or estimates of change in group performance.

Lord, Binning, Rush, & Thomas (1978) focused their study on the effect of performance cues for observers. Raters watched either of two 15 minute video tapes of four-person (two males and two females) groups. The designated male leader was coached to perform many initiating structure behaviors on one tape and few initiating structure behaviors on the other tape. Raters in two experimental groups were told either that the group performed the task well/correctly and was second best out of 24 groups or that the group performed the task poorly/incorrectly and was second worst. A control group of raters were given no performance cue information. The effects of the performance cue were significant for 11 of 12 LBDQ dimensions. There were no significant interactions suggesting that the performance cues had an effect that was independent of the behavioral stimuli on the videos. The level of stimulus behavior also had a significant main effect on the ratings suggesting that the performance cue was not used simply as a substitute for behavioral information.

Performance Cues and Group Research

Performance cue research has recently been applied in group research by Guzzo and his associates. Guzzo et al. (1986) conducted two experiments with observations and ratings of a videotaped group who were constructing a bridge over a pool of water. Raters received positive or negative outcome feedback cues or positive or negative process cues after observing the video. One set of raters received no feedback and served as the control group. They found that bogus positive process feedback cues were associated with higher outcome ratings and bogus positive outcome feedback cues were associated with higher process ratings. Unlike previous studies, they did not find a

relationship between feedback condition and memory accuracy though, also unlike previous studies, a multiple choice rather than free recall test was used to measure memory accuracy. A second study, controlling for feedback tone and including a manipulation check, confirmed these results and suggested no effect due to tone of feedback.

Martell and Guzzo (1991), using the same stimulus materials, attempted to account for the process underlying the performance cue effect in group evaluations. They found that evaluative ratings and behavioral recollections were distorted in accord with performance cues. In particular, they found that observers receiving positive feedback cues were more liberal in attributing effective behaviors to the group and more conservative in attributing ineffective behaviors to the group. Conversely, observers receiving negative feedback cues were more liberal in attributing ineffective behaviors and more conservative in attributing effective behaviors to the group. They suggested that the distortions in behavioral recollections were due to the adoption of a more liberal decision rule when deciding about the occurrence of cue consistent behavior and a more conservative policy when deciding about the occurrence of cue inconsistent behavior. The biasing effects of negative or positive cues were not significantly different in a delayed rating condition.

Summary

In sum, ratings of characteristics of a group and of group members are commonly associated with experimentally induced beliefs about the level of performance of the group. The performance cue effect seems to hold for actors and observers, and performance cues seem to affect ratings independently of but not in the complete absence of the

observations of behavior. Cognitive processing explanations of the findings have involved attribution theory (McElroy & Downey, 1982) and schemata or prototype models (DeNisi & Pritchard, 1978; Binning & Lord, 1980; and Martell & Guzzo, 1991). These explanations and others are pursued in greater depth in the next section.

Explanations of the Effect of Perceptions of Group Performance

A number of research studies have attempted to understand how perceptions of group performance can influence the ratings of individual and group characteristics. In one line of research, researchers have given performance cues before or after the observations of performance to determine the stage at which the performance cue effect occurs. In particular, studies have examined whether the effect occurs at attention/encoding or retrieval/evaluation. In other lines of research, researchers have examined whether attributional frameworks or schema processing models explain the effects of knowledge of performance on ratings. Research examining schema processing models has involved delayed rating conditions to determine how experimentally induced, increased reliance on memory alters the process by which knowledge of performance has its effect. The stage at which the performance cue effect occurs, attribution explanations, and schema/impression explanations of the performance cue effect are reviewed in the following sections.

The Stage at which the Performance Cue Effect Occurs

Mitchell et al. (1977) conducted three experiments examining the effect of performance cues on subsequent ratings of leader behavior and situational variables. Using Stogdill's (1963) Leader Behavior Description Questionnaire (LBDQ), they considered Initiating Structure

and Consideration as measures of leader behavior, and they used ratings of group atmosphere, power, and task structure as situational variables. Their findings differed across the three studies, but these differences were interpretable.

First, the performance cues were given before observations of the video in studies one and two and after observations in study three. Significant effects for performance cues on behavioral ratings (consideration and initiating structure) were found in studies one and two but not in study three. They hypothesized that the cues before observations in studies one and two may have served as a filter during the observation process so that observers' attention was biased toward behaviors that were consistent with the cue. The cue-consistent behaviors that received attention were the only cues encoded in memory, and subsequent behavior ratings reflected the cue-consistent data in memory. When the performance cue was given after the observations in study three, attention was not biased in the same way, and ratings of behavior did not demonstrate the influence of the performance cue. Mitchell et al. also noted the possibility of actor and observer differences in attribution across the three studies; this interpretation is reported in the attribution section that follows.

Larson (1982) hypothesized that the performance cue could affect raters either at the attention/observation or recall/evaluation stage of rating. In the first case, raters' implicit theories, prompted by performance cues, guide attentional and storage processes so behavior that is consistent with the cue is more likely to be coded into memory. In the latter case, raters selectively search through memory in a way that theory consistent behavior is more likely to be retrieved, or

raters may reconstruct how the leader probably behaved given the cue and not necessarily on recall of specific behaviors. Larson gave raters performance cues (success/failure) before or after their observations (pre- or post-observational). His hypothesis was not supported; regardless of the timing, ratees in the success condition were rated higher on Consideration and Initiating Structure. Given that the performance cues had an equally significant effect in the post-observation condition, Larson was left to conclude that pre-observation cues were not sufficient to create the performance cue effect. Following the same logic, Larson suggested that performance cues did not work solely by affecting raters at the time of attention.

Larson and colleagues (Larson et al., 1984) revisited this research question but employed a different method based on signal detection theory. They tested for memory accuracy in addition to ratings as dependent variables. The addition of the test for memory accuracy was critical since ratings only revealed the extent to which performance cues were associated with raters beliefs about what occurred and not necessarily what "actually" occurred. Larson et al. tested two explanations for the performance cue phenomenon including selective availability in memory--only behaviors consistent with the feedback cue are stored and/or retrieved from memory--or probabilistic response bias--observers receiving feedback reconstruct how individuals probably behaved consistent with the feedback cue but not necessarily with reality. Observers were given either success or failure group performance cues either before or after observing the exercise. Observers completed the LBDQ and responded to two questionnaires assessing observers' memory for specific behaviors related to LBDQ

dimensions. One questionnaire, the "leader quotation questionnaire," was designed to assess observers' recognition memory for what occurred in the group since it included quotations that were actually stated by the leader on the video. The other questionnaire, the "leader ideas questionnaire" included some ideas expressed by the leader but mostly ideas expressed by group members. Their use of pre and post-observational cues in accord with testing memory accuracy allowed them to test whether the performance cue effect could be explained by a selective availability in memory process or by a probabilistic response bias process.

Higher ratings were associated with the pre- and post-observation success cues than with the failure cues, although the effect of the pre-observation cues was somewhat weaker. Memory sensitivity scores produced no significant main effects for post-observation cues, but memory sensitivity was significantly greater for pre-observational cues and in the interaction between pre and post observational cues. In sum, performance cues were associated with ratings for both the pre-and post-observation conditions, but memory was more accurate when cues were given before observation. The interaction suggested that the combination of pre- and post-observational cues had an additive effect in improving accuracy of recall.

Their results suggested, first, that pre-observational feedback operated in a fashion that was consistent with the idea that there was selective availability of information in memory. In other words, pre-observational feedback prompted observers' attention such that incidents that were consistent with cues were stored in memory and recalled with higher accuracy. The dominance of cue consistent

information in memory resulted in ratings that were consistent with the cue. Second, observers in the post-observational success conditions over-attributed considerate, structuring, and problem solving behaviors to the leader. In other words, observers tended to report cue-consistent behavior but their reports were not accurate with what actually occurred. Larson et al. concluded that post-observational cues resulted in a probabilistic response bias but not a selective retrieval of cue consistent behaviors from memory, since ratings were more consistent with the cues but accuracy was lower.

Thus, elevations and depressions in the ratings are expected depending on the perceptions created about the performance level of the group and may occur as a result of attention/encoding or recall. Researchers have offered cognitive processing explanations involving attribution theory and schemata to explain the performance cue effect. Both of these will be conceptually explored in the following paragraphs, In the present study, I will examine whether attribution or schema explanations account for any effects of perceptions of group performance. The conceptual review of these explanations will lay the groundwork for predictions in the present study.

Attribution Explanations of the Performance Cue Effect

Causal attributions about individuals may account for the performance cue effect. This explanation suggests that raters who observe particular outcomes search for the causes of the outcomes. Thus, when raters believe that groups have performed well or poorly, raters attempt to specify either a person or the circumstances which caused the success or failure. McElroy and Downey's (1982) study and findings, cited previously, are good example of how attribution theory

can be applied to create specific hypotheses about performance cue effects.

Mitchell et al. (1977), cited previously, also interpreted their findings using actor and observer differences. Mitchell et al. suggested that actors/group members were more likely to interpret a leader's behavior as consistent with external circumstances (i.e. the performance cue) than observers who would tend to attribute group members' performance to persons rather than environmental cues. Raters were observers in studies one and two and actors in study three. Behavior ratings were significantly correlated with performance cues in studies one and two but a significant association was not found in study three. Thus, the significant findings in studies one and two may have reflected observers' tendency to attribute performance success or failure to the leaders' behaviors. It should be noted that situational ratings were associated with performance regardless of whether raters were observers or actors, so the prediction that performance cues would be associated with actors' but not observers' situational ratings did not hold.

Phillips and Lord (1981) spelled out the most detailed attribution explanation for the performance cue effect. Phillips and Lord (1981) began with a theoretical position based on Calder (1977), suggesting that leadership is attributed to actors when their performance is salient to observers and when they are seen as causal agents in the performance of others. In other words, when an observer observes or infers the achievement of results, the observer searches for possible causes of the results. If a particular actor is the most plausible cause for the outcome, the observer will perceive that the actor has

demonstrated leadership. Phillips and Lord, based on Kelley (1973), suggested that the probability that an observer would view an actor as a causal agent depends on: (a.) the availability of alternative causes and (b.) the salience of the actor. For example, regarding the availability of alternative causes, observers may hold the belief that low motivation is inconsistent with high performance. Therefore, the causal effect of a leader should be augmented if the group's performance is high but motivation or ability of non-leader, group members is perceived to be low. Regarding perceptual salience, observers may tend to rely on explanations that are most easily accessible; as actors are more salient, they are more likely to be seen as a leader.

In their study, Phillips and Lord (1981) varied salience of leadership by using two camera angles during video taping; the different angles affected the degree to which the assigned leaders were prominent on the video. They varied the configuration of plausible alternative causes by varying observers' perceptions of non-leader, group members' ability and motivation. In the leadership augmenting condition, perceptions were created that group members were low in ability and low in motivation so it was predicted that leaders would receive higher ratings. Lower ratings of the leader were expected in the discounting condition because non-leaders, portrayed as having high ability and high motivation, were more plausible causes of the group's performance. Performance cues were also given to observers to manipulate their perceptions of the groups' performance success or failure. Dependent variables included ratings of leader causality, Initiating Structure, Consideration, and a one item general leadership

impression index.

Phillips and Lord found that higher leader causality was ascribed to the target male leader in a four person group (two males, two females) in the high salience condition, augmenting, and high performance cue conditions. The interaction terms were most important to the hypotheses, since causal ascriptions were predicted to interact with perceptions of group performance. Salience of the leader and augmenting conditions resulted in higher Initiating structure and general leadership impression ratings. High performance cues and augmenting conditions were associated with higher Initiating Structure, Consideration, and general leadership impression ratings than low performance cue and discounting conditions. Thus, leaders were rated higher when the groups' performance was perceived as higher and when the target was the most plausible cause of the group's performance.

Phillips and Lord conducted further analyses to test whether the relationships between the experimental manipulations (salience, configuration, performance cue) and ratings were independently or jointly moderated by raters' attributions of causality to the leader and/or by a general leadership impression. In a series of hierarchical regressions, they found first that most of the variance in general leadership impressions created by the experimental manipulations could not be explained by attributions of causality. Second, when the general impression score was entered before the attributions of causality score in hierarchical regression, they found that general leadership impression accounted for almost all of the variance in the Initiating structure rating with attributions of causality accounting for non-significant amounts of variance. The same was not true for

Consideration; though general impression and attribution of causality each accounted for a significant portion of variance, significant variance remained unexplained after both variables were entered in the regression equation. Phillips and Lord's results suggested that ratings conformed to hypotheses based on causal schema or availability of alternative causes for performance, but a general leadership impression seemed to be a stronger explanation of these findings.

In sum, the studies here suggest that social judgments of leadership behavior in general and specifically performance cue effects can be understood within an attribution framework. Ratings of the target assessee tended to be higher (i.e. consistent with perceptions that the group performance was high) when the target was the most plausible cause of the group's high performance (i.e. in the augmenting composition A, L, L). Ratings of the target assessee tended to be lower (i.e. consistent with perceptions that the group performance was low) when the target was not the most plausible cause of the group's low performance (i.e. in the discounting composition A, H, H).

General Impression/Schema Explanations

A number of studies have examined whether observed data and knowledge of performance combine into a simplified general impression or schema that influences ratings. One example is Rush, Phillips and Lord's (1981) replication of Lord et al. (1978). Rush et al. added a temporal delay of 48 hours condition to the performance cue and behavioral information conditions. Their hypothesis was that the temporal delay would cause observers to rely on information-reduction heuristics decreasing the influence of behavioral information and increasing the influence of performance cues over time. In other

words, delayed conditions would cause raters to rely on a general impression or schema, formed in association with the performance cue, to facilitate memory.

Rush et al.'s results suggested that the delay led to a slight increase in the reliance on performance cues, but the behavioral information demonstrated a stable, significant effect despite the delay. Further analyses were used to explore the hypothesis that observers combined behavioral information and performance cues into a simplified impression (schema) and responded to the LBDQ based on this impression. A general leadership index (GLI) was calculated based on the subjects' responses to a single Likert scale item asking "how much leadership was exhibited by the group leader." When GLI was partialled out of LBDQ ratings, the previously significant effects of behavioral and performance information were reduced dramatically. A test of memory accuracy also suggested that recall accuracy decayed in the delayed condition as expected. All of these analyses suggested that observer's memory for actual behaviors decays over time, and they may maintain a simplified general impression of the person's performance and use the impression to infer probable levels of specific behavior when assessing leadership.

Phillips and Lord (1982) pursued further the hypothesis that a general impression intervenes between the observation and ratings of behavior. They hypothesized that observers would form and rely on a general impression to economically handle the wealth of incoming data. Instead of independently storing all information about each new leader, observers need only to equate incoming stimuli with an existing general impression or prototype of leadership. Despite the economy afforded by

this process, memory for specific behavioral occurrences is inhibited since observers have difficulty separating stimulus information from prototypes.

In their study, they manipulated performance cues, prototypicality of leader behavior at three levels--prototypically effective, prototypically ineffective, and non-prototypical--and presence/absence of actual behavior. Ratings were made immediately after observations. Study participants demonstrated the greatest accuracy in determining the presence or absence of non-prototypical behavior, though participants also demonstrated significant levels of accuracy in the prototypically effective condition. The effects of performance cues were quite pronounced in the prototypically effective and ineffective conditions. Specifically, participants tended to report more behaviors that were prototypically effective in the high performance cue condition and more behaviors that were prototypically ineffective in the low performance cue condition. These findings, overall, suggested that the biases toward reporting cue consistent behaviors were independent of the absence/presence of actual behavior.

Phillips (1984) also tested how the performance cue effect as mediated by a general impression was associated with memory accuracy. Performance cues were manipulated by informing the group that the assigned leader was very effective or very ineffective. Time of rating was manipulated by having observers rate immediately after observing or 24 hours after observing. Observers responded to a questionnaire of prototypically effective, prototypically ineffective, and nonprototypical behaviors, some of which occurred on the tape and some of which did not (presence/absence). There were no main effects for

time of cue. Consistent with previous research, observers reported more prototypically effective behavior when they were given effective performance cues and more prototypically ineffective behavior when ineffective cues were given. Overall, observers were significantly more accurate in reporting the occurrence of nonprototypic and prototypically effective behavior but not prototypically ineffective behaviors. Yet, in the delayed condition, observers were significantly more accurate in reporting the nonprototypic items but not the prototypically effective or ineffective behaviors. Also, the accuracy in reporting nonprototypic behaviors dropped considerably over time. Thus, observers showed a bias toward reporting behaviors that were prototypically consistent with the performance cue they were given. They showed better recall for behaviors that did not fit a particular prototype, but their accuracy decreased considerably (60% to 31% of variance) with only a 24 hour delay.

In sum, another plausible explanation of the performance cue effect is that perceptions of group performance contribute to the formation of a general impression of an individual as a leader. This general impression, then, may drive the ratings of the individual in such a way ratings are consistent with the general impression of the individual. In the present study, the assessor's general impression of the target as a leader was predicted to partially mediate the relationship between the perceptions of group performance and the ratings of the assessee.

Integration of Contrast and Assimilation Findings with Performance Cue Findings

Before summarizing this section, it is important to note that the literature reviewed to this point has revealed some different findings

regarding schema driven, information processing and accuracy in recalling consistent and/or inconsistent behavior. Findings presented in the contrast and assimilation section suggested that information that is inconsistent with its context was most salient and more accurately recalled (Maurer & Alexander, 1991). Findings presented in this section suggested similarly that behavior that is inconsistent with a post-observational performance cue was more accurately recalled (Larson et al. 1984). On the contrary, findings presented in this section also suggested that behavior that was consistent with a pre-observational cue was recalled more accurately than inconsistent behavior (Larson et al. 1984). A few comments here can integrate these seemingly contradictory findings and also shed some light on how these effects may manifest themselves in assessment centers.

Higgins and Bargh (1987) have emphasized the importance of differentiating between impression testing and impression formation when integrating the divergent empirical findings that either schema-consistent or schema-inconsistent information is remembered with greater accuracy. They suggested that when an impression is well formed consistent information is apt to be better recalled with inconsistent information being discarded as not being representative of the target. When the goal of the observer is to form an impression, however, there is greater likelihood that observers will attend to inconsistent behavior that distinguishes the target from its context. The findings that contrast effects may result from a focus on discrepant or inconsistent behavior (Balzer, 1986; Maurer & Alexander, 1991) and the finding that inconsistent information is recalled with greater accuracy in post-observational cue conditions (Larson et al.

1984) is compatible with an understanding of these situations as an impression formation process.

In assessment centers, steps are commonly taken to assign assessors to observe and rate persons with whom they have no previous experience or contact. In this way, assessment center evaluation is an impression formation process. This reinforces the prediction of a contrast effect for evaluations of assessees in groups since information that is inconsistent with it context is most likely to receive attention during impression formation (Maurer & Alexander, 1991). Further, assessors do not receive any information that would create expectations about the performance of a group before the group performance. Assessors are apt to form an impression of the group's performance dynamically, as they observe the progress of the group. Thus, their perceptions of group performance are more likely to conform to the findings of the post observational cue research. In other words, reports of behavior and evaluations of assessees tend to be consistent with cues although neither may accurately reflect the behaviors that occurred. Thus, in order for performance cue effects to occur in assessment centers, assessors must abandon the behavioral notes they have taken and make their evaluation in a way that is consistent with their knowledge of and attributions about performance.

Group Performance and Group Composition Hypotheses

All of the research cited here reinforces the importance of understanding the association between observers' beliefs about a group's performance and their responses to questionnaires or ratings regarding the group or individuals. Specifically, raters' perceptions about the success or failure of the group relative to performance

outcomes influence ratings of process and outcome behaviors and other characteristics of individuals and the group. Many studies cited here suggest that the performance cue effect is rather robust. Possible explanations of the effect suggest that behavioral or performance information may be combined into a general impression or schema that influences evaluations or that assessors' attributional tendencies are apt to reinforce raters' notions of which persons may be causal agents in performance.

Hypothesis 2. Ratings of the ambiguous target were expected to be associated with the experimental manipulations of perceptions of group performance. Assessors informed that the group is a high performing group were expected to rate the target assessee higher. Assessors informed that the group was a low performing group were expected to rate the target assessee lower.

Hypothesis 2A. It was predicted that assessors' general impression of the target in the role of the leader would partially mediate the relationship between the manipulation of perceptions of group performance and the ratings.

This hypothesis, depicted in Figure 3, was pursued in order to further explain the relationship between the group performance manipulation and ratings. Assessors were be asked a series of questions to access their general impression of the target's performance in his role as a leader. It has been found in performance cue research that behavioral data and contextual information is integrated into a simple, more easily remembered form and used to efficiently infer probable levels of specific leader behaviors (Phillips & Lord, 1981; Rush, Phillips, & Lord, 1981). In the items

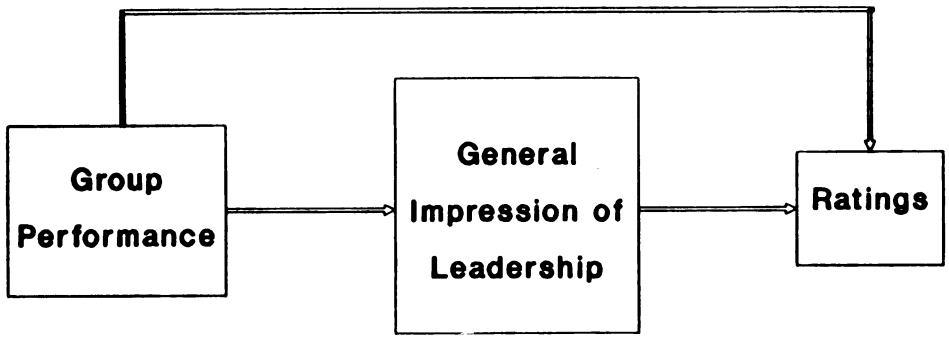

Figure 3

Hypothesis 2A: General Impression of Leadership as a Partial Mediator of the Relationship between Group Performance and Ratings

measuring general impression, assessors were asked about their perceptions of leadership without defining leadership. General impression measures of leadership such as this are thought to access impressions that differ from ratings of performance on behavioral anchored rating scales (Phillips & Lord, 1981), have been found to be related to objective indices of leadership (Lord, 1977, Lord, Phillips, & Rush, 1980), and have been closely related to behavioral reports and performance induced distortions in ratings (Rush, Phillips, & Lord, 1981).

Hypothesis 3. Knowledge of group performance created experimentally is predicted to interact with the group composition condition in predicting the ratings of the target.

The group compositions in this study included the average target with two high performing non-targets (A,H,H) and the average target with two low performing non-targets (A,L,L). These different group compositions were expected to influence assessors' attributions because the non-target individuals performing at different levels create different patterns of alternative causes to explain group performance. When the group performance manipulation was high and the group composition condition is low (A,L,L), the target was expected to get higher ratings than in the high performance, high composition condition (A,H,H) because he was the most plausible cause of the group's high performance. In the low group performance condition, assessors were expected to rate the target generally lower than in the high performance condition, but ratings were expected to be slightly lower when group composition was high (A,H,H,) than when group composition was low (A,L,L) because the target would be the most plausible cause of

the group's low performance. The expected mean differences are depicted in Figure 4.

Hypothesis 3A. It was predicted that assessors will attribute higher responsibility for the group's performance to the target and rate the target higher when group performance was perceived as high and when the target was the most plausible cause of group performance (i.e. in the low composition condition--A,L,L).

This hypothesis, depicted in Figure 5, was pursued in order to explain the relationship between group performance and group composition and ratings. Assessors were asked the extent to which they attributed responsibility for group performance to the target versus other group members. A number of research studies have suggested that attributional processes are employed in a stage of controlled processing of performance information (DeNisi et al. 1984; Feldman, 1981) and have confirmed that attributions mediate the relationship between observed data and ratings or conclusions drawn about observations (Green & Mitchell, 1979, Phillips & Lord, 1981).

Further, as was argued previously, the performance of other assessees in a group establish different conditions of plausible causes of group performance. This is even further reinforced by observers' attributional tendencies to search for internal, personal causes and not situational causes of performance involving others and by assessor training which teaches assessors that performance results from assessees' stable traits. Thus, assessors' attributions of responsibility for group performance to the target are hypothesized to partially mediate the relationship between the manipulations of group performance and group composition and the ratings.

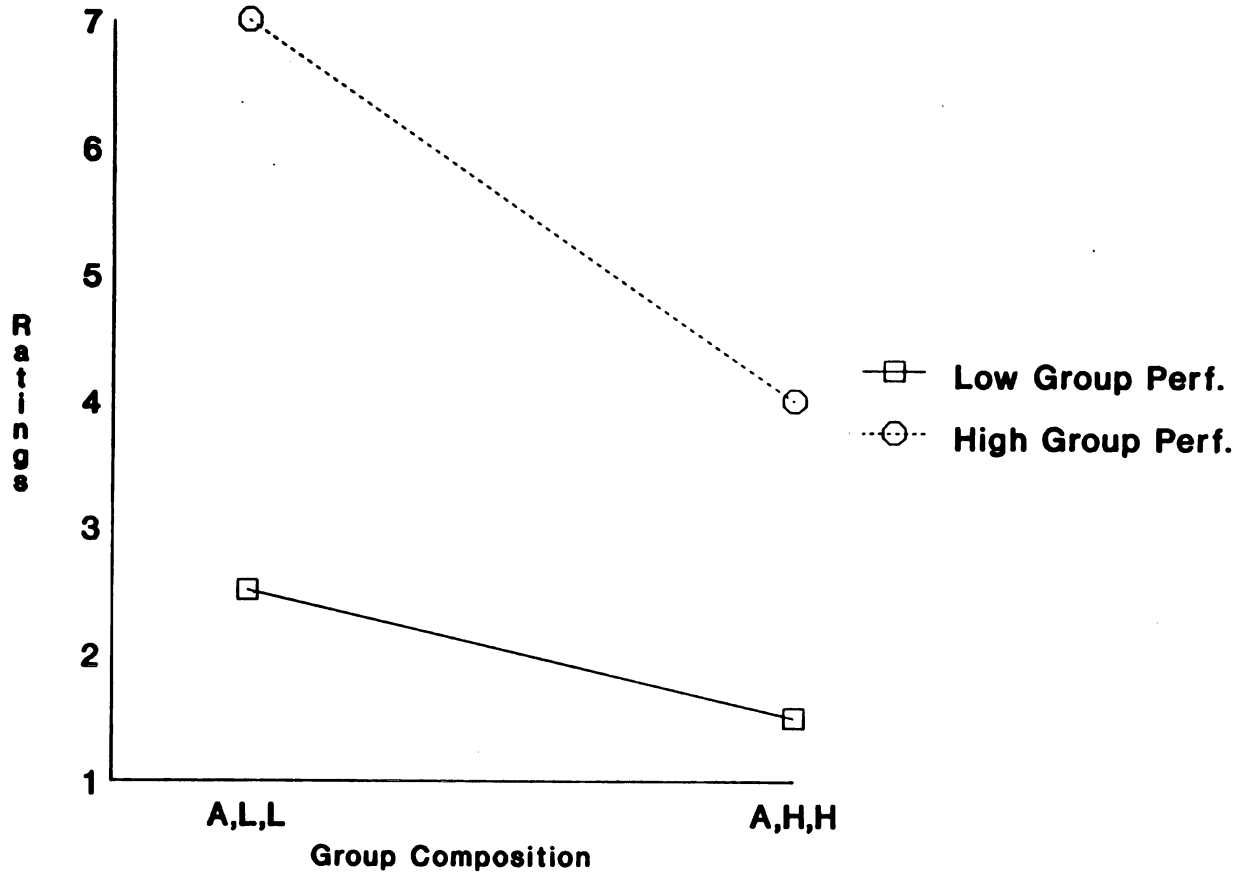

Figure 4

Hypothesis 3: Predicted Mean Differences in Ratings for the Group Composition by Group Performance Interaction

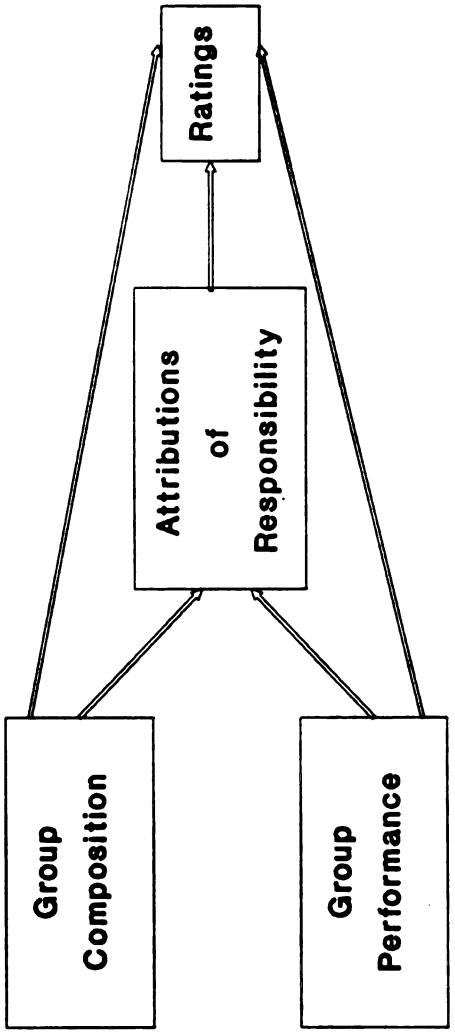

Figure 5

Hypothesis 3A: Attribution of Responsibility as a Partial Mediator of the Relationship Between Group Composition and Group Performance and Ratings

Summary

The hypotheses proposed so far have involved group composition and group performance and their combined effects. Contrast/assimilation and attributions have been offered as cognitive mediators of group composition and group performance effect. Figure 6 depicts the relationships predicted thus far. Figure 6 is presented for summary purposes only; no additional hypotheses are offered at this time.

Other members of a group whom an assessor is not assigned to rate may influence the ratings of a target assessee (i.e., a group composition effect). This effect may occur by means of a contrast effect, and thus, perceived dissimilarity between the target's performance and non-targets' performance was predicted to moderate the relationship between group composition and ratings. Group composition may also influence the attributions that assessors make about the targets and attributions were predicted to influence ratings. Group performance is predicted to have an effect on ratings of a target and to interact with group composition and affect ratings as mediated by attributions. Additionally, group performance is also predicted to result in the formation of a general impression of the target in a leadership role which in turn is predicted to influence ratings.

Task Interdependence

The previous sections have examined the extent to which beliefs that raters hold about personality traits, leadership, and group performance achievements influence their evaluations of individuals in groups. It is advanced here that assessors' beliefs about exercise tasks have the same kinds of influence. A number of researchers have suggested that group tasks cue up individual group members' beliefs

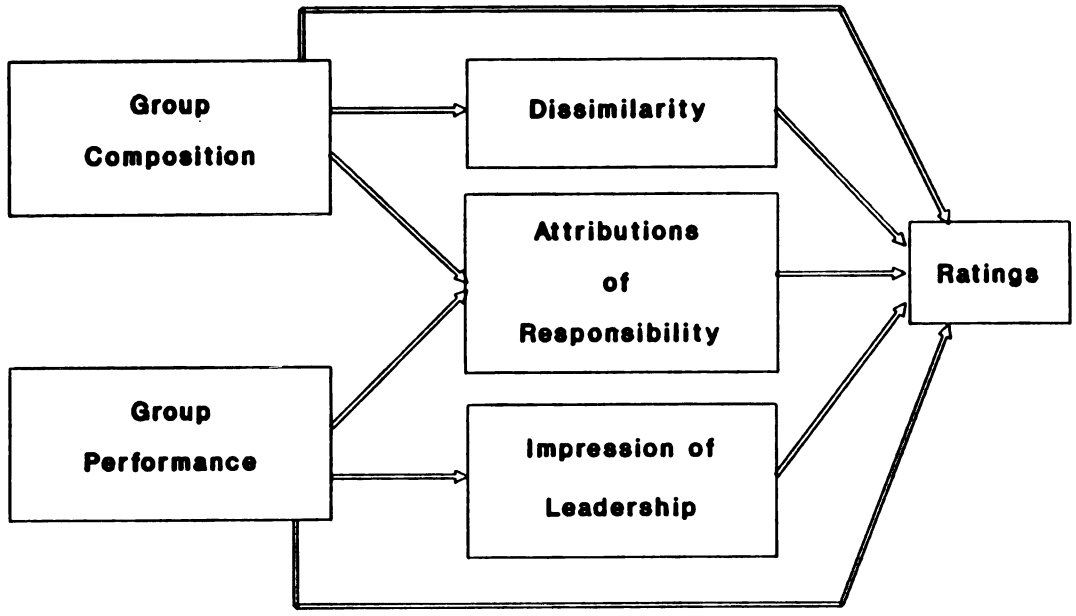

Figure 6

Summary of Main and Mediated Effects of Group Composition and Group Performance on Ratings

about what are proper strategies and responses to the task (Kabanoff & O'Brien, 1979; Mitchell & Silver, 1990). Similarly, assessors' understanding of exercise task characteristics may influence their expectations for and interpretations of assessees' behaviors.

The assigned role, group discussion was chosen as the prototype for the exercises in the present study because Thornton and Byham's (1982) statistics suggest that it is a commonly used group exercise, and Thornton (1992) suggested that it is content valid for managerial jobs. Different exercise tasks were created within the assigned leader group discussion by manipulating assessors' perceptions of the interdependence among group members. Specifically, perceptions of interdependence were varied by varying what assessors were told about the distribution of problem-related information to assessees.

Manipulating task interdependence by means of changing perceptions about the distribution of information is consistent with how the literature defines task interdependence. A number of researchers have suggested that interdependence is related to the distribution and scarcity of resources (Kiggundu, 1981; Thompson, 1967). Thompson (1967), for example, has suggested that dependence and power are related. As organizations attempt to deal with environmental uncertainty, units who have resources to deal with uncertainty can become powerful and units who need access to resources become dependent. In the same way, distribution of information resources in assessment center groups creates power/dependency relationships. When information is not distributed to all assessees in a group (i.e. high interdependence), each individual is dependent on each other for an important and scarce resource.

Support for the hypothesized effect of raters' perceptions of interdependence on ratings comes from Liden and Mitchell (1983). Liden and Mitchell suggested that it may be difficult for supervisors to evaluate the contributions of individual group members when group members are perceived as highly interdependent. They also suggested that an empirical finding similar to a contrast effect may occur in evaluations of members in an interdependent group--a poor performer is apt to be rated higher and a good performer lower when the group is highly interdependent.

In their study, undergraduate raters read and rated written descriptions of three person groups consisting of two good performers and one poor performer. Liden and Mitchell manipulated the information given to raters about the interdependence of work group members. Raters in the high interdependence condition were told that three page written documents were created from a collection of ideas from all group members and endorsed by all group members (high task interdependence). Raters in the low interdependence condition were told that each group member prepared one page of the three page document. The poor performer was rated higher (i.e. more similar to the good performers) in the high task interdependent condition than in the low task interdependence condition. Their findings supported the notion that assessors, viewing individuals performing in a cooperative--high task interdependent group, may make similar attributions to all individuals in the group. In competitive, low task interdependent exercises, assessors may be more likely to attribute group performance outcomes to specific individuals and not to all members of the group.

In sum, task designs that impose interdependence through information distribution can affect the behaviors of individuals in groups. The work of Liden and Mitchell (1983) suggests further that raters' perceptions of task interdependence can affect their ratings, even when their ratings are based on the behavioral data is held constant. Effects associated with assessment center exercises may be a function of assessor perceptions of the task and the kinds of expectations and norms for performance that are cued by knowledge about the task. The examination of assessors' perceptions of interdependence among group members in the present study can make a significant contribution to understanding exercise effects in assessment centers.

Hypothesis 4. Assessors' perceptions of interdependence among group members is expected to moderate the relationship between group composition and ratings. The expected trends in the mean ratings is depicted in Figure 7.

Based on Liden and Mitchell (1983), target assessees will be rated in contrast to the scripted performance levels of the non-target assessees when interdependence among group members is perceived as low. In other words, a contrast effect will occur in which the target is rated lower when working with high performing, non-target assessees and higher when working with low performing, non-target assessees. When interdependence among group members is perceived as high, target assessees' ratings will be assimilated to the level of the scripted performance of non-target assessees. In other words, the target will be rated higher when working with high performing non-target assessees and lower when working with low performing, non-target assessees. This hypothesis competes with the predictions of hypothesis one, since task

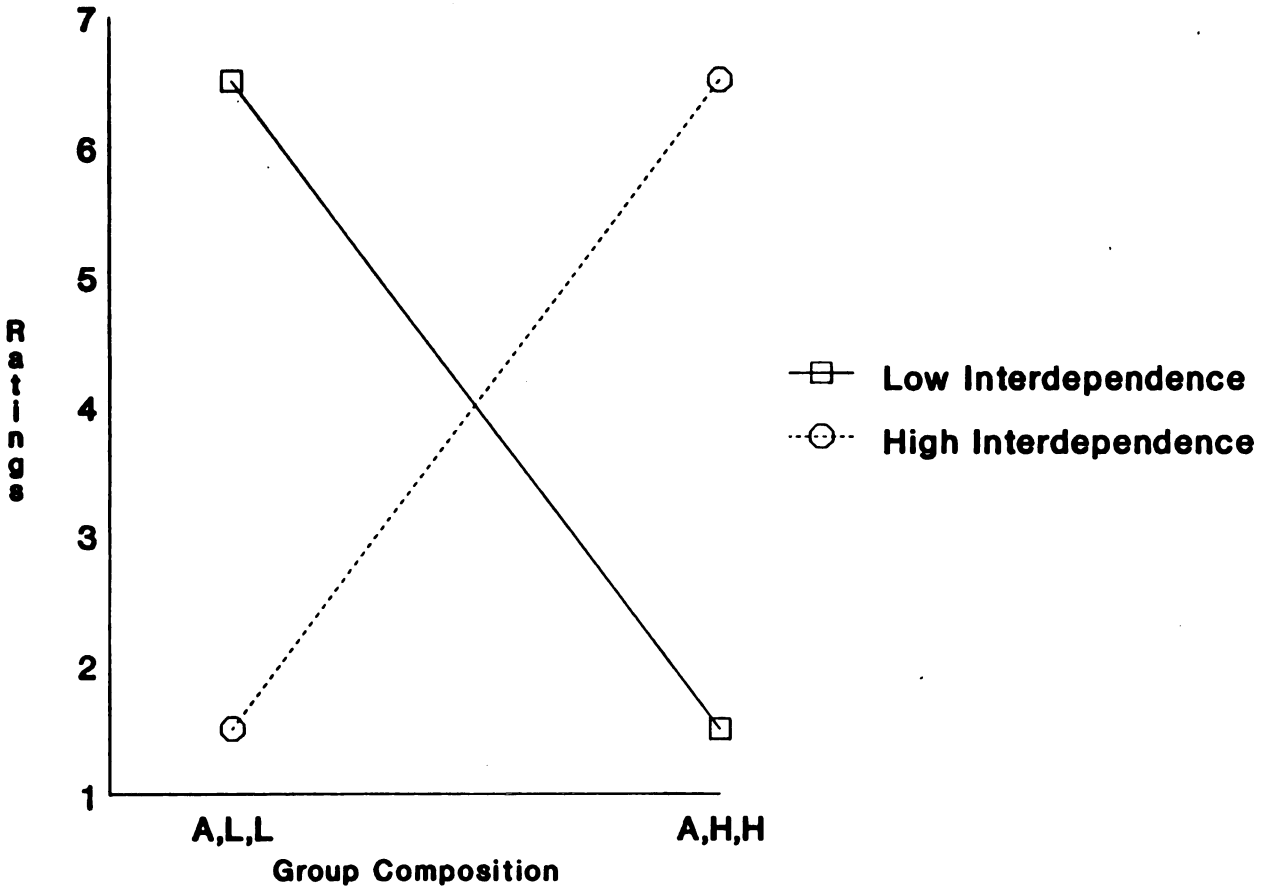

Figure 7
Hypothesis 4: Predicted Mean Differences in Ratings for the Group Composition by Interdependence Interaction

interdependence is hypothesized here alter whether a contrast or assimilation effect is predicted.

Summary and Conclusions

Ratings of assessees in groups may be influenced by a variety of stimuli and not only the target assessee's behavior. This review has shown that other individuals in a group exercise, group performance, and interdependence are available for assessors' attention and may influence ratings.

Specific Predictions

First, assessors' attention, recall, and judgments are apt to be influenced by beliefs about and standards of performance held by assessors. Though it is thought that assessor training sets this standard normatively, past experience with persons similar to the assessee, the assessors' self-concept, or other assessees observed at the same time may influence standards assessors employ. Particularly in group exercises, ratings of assessees may be influenced by simultaneous observations of other assessees in the group. This effect may occur by means of either or both of two cognitive processes--contrast/assimilation or attributions.

Second, assessors' knowledge of the group's success or failure can influence attributions about and evaluations of assessees. On one hand, perceptions of group performance may be combined with behavioral data into a general impression of the assessee's performance and this impression may influence ratings. On the other hand, assessors perceiving group performance achievements and holding the implicit belief that group performance is initiated by a causal agent or leader are apt to search for a particular assessee who is the most plausible

cause of the group's performance. Attributions of responsibility are apt to be shaped not only by assessors' perceptions of group performance but also by the performance of other assessees who represent alternative, plausible causes of performance.

Third, assessors' beliefs about interdependence among group members are apt to influence ratings. When group members are perceived as interdependent, it may be difficult for assessors to attribute performance and behavior to one particular individual without crediting another. Specifically, when group members are perceived as interdependent, it may be difficult for raters to attribute performance to one particular assessee without crediting another and an assimilation effect may result. Assessors may not have the same difficulty distinguishing the performance of one individual from another when they assume group members are independent, and a contrast effect may result.

Potential Contributions

My pursuit of the hypotheses in this study has potential to make a number of contributions. First, the many assessees who are evaluated annually have a right to be evaluated as accurately as possible. Therefore, an understanding of contextual, non-behavioral influences on assessee ratings is important, and it may be possible to reduce problematic contextual biases once they are understood. Second, assessment center researchers have not been able to answer construct validity-oriented questions about what assessment centers measure, despite many different factor analytic examinations. This study takes a different approach by attempting to model the assessor-cognitive process of rating. A better understanding of how assessors arrive at their

ratings may ultimately put researchers in a better position to answer the question of what assessment centers measure.

METHOD

Overview and Design

This experiment was a two (high vs. low group composition) by two (high vs. low group performance) by two (high vs. low interdependence between group members) between subjects design with an appended control group. Assessors, the participants in the study, observed and rated the videotaped performance of a single, target assessee in three-person group discussion exercises. Group composition for the experimental conditions was manipulated by scripting the performances of the two, non-target assessees at high levels in one tape and at low levels in another. Group performance and interdependence among group members were manipulated through information given in assessor training. For the group performance manipulation, assessors were shown frequency distributions of the number of cases typically finished by groups to create perceptions that the number of cases finished by the group they observed was either above average (high) or below average (low). Interdependence was manipulated by telling assessors either that each assessee had unique information about the case studies in the exercise and that sharing information was important for accurate problem solving (high interdependence) or that assessees had equal information that was sufficient to solve the problems without the help of the others (low interdependence). A control group of assessees viewed a video tape in which the target assessee was paired with one high performer and one

low performer, creating a neutral condition of group composition. No information about group performance or interdependence of group members was given to the control group.

The dimension ratings made by the assessors were the primary dependent variables in the study. Assessors also responded to a questionnaire which included items checking the manipulations of the independent variables and measuring three psychological variables--impression of leadership, attributions of responsibility for group performance, and dissimilarity of the performance of the target with other group members. Analyses were conducted to determine whether these variables partially mediated the relationship between the independent and dependent variables.

The construction of the stimulus materials, the development of measures, and the procedure for this study are described in the paragraphs that follow. The focal study was conducted with 167 participants (assessors). Two pilot studies were conducted since the materials and measures had not been used prior to this study. The first pilot study was conducted with 41 participants, and the results from this sample were used to confirm the effectiveness of and/or modify the stimulus materials and the measures. A second pilot study was conducted with 13 participants to determine if the modified stimulus materials and measures produced the expected results. Since the second pilot study confirmed that the manipulations and measures conformed to expectations, the majority of this section is devoted to interpreting the results from the first pilot study and explaining the modifications made as a result of the first pilot study.

Participants

The participant/assessors in both pilot studies and in the full experiment were undergraduate students. Students participated to fulfill course requirements for upper level industrial and organizational and for general psychology classes. Participants in the first pilot included 16 males and 25 females. Participants in the second pilot included 5 males and 8 females. Participants in the study included 71 males and 92 females; four participants did not report their gender. Participants were informed about the procedure and consented to participate in the study before the session and were debriefed about the intentions of the research after the study. Consent and debriefing materials can be found in Appendix A. Participants were assigned randomly to a particular combination of the independent variables or to the control condition and watched only one exercise. Moderate effect sizes were expected in this study, and statistical power was estimated to be .80 with sixteen participants per cell (Cohen, 1988). This estimate resulted in a target of one hundred, forty four assessors distributed across the nine cells in this study. This target was exceeded when the study was conducted. All of the cells had over seventeen participants.

Students were chosen as participants of this study so that previous experience would not interfere with the manipulation of the independent variables. For example, if the participants in this study had previous experience as assessors, they may have relied on their experience instead of the information given to them in training to judge whether a group's achievements were high or low. Compared to the typical practice of using assessment centers in organizations, the

students were thought to be more similar to internal organizational assessors than to experienced assessors such as psychologists. Internal organizational assessors--managers, supervisors, or incumbents--know the performance domain well but may only serve one to two times per year and two to five times during their tenure (Schmitt, Schneider, & Cohen, 1990). Like organizational assessors, the student assessors were experienced and familiar with the domain of task performance (i.e. students working on a class project/case studies in industrial psychology) and did not have extensive experience as assessors. Students in this study may have differed in their motivation to make accurate assessments from organizational member assessors whose ratings are considered in decisions that affect their organization or work unit.

Dimensions

As is common in assessment centers that are designed for organizations, the development of the exercises and scoring guidelines in this study began by defining the performance domain to be assessed. The four dimensions initially selected were Problem Solving, Empathy, Setting Directions and Goals, and Motivation. These four dimensions were based on three factors--an intellectual ability/problem solving factor, two dimensions from an interpersonal/leadership factor, and a motivation/activity factor. Factor analysis studies of assessment center final ratings have commonly demonstrated this three factor solution (Russell, 1985; Sackett & Hakel, 1979; Schmitt 1977).

For the purposes of this study, the interpersonal/leadership factor was broken down into two dimensions that resembled the LBDQ's consideration and initiating structure dimensions. Some support for

this break down has also been found in assessment center factor analyses (Archambeau, 1979; Sackett & Hakel, 1979). This break down was also consistent with factor analyses on leadership (Fleishman & Harris, 1962), and its use in this study facilitated the merging of this research with the theory and findings from the leadership research on performance cue effects. The labels "leadership," "initiating structure," or "consideration" were not used in order to avoid specifically cuing assessors theories of leadership and to avoid a confound with the measures of general impressions of leadership.

The four dimensions were defined with behavioral examples generated from dimensions used in other assessment centers, particularly from a center developed by Schneider (1990). The Schneider center was used to evaluate student performance against educational objectives. Problem solving behaviors included seeking and integrating information, recommending alternative solutions, and making decisions through logical reasoning and analysis. Empathy behaviors included reinforcing and promoting teamwork and rapport with others, compromising in support of team objectives, and non-verbally demonstrating warmth and openness. Setting Directions and Goals behaviors included stating ideas confidently and directly, and setting direction and goals for others. Motivation behaviors included demonstrating motivation through active participation. Dimension definitions and high, average, and low behaviorally anchored scoring guidelines can be found in Appendix B.

During the pilot study, it was observed that assessors had difficulty distinguishing between Setting Directions and Goals and Motivation, and the data confirmed this observation. Setting

Directions and Goals and Motivation were correlated .69 (N = 38, p < .001--the three subjects in the control group were not included in this analysis) so these dimensions were combined into a single dimension called Initiative/Assertiveness. The label for the Empathy dimension was changed to Managing Interpersonal Relationships because it was observed that the Empathy label tended to bias assessors' observations and ratings in favor of anchor behaviors such as listening and against anchor behaviors such as dealing with conflict. The label, Managing Interpersonal Relationships, more comprehensively captured the range of behaviors defining this performance dimension. The definitions for the modified dimensions appear in Table 1, and the scoring guidelines for the three modified dimensions also appear in Appendix B.

Case Study Exercise

The assigned role, group discussion exercise was the prototype for the exercise in this study. Assessors were told that the target assessee was given the role of assigned leader in the three person group. The content of the exercise consisted of seven industrial and organizational psychology case studies. All of the cases were described in a packet of written materials that was given to assessors. The written materials for each case described an organization with a human resources-related problem that could be solved by applying theories from industrial and organizational psychology. The assessor packet also included materials describing the correct solution to each case. Assessors were led to believe that they were observing and rating assessees from an advanced class in industrial and organizational psychology. The materials describing the seven cases can be found in Appendix C.

Table 1

Behavioral Definitions for the Final Set of Dimensions

Problem Solving
-- Makes factual information known to the group
-- Adds own experience to facts to help solve the case
-- States and applies Organizational Psychology Theory
-- Offers practical recommendations to the problem
Managing Interpersonal Relationships
-- Cooperates (but does not go along with incorrect answers)
-- Tries to lighten up the group with sense of humor and wit
-- Asks others for their opinions--gets quiet members talking
-- Deals with conflict among group members
-- Listens to and maintains eye contact with the speaker
Initiative/Assertiveness
-- States opinion forcefully and confidently
-- Provides structure to the assignment, gives the group direction
-- Remains active and involved
-- Provides directions to the group when its stuck or stalled

Assessor Training

Assessors participated in a one and one-half hour assessor training program before watching the exercise condition to which they were assigned. The agenda and scripts for the assessor training can be found in Appendix D. The program included forty minutes of instruction in understanding dimensions and cases, observing and rating behavior, and avoiding halo, leniency, and central tendency errors. Fifty minutes of the program were devoted to practice during which assessors observed, used the scoring guidelines, rated and discussed their ratings of three performance excerpts. One practice excerpt showed the practice target performing a majority of high behaviors; one practice excerpt showed the practice target performing at an ambiguous level involving a mix of low and high behaviors at relatively equal frequency; and one practice excerpt showed the practice target performing a majority of low behaviors. In all of the three practice excerpts, assessors observed and rated a male target who was the assigned leader of a group composed of the target, one male, and one female.

A number of steps were taken so that the instruction and the observations and ratings during the practice excerpts would not interfere with the manipulation of the independent variables in the experimental and control conditions. First, the majority of the training was video taped including the segments of the training during which information creating the group performance and interdependence manipulations was provided. This ensured that the manipulations were consistent across trainee groups and prevented experimenter/trainer bias. Second, the practice assessees worked on the three of the seven

cases that were not in the experimental or control conditions, and excerpts rather than full length exercises were used for practice. In this way, an attempt was made to use case-related performance in the practice excerpts that would not establish particular expectations for the performance of assessees or about the number of cases that a group finished. Third, the three actors in the practice tapes were not the same actors in the experimental and control tapes so that there would be no priming effect due to previous exposure to the same assessees. Fourth, the order of presentation of the practice excerpts was counterbalanced across all of the assessor groups.

Assessor training was also used as the vehicle to create perceptions about group performance achievements and about the interdependence among group members. These manipulations will be described in more detail in the section titled, independent variables.

Independent Variables

The following section describes the development of the stimulus materials used to manipulate group composition, perceptions of group performance, and perceptions of interdependence among group members. The first pilot was used as a means to assess the strength of these manipulations prior to conducting the full experiment. The second pilot was used to verify that the modifications made after the first pilot led to the predicted results. For the first pilot, participants participated in three of the potential nine cells in the design. One cell employed in the pilot involved low composition (i.e. low, non-target performance), low group performance, and low interdependence. Another cell employed in the first pilot involved the control condition involving one high and one low non-target performers and no group

performance or interdependence information. The final cell in the first pilot involved non-target performances at high levels (A,H,H), high group performance, and high interdependence. The chief goal of the second pilot study was to evaluate the new items in the interdependence, group performance, attributions, and dissimilarity scales. No changes were made regarding the group composition manipulation so the data used in evaluating the manipulation in the second pilot is not presented. The cells employed in the second pilot study included one cell with high group performance high interdependence and high group composition and a second cell included low group performance, low interdependence, and high group composition.

Group Composition

Group composition was manipulated by scripting the performance of the actors on the video tapes. The end goal was to create three exercise videos representing two experimental conditions (high vs. low group composition) and a control condition. Another important goal was to create assessees whose behavior was within the boundaries of common occurrence and that, in meeting the objectives of the manipulation, did not result in unrealistic mixes of behaviors. As preparation for creating the tapes, the researcher viewed video tapes of 15 leaderless group discussion exercises from a center in which students discussed education-related issues (Schneider, 1990). Counts of anchored behavior were compared with ratings from the center issued by trained assessors in order to determine what mix of high and low anchored behaviors was associated with different levels of ratings. Before scripting the actual behaviors, an overall description of the assessees' personality was created so that the scripted behaviors would

result in an assessee whose behaviors showed some consistency. The scripted behaviors for the target and non-targets were created based on the dimension anchors to achieve the different levels of performance. The scripts can be found in Appendix E.

Professional actors, who were undergraduates, were hired to play scripted roles in the stimulus videos. The target assessee was a male who repeated the same performance in all three conditions. A female and a male were cast in the roles of the other group members. In the low group composition condition, the two others were scripted to perform a majority of low anchored behaviors. In the high group composition condition, the two others performed a majority of high anchored behaviors. In the control condition, the female repeated her high performance and the male repeated his low performance.

The male, target assessee was scripted to perform an even mix of high and low anchored behavior for all tapes. This mix of high and low behaviors established him as an ambiguous target whose ratings would be most susceptible to contextual influences. The target assessee performed twelve high anchored and twelve low anchored Problem Solving behaviors, four high and four low Managing Interpersonal Relations behaviors, seven high and seven low Initiative/Assertiveness behaviors. In addition, the target shifted his non-verbal behavior between being attentive, involved, and maintaining eye contact and being withdrawn and shuffling through papers. The target repeated this same performance across all experimental and control conditions.

One of the non-target, group members was a female who performed at a low level in the low composition video and at a high level in the high composition video, and she repeated the high performance in the

video for the control condition. In the low performance condition, she performed twelve low and ten high Problem Solving behaviors, twelve low and zero high Managing Interpersonal Relations behaviors, and four low and two high Initiative/Assertiveness behaviors. As the counts of behavior indicate, she offered many good ideas for solving problems that were balanced by poor ideas. She was very abrupt and critical interpersonally and initiated and sustained conflicts with the other non-target group member. In the high condition, she performed sixteen high anchored and six low anchored problem solving behaviors, seven high and three low Managing Interpersonal Relations behaviors, and six high and three low Initiative/Assertiveness behaviors. Approximately seventy-five percent of her Problem Solving behaviors, thirty percent of her Managing Interpersonal Relations behaviors, and forty percent of her Initiative/Assertiveness behaviors were exactly the same in the high and low conditions. Her high performance differed from her low performance in that she stated many of her ideas and solutions with much more tact and grace. She caused less conflict in the high condition but still created some tension by using sarcasm when criticizing others. In this way, her performance in the high condition was not a perfect performance in stark contrast to her poor performance in the low condition.

The other non-target, group member was a male. He performed at low level in the low composition condition and at a high level in the high composition condition, and he repeated the low performance for the control condition. In the low performance condition, he performed twelve low and nine high Problem Solving Behaviors, fourteen low and zero high Managing Interpersonal Relations behaviors, and five low and

two high Initiative/Assertiveness behaviors. Overall, most of his ideas and solutions were incorrect. When criticized, particularly by the female, non-target, he either attacked or withdrew. He said nothing during the discussion of one of the four cases. In the high condition, he performed sixteen high anchored and four low anchored Problem Solving behaviors, six high and three low Managing Interpersonal Relations behaviors, and five high and two low Initiative/Assertiveness behaviors. Fifty percent of his Problem Solving behaviors, ten percent of his Managing Interpersonal Relations behaviors, and thirty percent of his Initiative/Assertiveness behaviors were exactly the same in the high and low conditions. He was much more cooperative, less withdrawn, and less involved in conflict in the high condition than in the low condition.

The group composition manipulations were checked in the pilot study. Four different video tapes were used in the first pilot--one tape representing the low composition condition, two different tapes representing the high composition condition, and one tape representing the control condition. Two different high composition tapes were tested because of uncertainty about which tape would best achieve the manipulation. Assessors were asked in the questionnaire to rate the non-target assessees on the four dimensions. Means and standard deviations of these ratings for the levels of group composition are listed in Table 2.

Overall, the manipulations of the non-target's performances were successful in creating the desired group composition effects. The male's and the female's mean ratings on all the dimensions were low (i.e. under four on a seven point scale) in the low composition

Table 2

Mean Ratings of Non-target Assessees in Various Group Composition Conditions from the First Pilot Study

	ALL (N=11)	AHH_1 (N=18)	AHH_2 (N=9)	AHL (N=3)
Problem Solving				
Male	2.45(.52)	4.83(1.20)	3.44(1.59)	2.00(1.00)
Female	3.36(1.21)	4.50(1.50)	4.56(1.33)	4.33(.58)
Empathy				
Male	1.64(.67)	3.28(.96)	2.78(1.20)	1.33(.58)
Female	1.91(.94)	3.00(1.81)	3.33(1.22)	4.33(.58)
Setting Direction and Goals				
Male	2.09(.83)	4.39(1.72)	2.89(1.27)	1.67(.58)
Female	2.82(.98)	4.00(1.50)	4.00(1.50)	5.67(.58)
Motivation				
Male	2.36(.92)	4.69(1.49)	2.89(1.76)	1.33(.58)
Female	2.91(1.45)	4.54(1.61)	4.78(1.72)	6.67(.58)

Note: ALL = Target with two low performing group members; AHH_1 & AHH_2 = Two different videos of the Target with two high performing group members, AHL = Target with one high (female) and one low (male) performing group member.

condition (A,L,L). The findings were similarly supportive in the A,H,H_1, high composition condition so this tape was chosen as the high composition tape. The exception was the empathy ratings which were not above four for either the male or the female. This finding led partly to the change of the title from Empathy to Managing Interpersonal Relations. Both the male and female were assertive and engaged in some conflict in the high composition condition. Based on feedback from the participants, it was hypothesized that the male and the female would be not be penalized as much (i.e. rated lower) if the dimension label was changed to reflect the wide range of interpersonal skills encompassed by the dimension and not just the empathy-oriented behaviors.

A one-way analysis of variance was conducted to determine whether the mean differences for the high and low composition conditions were significant. The dependent variables were means calculated from the male's and female's ratings for each dimension. The correlations between the male's and female's rating were Problem solving = .39, Empathy = .27, Setting Direction and Goals = .31, and Motivation = .35. Mean differences were significant for Problem Solving (F (1,27) = 27.2, $p \leq .001$), Empathy (F (1,27) = 13.7, $p \leq .001$), Setting Directions and Goals (F (1,27) = 18.7, $p \leq .001$), and Motivation (F (1,27) = 22.7, $p \leq .001$). The means in the control group were also at hypothesized levels, though it is worth noting that the female was rated much higher in the control condition when the non-target male's performance was low than in the AHH_1 condition when the non-target male's performance was high, even though the female's high performance was virtually the same.

Group Performance

The manipulation of group performance was accomplished by showing the assessors different distributions of the number of cases that groups "typically" finish in the exercise. The groups on the experimental and control videos finished three cases totally and finished about 1/3 of a fourth case. In the high group performance condition, assessors were shown a distribution that suggested that a typical group finishes only one or two cases, hopefully leading assessors to believe that the group on tape was a highly productive group. In the low group performance condition, assessors were shown a distribution that suggested that a typical group finished six or seven cases, leading assessors to believe that the group on tape was a highly unproductive group.

Three items were included in the questionnaire asking assessors about their perceptions of the group's performance. Item intercorrelations and alpha for the scale were not high, so the means and item intercorrelations were examined. The means, appearing in Table 3, were in the predicted direction for item 19 but not for items 20 and 21.

Discussions with participants revealed that they rated items 20 and 21 based on performance factors other than the number of cases finished, including the extent to which they believed that the group solved the cases accurately. Based on the data and the feedback from participants, item 19 was retained. Two new items were written and were worded so that the participants would respond to the items based on the number of cases that the group finished. The final versions of the items in this scale appear in Table 4.

Table 3

First Pilot Results Examining the Manipulation of Perceptions of Group Performance

Item	High Mean(SD)	Low Mean(SD)
19. How do you rate this group in terms of the number of cases they finished?	3.37(.88)	2.64(.92)
20. How would you rate the overall performance of this group?	1.96(.71)	2.45(.82)
21. All things considered, this group was	2.33(.88)	2.64(1.12)

Note: Response format was a five point scale with 1 = Far Below Average and 5 = Far above average. Means are listed first with standard deviations in parentheses. High group performance condition (N=27); Low group performance condition (L=11).

Table 4

Group Performance: Final Items

16. How do you rate this group in terms of the number of cases they finished?
17. How productive was this group given what you learned during the training about how many cases groups typically finish?
18. Based on what you have learned about the number of cases that groups typically finish, how do you rate this group's level of achievement?

Note: Response format was a five point scale with 1 = Far Below Average, 2 = Below Average, 3 = Average, 4 = Above Average, 5 = Far Above Average.

Discussions with the assessors in the pilot also revealed that the group's level of performance on the experimental and control videos looked poor compared to the group who performed in the practice excerpts. Because of this perception, a new practice excerpt was created to replace the low practice excerpt. In the previous low performance, practice excerpt, the practice target's performance was low, but the practice, non-targets' performances were high resulting in relatively high productivity for the group as a whole. In the new, low, practice excerpt, all of the practice assessees performed at a low level, and the resulting level of productivity for the group was low. The intent in creating the tape was to lower assessors' lower boundary for group performance. In this way, the group's performance in the experimental and control videos would not be so clearly perceived aslow compared with the group's performance in the practice excerpts.

The new group performance items and training excerpt were used in the second pilot. Alpha for the new scale was .83 (n=13) and the mean differences on the five point scales was in the appropriate direction (High: Mean = 3.11, SD = .73, n = 9; Low: Mean = 2.08, SD = .63, n = 4).

Interdependence among Group Members

Interdependence among group members was also manipulated during the assessor training. The materials for the case studies included a page with three different facts about the case that were marked A, B, or C. Assessors in the high interdependence condition were informed that each assessee had been given the A, B, or C fact. Assessors were told that the assessees had to depend on each other to solve the cases accurately since each assessee had a unique piece of information. In

the low interdependence condition, assessors were told that all three pieces of information--A, B, and C--were given to all of the assessees for all of the cases. Assessors were told that any one assessee could solve the problem entirely on his/her own and did not need to share information because each assessee had all relevant information.

A number of items were included in the questionnaire asking assessors about their perceptions of the interdependence among group members. An examination of the means and intercorrelations of these items suggested that the internal consistency of these items was not high. The means on the five point scale for the high and low interdependence condition are depicted in Table 5. The means were in the correct direction for three of the items--item 31, item 32, and item 34. Items 31 and 34 were also highly correlated ($r=.48$, $n=19$) so both of these items were retained in the final scale. Item 32 and an item which had a similar stem, item 35, were not retained since the language of these items included a reference to achievement, and it was desirable to maintain a semantic distinction between the measurement of interdependence and of group performance. The means were not in the correct direction for Items 30 and 33 in which assessors were asked about their perceptions of information distribution and dependency relationships. These items were modified significantly into seven new items asking assessors about their perceptions of dependency relationships and information distribution. The retained and new items appear in Table 6.

Table 5

First Pilot Results Examining the Manipulation of Perceptions of Interdependence between Group Members

Item	High Mean(SD)	Low Mean(SD)
30. The Target had important information that the Other Group Members did not have.	1.85(1.35)	2.27(1.27)
31. The Other Group members had important information that the Target did not have.	3.52(1.70)	2.27(1.49)
32. The Target could have achieved exactly the same or more without the contributions of the Other Group Members.(R)	4.63(.69)	3.09(1.22)
33. The other members of the group were highly dependent on the Target.	1.48(1.01)	2.64(1.12)
34. The Target was highly dependent on the Other Group Members.	4.15(.99)	2.45(1.13)
35. The Other Group Members could have achieved exactly the same or more without the contributions of the Target.(R)	1.78(.89)	3.91(.94)

Note: Response format was a five point scale with 1 = disagree and 5 = agree. Means are listed first with standard deviations in parentheses. High interdependence condition (N=27); Low interdependence condition (L=11). (R) = Reverse scored

Table 6

Interdependence among Group Members: Final Items

25.	Sharing information among group members was important for accurate problem solving because all of the information was not given to all of the group members.
26.	Sharing information among group members was not important for accurate problem solving because all of the information was given to all of the group members. (R)
27.	Because of the distribution of information to members of the group, group members had to depend on each other to solve problems accurately.
28.	Group members did not have to depend on each other at all; Any one of the group members could have solved the problem just as completely and accurately without access to others' information. (R)
29.	The Target was highly dependent on the Other Group Members.
30.	The group members each had sufficient information to solve the case on his/her own. (R)
31.	The Other Group members had important information that the Target did not have.
32.	Group members needed to depend on each other to provide important information about the case.
33.	Since all team members had the same information about the cases, it didn't matter whether the group worked together as a team or not. (R)

Note: Response format was a five point scale with 1 = Disagree, 2 = Disagree Somewhat, 3 = Neither Agree nor Disagree, 4 = Agree Somewhat, 5 = Agree (R) = Reverse scored.

The new interdependence items were employed in the second pilot study. Alpha for the new scale was .76 (n = 13), and the means based on the five point scale were in the predicted direction for the different experimental conditions (High: Mean=3.98; SD=.82 n=7; Low: Mean=3.29 SD=.81; n=6).

Mediating and Dependent Variables

Dimension Ratings

It was explained previously that four dimensions were part of the original conceptualization, and three dimensions--Problem Solving, Managing Interpersonal Relationships, and Initiative/Assertiveness--were used in the focal study. In this section, I explain how assessors were trained to arrive at their ratings.

Assessment center rating procedures range from detailed behavior-oriented note taking to computerized checklist systems and from post-center consensus discussions to post-exercise ratings with no consensus procedure. No one of these practices can be called "the" assessment center method (Moses, 1992). I chose to use a method of recording and categorizing behaviors that is commonly used and has an extensive history. The method of rating without a consensus discussion was chosen since the post exercise rating design is necessary to examine the hypothesized effects in studies of this type (e.g. Sackett & Dreher, 1982). These methods of observing, categorizing, and rating are within the range of common practice in assessment centers.

Specifically, assessors in the study were taught to take notes of the target assessee's behavior. Assessors were urged to record behavioral data and not general impressions or evaluations of the assessees' behavior. Assessors were instructed to categorize the

behaviors according to the three dimensions and then, compare the behaviors categorized according to a dimension with behaviorally anchored scoring guidelines. Scoring guidelines provided anchors at high, typical, and poor levels and thus, assisted assessors in making their evaluations. Finally, assessors decided on an integer rating on a one (low) to seven (high) point scale for each of the three dimensions. Assessors were instructed not to discuss their ratings or observations with each other at any time during the observation, categorization, or evaluation processes. Thus, assessors' ratings were independent of each other and not a product of a consensus discussion. The instruction that assessors received about recording, categorizing, and rating behavior appears in the assessor training materials in Appendix D. The behavioral anchored scoring guidelines can be found in Appendix B.

Impression of Leadership

Assessors were asked a series of items regarding their perceptions of the target in fulfilling his role as the group leader. This scale was based on Phillips and Lord (1981) who found that a single item measure of general impression of leadership mediated the relationship between performance cue information and trends in ratings. The results of these questions from the first pilot appear in Table 7. The intercorrelations of these items were quite high. Alpha for the scale was .95. The five items were retained in the final version of the questionnaire.

Attributions of Responsibility for Group Performance

Assessors were asked a series of questions regarding whether they attributed responsibility for the group's performance to the target or

Table 7

Pilot Results Examining Impression of Leadership Scale Items

Item	Mean(SD)	r_{it}
45. I considered the Target an excellent leader.	1.45(.86)	.91
46. The Target exhibited a great deal of leadership.	1.50(.80)	.90
47. The group was led well by the Target.	1.87(1.07)	.85
48. I considered the Target a poor leader. (R)	1.68(1.04)	.94
49. The group was led poorly by the Target. (R)	1.73(1.24)	.81

Note: Response format was a five point scale with 1 = disagree and 5 = agree. Means are listed first followed by Standard Deviations (in parentheses). r_{it} = Corrected Item-Total Correlations. (R)=Reverse scored.

the other group members. Also as in Phillips and Lord (1981), I intended to determine whether assessors' causal attributions mediated the relationships between independent variables and ratings. The results from these questions appear in Table 8. All of the item intercorrelations and item-total correlations were in an acceptable range, except for items 41 and 42. It was hypothesized that the meaning of the dual stem--success/failure--may not have been interpreted as it was intended. Items 41 and 42 were not retained in the scale. All of the remaining items were retained except for Item 43 which was modified to resemble item 44 by having "others" as the subject of the stem but which asked about assessors' attributions regarding the ability of the target. This distinction of ability and effort in the two stems resulted in more items that more closely resembled the items used by Phillips & Lord (1981) and equated thenumber of items with the target as the subject of the stem and with the others as the subject of the stem. The new scale appears in Table 9. The new scale was examined in the second pilot study. Alpha for the new scale was .87 (n = 13) in the second pilot study.

Dissimilarity of the Target's versus Non-targets' Performance

Participants were asked about the extent to which they viewed the other target as similar to or not similar to the other group members in the four performance dimensions. The items and format for the responses are depicted in Table 10. Participants were first asked to indicated the extent to which they viewed the target as similar to or not similar to the non-target assessees. Then, participants were directed to one of two items in which they indicated how similar or dissimilar the target assessee's performance was to the non-targets'

Table 8

Pilot Results Examining Attributions of Responsibility Scale Items

Item	Mean(SD)	r_{it}
36. This group's accomplishments were clearly due to the ability of the Target	1.84(1.10)	.72
37. This group's accomplishments were clearly due to the ability of the Other Group Members. (R)	2.29(1.27)	.56
38. Compared to the Other Group Members, the Target contributed the most to the accomplishments of the group.	2.00(1.34)	.76
39. The Target was an obstacle to the group's performance. (R)	2.66(1.24)	.56
40. The Other Group Members were an obstacle to the group's performance.	2.50(1.29)	.51
41. The Target was solely responsible for this group's success/failure.	2.45(1.13)	.22
42. The Other Group Members were solely responsible for this group's success/failure. (R)	3.18(1.31)	.20
43. The Target would have accomplished more if the others in the group had put in more effort.	2.29(1.23)	.70
44. The others in the group would have accomplished more if the Target had put in more effort. (R)	2.13(1.26)	.69

Note: Response format was a five point scale with 1 = disagree and 5 = agree. Means are listed first followed by Standard Deviations (in Parentheses). r_{it} = Corrected Item-Total Correlations. (R) = Reverse scored.

Table 9

Attributions of Responsibility: Final Items

34. This group's accomplishments were clearly due to the ability of the Target.
35. This group's accomplishments were clearly due to the ability of the Other Group Members. (R)
36. Compared to the Other Group Members, the Target contributed the most to the accomplishments of the group.
37. Compared to the Other Group Members, the Target was most responsible for the group's success/failure.
38. The Target was an obstacle to the group's performance. (R)
39. The Other Group Members were an obstacle to the group's performance.
40. The others in the group would have accomplished more if the Target had put in more effort. (R)
41. The Others in the group would have accomplished more if the Target had greater ability. (R)

Note: Response format was a five point scale with 1 = Disagree, 2 = Disagree Somewhat, 3 = Neither Agree nor Disagree, 4 = Agree Somewhat, 5 = Agree. (R) = Reverse scored.

Table 10

Format for the Dissimilarity Scale Used in the First Pilot

---

50. How did you view the Target's Problem Solving Skills relative to the Other Group Members?
    a. Similar--go to # 51 b. Not Similar Go to # 52

51. If you marked Similar was the Target's Problem Solving behavior
    a. Similar but Slightly Worse?
    b. Moderately Similar?
    c. Highly Similar?
    d. Similar but Slightly Better?

52. If you marked Not Similar was the Target's Problem Solving behavior
    a. Much Worse
    b. Worse
    c. Better
    d. Much Better

53. How did you view the Target's Empathy relative to the Other Group Members?
    a. Similar--Go to # 54 b. Not Similar--Go to #55

54. If you marked Similar was the Target's Empathy
    (Same response format as #51)

55. If you marked Not Similar was the Target's Empathy behavior
    (Same response format as #52)

56. How did you view the Target's Setting Directions & Goals relative to the Other Group Members?
    a. Similar--Go to # 57 b. Not Similar--Go to # 58

57. If you marked Similar was the Target's Setting Directions & Goals Behavior (Same response format as #51)

58. If you marked Not Similar was the Target's Setting Directions & Goals Behavior (Same response format as #52)

59. How did you view the Target's Motivation relative to the Other Group Members?
    a. Similar--Go to # 60 b. Not Similar--Go to # 61

60. If you marked Similar was the Target's Motivation
    (Same response format as #51)

61. If you marked Not Similar was the Target's Motivation
    (Same response format as #52)

---

performance. First, it should be noted that a number of the participants did not follow the directions and provided responses to every item. A clearer item format was necessary. The intercorrelations between the items also suggested that the responses were not highly consistent with each other, and some items were correlated negatively.

The item format was changed considerably for the final version of these items. The modified format and items are depicted in Table 11. The second pilot study confirmed that the new item format was more clearly understood. The collection of these items did not seem to be unidimensional, however. Although items 46 and 50 correlated .29 (n = 12), both of these items were correlated near zero with item 48. I proceeded to measure this variable as it was originally conceived, though the items may not have represented a unidimensional scale.

Assessment of Common Method Bias

All of the dependent measures were subjective ratings collected from the same individual. As such, an attempt was made to assess the extent to which assessors may have been responding in an indiscriminate manner. A number of items in the questionnaire asked the participants about their level of satisfaction with their current or recent job. Participants answered these items after making their ratings and before responding to the questions regarding the manipulations and psychological constructs. It was thought that responding to this questionnaire would also serve as a distractor between the collection of the different dependent variables. Further, the relationship between the participants' responses to the job satisfaction questions should be uncorrelated with their ratings and questionnaire responses

Table 11

Dissimilarity: Final Items

---

47. How did you view the Target's Problem Solving Skills relative to the Other Group Members?
    a. Similar b. Not Similar

48. If you marked Similar was the Target's Problem Solving behavior
    48a. Similar but Slightly Worse?
    48b. Moderately Similar?
    48c. Highly Similar?
    48d. Similar but Slightly Better?

    If you marked Not Similar was the Target's Problem Solving behavior
    48e. Much Worse
    48f. Worse
    48g. Better
    48h. Much Better

49. How did you view the Target's Managing Interpersonal Relationships behavior relative to the Other Group Members?
    a. Similar b. Not Similar

50. If you marked Similar was the Target's Managing Interpersonal Relationships behavior
    (Same Response Format as #48a-d)

    If you marked Not Similar was the Target's Managing Interpersonal Relationships behavior
    (Same Response Format as #48e-h)

51. How did you view the Target's Initiative/Assertiveness relative to the Other Group Members?
    a. Similar b. Not Similar

52. If you marked Similar was the Target's Initiative/Assertiveness Behavior
    (Same Response Format as #48a-d)

52. If you marked Not Similar was the Target's Initiative/Assertiveness Behavior
    (Same Response Format as #48e-h)

Note: Items in the format of number 48, 50, and 52 were coded so that a = 3, b = 2, c = 1, d = 3, e = 5, f = 4, g = 4, h = 5. High scores indicated the extent to which participants viewed the target's performance as not similar to the performances of the non-targets.

if they were attending to the content of the items. Thus, finding no evidence of a relationship would offer evidence that participants were responding in a discriminating manner.

The results from the first pilot test of the job satisfaction items appear in Table 12. Many of the items demonstrated high internal consistency with alpha for the scale = .85 (n = 37). As such, many of the items were retained in the final version of the questionnaire. Some additional items were added with stems asking about intrinsic motivational elements of their job such as self-esteem and self-actualization. Also, some extreme items were added asking the students the extent to which their job and salary was ideal in hopes that these items would identify students who were responding to the questionnaire indiscriminately. The final version of the job satisfaction items appears in Table 13.

Summary

The pilot studies resulted in some modifications of the measures, materials, and the procedure. A number of new items were written for the questionnaire to improve the measurement properties of the variables. The changes also involved the creation of an additional, low condition for the training tape so that assessors would develop lower boundaries for their expectations of group and individual performance. Second, the Motivation and Setting Directions and Goals dimensions were combined into a single dimension called Initiative/Assertiveness. The title for the Empathy dimension was changed to Managing Interpersonal Relations. The revised items were used in a second pilot study with a small number of subjects, and the results confirmed that the new items performed as predicted and could

Table 12

Pilot Results Examining Job Satisfaction Scale Items

Item	Mean(SD)	r_{it}
5. The amount of pay and fringe benefits I receive	3.40(1.48)	.53
6. The degree of respect I get from my boss	3.95(1.20)	.52
7. The feelings of worth and accomplishment I get from my job	3.62(1.38)	.78
11. The degree to which I am paid fairly in this job	3.65(1.46)	.62
12. The amount of independent thought and action I can exercise in this job	3.38(1.30)	.59
13. The amount of challenge in my job	3.00(1.27)	.49
14. The overall quality of the supervision I receive	3.40(1.36)	.68

Note: Response format was on a five point scale with 1 = Dissatisfied and 5 = Satisfied. Means are listed first followed by Standard Deviations (in Parentheses). n = 38. r_{it} = Corrected Item-Total Correlations.

Table 13

Job Satisfaction: Final Items

4.	Its hard for me to care very much about whether the work gets done right on this job. (R)
5.	My opinion of myself goes up when I do this job well.
6.	I am constantly feel as sense of fulfillment and self-actualization in my work.
7.	Most of the things I do on this job seem very trivial. (R)
8.	I can't imagine having a job that is more satisfying that the one I have now.
9.	This job is my ideal career job. Its all I ever wanted in life.
10.	The degree of respect I get from my boss.
11.	The feelings of worth and accomplishment I get from my job.
12.	The degree to which I am paid an ideal salary in this job.
13.	The amount of independent thought and action I can exercise in this job.
14.	The amount of challenge in my job.
15.	The overall quality of the supervision I receive.

Note: Response format was a five point scale with 1 = Dissatisfied, 2 = Slightly Dissatisfied, 3 = Neutral, 4 = Slightly Satisfied, 5 = Satisfied. (R) = Reverse scored

be used in the focal study. Copies of the complete questionnaires used in the first pilot and in the experiment can be found in Appendix F.

Procedure

Participants were scheduled into two-hour sessions, and the experimental conditions were randomly assigned to the sessions. Sessions were conducted with groups of six to eight assessors with all assessors in the session exposed to the same conditions of the independent variables. Roughly, three sessions were conducted for each cell in order to arrive at the intended sample size. When participants arrived, they were verbally informed of the study's procedure and their rights as participants. Then, they were asked to indicate their consent for participation by signing a consent form which explained the procedure and their rights in writing. Participants were told that they were being trained to rate an individual who was the assigned leader of a team of students working on industrial and organizational psychology cases. Participants were fully debriefed after the study including being informed that the assessees on tape were paid actors.

Once the assessor training portion of the experiment began, participants watched a fifteen minute video about the dimensions and observing, categorizing, rating, and avoiding rater errors. The participants then watched a video introducing them to the case materials including the manipulation of group performance and interdependence among group members. Specifically for group performance, participants were shown the distribution of cases that groups typically finished to create the group performance manipulation as described above. Regarding interdependence, participants were told about the information distribution and about the dependency

relationships that were created from the distribution of information.

Next, the background information and correct answers for three of the seven cases were explained on video, followed by practice observation and rating of the same three cases. Participants were then introduced to the background and correct answers for the remaining four cases. This was followed by a two minute video summarizing key components of the training content and including a brief reminder of the group performance and interdependence information.

Finally, assessors watched one of the experimental or control videos depending on the condition to which they were assigned. Participants responded to the questionnaire immediately after viewing the video. They first rated the target on the three dimensions. Scoring guidelines were taken from assessors at this time so that the remainder of the items such as those measuring group performance and the performance of non-target assessees would be rated from memory and not with access to dimension anchors. Second, they responded to the job satisfaction items. Then, they completed the group performance items, the ratings of the non-targets, the interdependence items, the attribution items, the impression of leadership items, and the contrast and assimilation items. The session ended with a debriefing session during which the purpose of the experiment was explained and the participants were offered the opportunity to ask questions and to discuss their reactions.

Data Analyses

Effects of Group Variables on ratings

The facets in the design included: group composition (A,H,H vs. A,L,L); assessors' perceptions of group performance (high vs. low); and

assessors' perceptions of interdependence of individuals in the group (high vs. low). Mean differences in the ratings for the independent variable conditions were analyzed in a 2 X 2 X 2 analysis of variance (ANOVA), and means for specific cells were compared to test specific hypotheses.

Kravitz and Balzer (1992) have illustrated the importance of using a control group in contrast and assimilation research. The ratings from the control group were treated according to Winer's (1971) recommended procedure for a factorial experiment with a single control group. Much of the analysis followed the usual computation procedures for any factorial experiment. The ratings of the control group were used primarily in calculating error terms. The within cell variation was obtained by pooling the within cell variation from the factorial part of the experiment with the within cell variation from the control group. The between cell variation is based on the sum of ratings from the factorial part of the experiment and the sum of ratings from the control group. Thus, statistical tests of the effects of the independent variables were based on a contrast with a condition where the manipulations were neutral or absent. The control group ratings were only used in the ANOVA procedure.

Assessor Cognitive Process Explanations of the Findings

Analyses were also conducted to determine whether the effects of the independent variables on ratings were explained by the hypothesized cognitive processes. The questionnaire responses to the dissimilarity of target's and non-target's performance, general impression of leadership, and or attribution of responsibility scales were tested as partial mediators of the effects of the independent variables on

ratings according to the respective hypotheses. Series of hierarchical regression analyses were used for this test according to the procedure recommended by Alwin and Hauser (1975).

The procedure for conducting these analyses is described generally here and in more detail later when each hypothesis is examined in the results section. Dummy coded variables were used to represent the effects of the independent variables. For each hypothesis, the overall dimension rating was regressed on the dummy coded variable representing the experimental manipulation in the first step, followed by the scale score for the mediating psychological variable in the second step. This sequence allowed for the estimate of the change in R^2 with the addition of the mediating variable in step two. The steps were reversed to determine whether the experimental condition significantly accounted for variance in the ratings when entered after the mediating variable. A significant change in R^2 at step two in the first equation but no significant change in R^2 at step two of the second equation was considered evidence of a full mediating relationship. Finally, path models for the same set of variables were constructed based on the procedure recommended by Cohen and Cohen (1977) to facilitate the understanding of the combined effects of the variables in each hypothesis.

RESULTS

Independent Variables and Manipulation Check

Group Composition

Group composition was manipulated by scripting the performance of non-target actors/assessees in three different videotapes. As a manipulation check, participants were asked to rate the non-target assessees from memory. Ratings of the non-target male and the non-target female are depicted in Table 14. A one-way analysis of variance was used to verify that the differences in the means between the high and low composition conditions were significant. A single index created by averaging the ratings of all three dimension ratings for both the male and the female was the dependent variable. Alpha for the index based on six items was .76 (n = 145). Mean differences on the seven point scale between the high and low conditions, as shown in Table 14, were in the predicted direction and were statistically significant (F (1,145) = 76.9, p < .001).

The statistics suggested that most of the mean differences were in the predicted direction, but further examination of these ratings reveals that the ratings of the non-targets did not exactly conform to intended levels. In the low composition (A,L,L) condition, ratings of the male and female non-targets were all below four, the midpoint of the seven point rating scale. As predicted in the control group, ratings of the female's high performance were above four and ratings of

Table 14

Manipulation Check for Group Composition: Assessors ratings of Non-Target Assessees

Male	ALL (N = 76)	AHH (N = 70)	AHL (N = 21)
Problem Solving	3.24(1.60)	4.16(.81)	2.48(1.12)
Managing Interpersonal Relationships	2.11(1.10)	4.51(1.23)	1.76(1.22)
Initiative /Assertiveness	3.33(1.57)	3.32(1.36)	2.57(1.16)
Female			
Problem Solving	3.04(1.63)	4.84(1.11)	5.00(1.09)
Managing Interpersonal Relationships	1.91(1.27)	3.25(1.52)	4.90(1.18)
Initiative /Assertiveness	3.81(1.88)	4.87(1.32)	5.43(.93)
Composite	2.91(.91)	4.16(.81)	3.11(1.05)

Note: ALL = Target with two low performers; AHH = Target with two high performers; AHL = Control Condition--Target with one high (Female) and one low (Male) performer.

the male's low performance were below four. In the high composition condition (A,H,H), ratings of the male and female were abovefour, for the most part. The exceptions were the rating of the male for Initiative/Assertiveness which was not higher than the same rating in the low condition and the rating of the female of Managing Interpersonal Relationships which was not above four but which was higher than the same rating in the low composition condition.

Although alpha was .76 (n = 145), factor analyses (presented later) demonstrated that the ratings of the non-target male and female shared some variance (intercorrelations ranged from .28 to .35) but represented two factors. This can been seen in the mean ratings in that the rank ordering of the non-target male and female was not consistent across conditions. In the high condition, for example, the male non-target's performance was a stronger manipulation (i.e. rated higher) of Managing Interpersonal Relationships whereas the female was a stronger manipulation of Problem Solving and Initiative/Assertiveness.

These results indicated that the manipulations did not go completely as intended. It was thought that assessors would not make fine discriminations in rating the dimensions for the non-targets and would rate both as either wholly above average or wholly below average. Although this did not occur, the ratings were generally consistent with the levels of performance of the non-targets. For example, the female was scripted in the high condition to take the lead of the group in the absence of leadership on the part of the target, so it is not surprising that the female was rated higher than all assessees on Initiative/Assertiveness. In contrast, the female in the high

condition was scripted to subtly attack other assessees with some sarcastic remarks, and her lower rating relative to the non-target male on Managing Interpersonal Relations may have reflected these differences in performance.

Despite these discrepancies from expectations, the manipulation of group composition worked, overall. The non-targets were perceived as higher performers in the high condition than in the low condition. The impact of the issues discussed in this section on the results are examined in the discussion section.

Perceptions of Group Performance

Assessors' responses to the questionnaire items regarding their perceptions of group performance were examined for reliability and for the predicted mean differences. Higher ratings on the items indicated perceptions of higher group performance. Alpha for the group performance items was .86 (n = 146). Means calculated across all three items were entered into a one-way analysis of variance. Mean differences on this five point scale were significant (F (1,145) = 67.9, $p < .001$) and were in the predicted direction (high: mean = .34, sd = .86 n = 70; low: mean = 2.27, sd = .72, n = 76).

Perceptions of Interdependence

Assessors' responses to the questionnaire items regarding their perceptions of interdependence were also examined for reliability and for the predicted mean differences. Higher scores suggested that the assessors perceived the group members as highly interdependent. Alpha for the items in the scale was .90 (n=146). Average scores were computed for the eight item scale and were entered in a one way analysis of variance to determine whether assessors' perceptions were

consistent with the experimental conditions. Mean differences on the five point scale were significant ($F(1,145) = 102.2$, $p < .001$) and were in the predicted direction (high: mean = 4.11, sd = .77, n = 74; low: mean = 2.68, sd = .94, n = 72).

Summary of Independent Variables

In sum, the results presented to this point suggested that the manipulations of the independent variables in this study were successful. Assessors' ratings suggested that they were cognizant of the non-targets' performance, and their impressions of the non-targets' performance were generally consistent with the high and low composition manipulations. Assessors' responses to questionnaire items suggested that their perceptions of the performance achievements of the group and their perceptions of the interdependence among group members were also consistent with the experimental manipulations.

Further summary regarding the independent variables and manipulation checks is provided in Table 15. Intercorrelations in the table confirm that assessors were cognizant of the experimental manipulations. Each experimental condition was significantly correlated with the corresponding variable as measured by the questionnaire. Interdependence as measured by the questionnaire was statistically independent of participant's perceptions of group performance and of the performance of the non-target group members. There was relatively low ($r = .24$) but significant correlation between perceptions of group performance and group composition. Group performance was rated higher when the non-target performers were rated higher and lower when the non-target performers were rated lower. Perhaps participants' perceptions of group performance was partly

Table 15

Means, Standard Deviations, and Intercorrelations for Independent Variables and Manipulation Check Variables

	Mean(SD)	1	2	3	4	5	6
Experimental Manipulations							
1. Group Performance	1.49(.50)						
2. Interdependence	1.51(.50)	-.01					
3. Group Composition	1.52(.50)	-.01	-.01				
Questionnaire Responses							
4. Group Performance	2.78(.95)	.56[b]	.05	-.01	(.86)		
5. Interdependence	3.40(1.12)	.12	.65[b]	.15	.14	(.90)	
6. Group Composition	3.56(1.06)	-.01	-.08	.59[b]	.24[b]	.08	(.76)

Note: [b] = Significant at $p < .01$. SD = standard deviation. Reliabilities for perceptual measures are on the diagonal and in parentheses. Underlined values represent manipulation checks. Number of participants = 146 for all values.

shaped by how they viewed the performance of members of the group. These perceptions may not have been solely rooted in behavioral data, however, since the correlation between the perceptions of group performance and the experimental, behavior-based manipulation of non-targets' performance (i.e. group composition) was not significant.

Measurement of Mediating and Dependent Variables

Dimension Ratings

In many assessment centers (Bycio, et al. 1987; Sackett & Dreher, 1982), the ratings of different dimensions tend to be highly intercorrelated. The intercorrelation between the dimension ratings in this study was also high, ranging from .67 to .69. In fact, high dimension intercorrelations were expected since the behavior of the target was manipulated at a uniform, average level for all three dimensions. Because of the high intercorrelations, a mean rating based on the three different dimension ratings was employed as a single composite dimension rating in all of the analyses that follow. Alpha for the three item composite was .86 (n = 146).

Attributions of Responsibility

Assessors' responses to the questionnaire items were examined for internal consistency. The items were scored so that higher scores indicated the extent to which assessors attributed responsibility for the group's performance to the target. Lower scores indicated the extent to which assessors attributed responsibility for group performance to other group members. Alpha for the eight-item scale was .71.

Impressions of the Target's Leadership

Responses to the questionnaire items asking assessors about their impressions of the target's performance were also examined for internal consistency. Item intercorrelations were high, and alpha for the five item scale was .93. High scores on the scale suggested that assessors perceived the target to be above average and low scores suggested that assessors perceived the target to be below average in fulfilling his role as a leader.

Dissimilarity of the Target and Non-Target Assessees

Responses to the questionnaire items asking assessors about the extent to which they perceived the target as similar to or different from the non-target assessees on the three performance dimensions were also examined for reliability. The responses to the items were coded so that higher ratings indicated that the assessor viewed the target's performance as dissimilar to the performances of the non-target (i.e. a contrast). Lower ratings indicated that the assessor viewed the target's performance as similar to others (i.e. an assimilation). Intercorrelations among the items ranged from .25 to .31, so alpha for the three item scale was moderately low at .53. The low intercorrelations and alpha here was not surprising for a number of reasons. First, assessors did not perceive the performance of the two non-target assessees as completely consistent, so it is possible that asking assessors to rate target's performance versus the "other non-targets" was not a clearly executable task. Further, assessees' ratings also reflected differences in the way the target and non-target assessees were ranked on the different dimensions. This tendency would have caused low reliability for this scale, even though assessors'

rankings may have been consistent with the levels of performance of the assessees on the different dimensions. Acknowledging its low reliability, this scale was maintained and examined as hypothesized, and the analyses involving the scale were conducted as planned.

Job Satisfaction Scale

A number of questionnaire items asked assessors about their feelings of satisfaction toward their current or most recent job. These items were included to assess the extent to which common method bias may have affected the results. One would not expect assessors' ratings of the elements of their job to be significantly associated with their responses to items measuring their perceptions about various aspects of the performance they observed. As a result, evidence suggesting no significant relationship with the other variables measured in the questionnaire was considered as evidence that assessors responded to the questionnaire items in a discriminating fashion.

Alpha for the twelve item scale was .79. Items 8, 9, and 12 were intended to be extreme items regarding job satisfaction, and somewhat low intercorrelations (ranging from .05 to .30) of these items with the other items of the scale likely reduced alpha. These items asked participants the extent to which they found their job ideal in terms of salary and their career aspirations. Low means were expected on these items since it was thought that undergraduates would not be in jobs that they considered ideal. The low means for all three of these items, particularly the means below 2.0 for items 8 and 9, supported the position that assessors did not respond to the questionnaire items haphazardly.

Summary of Mediating and Dependent Variables

In summary, Table 16 depicts the means, standard deviations, intercorrelations for the mediating and dependent variables in the study. As expected, the job satisfaction scale was uncorrelated with the responses to the items regarding the performance that assessors observed. This is offered as evidence that participants considered the content when responding to the questionnaire items. Attributions of responsibility, impressions of the target assessee in his role as leader, and the composite rating were all significantly related. As participants attributed greater responsibility for the group's performance to the target, they rated him higher in terms of their impression of him in the role of leader of the group. Higher ratings were also associated with higher attributions of responsibility and impressions of the target as a leader.

Summary of All Measures

Further summary regarding the measurement of the independent, mediating, and dependent variables is provided here. First, I present factor analysis results that allow for an examination of the convergent and discriminant validity of the different variables. Second, I present the correlations among the variables for additional examination of the convergent and discriminant validity of the variables.

Factor Analyses

A principal components factor analysis with an orthogonal (varimax) rotation was conducted with all of the questionnaire item responses. The number of factors was fixed initially at eight to represent the following eight factors--ratings of the non-target female, ratings of the non-target male, group performance,

Table 16

Means, Standard Deviations, and Intercorrelations for Mediating and Dependent Variables

Variable	Mean(SD)	1	2	3	4	5
1. Composite Rating	3.09(1.27)	(.86)				
2. Dissimilarity	3.43(.78)	-.16	(.53)			
3. Impression of Leadership	1.89(.96)	.72[b]	-.01	(.93)		
4. Attributions of Responsibility	2.55(.70)	.51[b]	-.14	.59[b]	(.71)	
5. Job Satisfaction	3.16(.64)	-.01	.06	-.06	.09	(.79)

Note: [b] = Significant at $p < .01$. SD = standard deviation. Reliabilities are on the diagonal and in parentheses. Number of participants = 146 for all values.

interdependence among group members, attributions of responsibility for group performance, impression of leadership, dissimilarity, and job satisfaction. Attributions of responsibility and impression of leadership were represented by one factor in this analysis, as would be expected based on the high correlation between these scale scores presented previously. The same analysis was repeated with the number of factors fixed at seven. The results are depicted in Table 17.

The factor analysis generally supported the discriminant validity of the constructs measured in the questionnaire. The chief differences between the hypothesized factor structure and the findings were that the non-target male's and female's ratings represented two factors and that attributions of responsibility and impressions of leadership items loaded on a single factor. As explained previously, the two factors representing the non-targets' ratings departed from expectations but were consistent with their performance. The single factor representing the attributions of responsibility for group performance and impressions of leadership scales can be explained on theoretical grounds. As the review of the literature on leadership has shown, observers tend to attribute leadership to persons whom they perceive to be a cause of group performance. There were a few items for which the highest loading was not on the hypothesized factor. These items are discussed in the paragraphs that follow, but the discrepant findings were not seen as significant enough to justify changes in the way the scales were constructed.

Item 39 was one of the items that did not load on the attributions factor as predicted. The item "the other group memberswere an obstacle to the group's performance" does not

Table 17

Factor Analysis of Questionnaire Items

Variable	Factor 1	2	3	4	5	6	7
Attributions							
AT34	.46	- .12	.22	.03	-.18	-.23	.04
AT35	.36	-.08	.11	-.11	-.57	-.07	-.01
AT36	.61	-.05	.12	-.05	-.39	-.22	.04
AT37	.18	-.08	-.05	-.11	-.11	-.04	.43
AT38	.55	-.09	-.05	-.10	-.16	-.08	-.33
AT39	.05	.06	.07	-.69	.06	-.14	-.05
AT40	.63	.08	-.04	.01	.01	-.12	-.00
AT41	.69	.17	.00	.03	.01	-.04	.06
Leadership							
LD42	.83	.09	-.04	.11	.02	-.07	.08
LD43	.85	.09	-.02	.03	.01	.06	.05
LD44	.83	-.04	-.05	.12	-.05	.13	.00
LD45	.88	-.04	-.05	.04	-.01	.14	-.03
LD46	.86	-.03	-.04	.05	.03	.10	-.01
Interdependence							
IN25	-.02	.80	.15	-.02	.11	-.01	-.01
IN26	-.02	.70	.23	-.11	.03	-.14	.03
IN27	.01	.85	.06	.05	.13	-.05	.02
IN28	.06	.83	.04	.12	-.14	.06	.06
IN29	-.32	.32	-.12	.10	.28	.14	.05
IN30	.04	.79	.02	.09	-.10	-.03	.04
IN31	.04	.77	.04	-.02	.11	-.05	.12
IN32	.03	.82	.05	.12	.08	-.02	-.03
IN33	.04	.65	.02	-.03	-.01	.18	.00

Table 17 (cont'd)

Variable	Factor 1	2	3	4	5	6	7
J Satisfaction							
S4	.01	.12	.55	-.15	-.07	.13	.08
S5	-.06	.07	.58	-.22	-.09	-.04	-.15
S6	.01	.16	.74	-.06	-.04	.07	-.05
S7	-.01	.06	.54	.11	-.01	.10	.12
S8	.01	.16	.40	-.12	.28	-.16	.15
S9	-.04	.06	.34	-.08	.26	-.35	.32
S10	.05	.09	.61	.10	-.02	-.13	-.03
S11	.05	-.03	.74	-.09	-.11	-.05	-.05
S12	.15	-.30	.27	.15	-.16	-.03	.31
S13	-.09	.06	.61	-.14	.09	-.03	.19
S14	-.16	-.06	.69	-.05	.02	-.11	.03
S15	.07	.01	.53	.04	.00	-.22	-.24
Gr Performance							
G16	.06	.11	-.10	.87	.00	.04	.04
G17	.08	.08	-.18	.80	.22	.10	-.09
G18	.10	.05	-.09	.89	.11	.04	.01
Female Non-Target							
PSF	-.01	-.01	.00	.12	.76	.12	-.01
MIPF	.06	.07	.00	.06	.77	.00	.05
IAF	-.11	.05	.05	.09	.58	.24	.03
Male Non-Target							
PSM	-.04	.04	-.02	.06	.24	.82	-.04
MIPM	-.05	.03	-.07	.11	.32	.63	.09
IAM	-.01	-.12	-.07	.01	.13	.81	.05
Dissimilarity							
CA48	-.17	.17	.03	-.01	.03	.04	.63
CA50	.19	.13	.04	.10	.12	.10	.65
CA52	-.17	.13	.04	-.02	.10	-.03	.54

Note: Underlined values represent the loading of the highest absolute magnitude.

specifically mention the target assessee and thus, may not co-vary with the other items that state attributions more directly in terms of the target's behavior. Item 39 was negatively associated with the group performance factor (factor 4) suggesting that there may have been an association between perceptions that other group members were not obstacles to performance and group performance. The stem for item 37 directed the participants to compare the target with the other group members, so it was not surprising that the item also loaded on the dissimilarity scale (factor 7). Perhaps item 38 for which the stem was "the target was an obstacle to the group's performance also stimulated a contrast between the target and other group members, thus creating an association with the items on the dissimilarity scale (factor 7). The reverse scored item 35 was negatively associated with the ratings of the non-target female. It is not surprising that lower attributions of responsibility to the target (i.e. higher attributions of responsibility to the non-targets) would be associated with higher ratings of the non-target female.

The loadings for interdependence item 29--"the target was highly dependent on other group members"--were distributed across many factors. The loading on attributions/leadership factor (factor 1) was as high as the loading on the interdependence factor (factor 2). The attribution items created an oppositional relationship between the target and non-targets (high attribution scores meant that assessors attributed responsibility for group performance to the target and low scores meant that group performance was attributed to the non-targets). Thus, as assessors saw the target and non-targets as more dependent, targets were seen as poorer leaders and were attributed less

responsibility for group performance. Additionally, the positive loading on factor four suggests that group performance was perceived as higher under the same circumstances of high dependence.

The job satisfaction item 9, "This is my ideal career job. It is all I ever wanted in life," and item 12, "the degree to which I am paid an ideal salary in this job," loaded on multiple factors. It is understandable that these items did not load on factor 3 in a consistent way with other job satisfaction items because they inquired about more extreme (ideal) circumstances.

Though the items on the dissimilarity scale loaded highest on the factor as predicted, there was evidence of an association between these items and the attributions/leadership factor (factor 1). There was a negative association between the dissimilarity scale items for Problem Solving and Initiative/Assertiveness and factor 1. This suggested that more leadership and attributions of responsibility for group performance were associated with the target when he was perceived as more similar to the non-targets on problem solving and Initiative/Assertiveness. The opposite was true for Managing Interpersonal Relationships; more leadership and attributions of responsibility were associated with the target when his performance was seen as dissimilar from the non-targets' performance. The latter finding may not be so surprising given that there was a sharp contrast between the target's and non-targets' performance on Managing Interpersonal Relationships, particularly in the low composition condition in which the non-targets engaged extensively in conflict. The contrast between the target's and non-targets' performances in this condition may have boosted the assessors' views of the target.

In sum, the factor analyses provided evidence of support for the convergent and discriminant validity of the different variables measured by the items in the questionnaire. No changes were made in the way the scales were constructed. The few moderate discrepancies between the hypothesized factors and the results were explained for the most part on theoretical or logical grounds. Minor discrepancies were likely due to sampling error.

Correlations between Independent, Mediating, and Dependent Variables

Correlations between independent, dependent, and mediating variables are depicted in Table 18. This table, combined with Tables 15 and 16, provides an examination of the relationships between all of the measured variables. Some interesting findings emerged from these results. First, attributions of responsibility were negatively associated with the experimental manipulation and the questionnaire measure of group composition. Consistent with a component of hypothesis 3A, more responsibility for the group performance was attributed to the target assessee as the non-targets' performance was manipulated or perceived at low levels. Second, though no relationship existed between group performance and group composition manipulations and the composite performance rating, there was a significant but relatively small association between the questionnaire measures of these variables and the composite rating. The significant, positive correlation with group performance suggested that the target assessee was rated higher when the groups' performance was perceived higher as predicted in hypothesis 2. The significant, positive correlation with group composition suggested an assimilation effect, contrary to the prediction in hypothesis 1. In other words, the target assessee was

Table 18

Correlations between Independent, Mediating, and Dependent Variables

	Mediating and Dependent Variables				
	Comp. Rating	Dissim.	Lead.	Attr.	Jsatsf.
Experimental Conditions					
Group Performance	-.02	.18[a]	.02	-.06	.00
Interdependence	-.02	.01	.06	.04	.02
Group Composition	-.01	.07	-.16	-.50[b]	-.08
Questionnaire Responses					
Group Performance	.18[a]	.04	.14	-.07	-.20[b]
Interdependence	-.06	.16[a]	-.01	-.06	.11
Group Composition	.16[a]	.15	-.05	-.42[b]	-.11

Note: [a] = Significant at $p < .05$. [b] = Significant at $p < .01$. Number of participants = 146 for all values. Comp. = Composite, Dissim. = Dissimilarity of Target's vs. Non-targets' performance, Lead. = Impression of the Target in the role of leader, Attr. = Attributions of Responsibility for group performance, Jsatsf. = Job satisfaction.

rated higher in association with the perceived higher performance of the non-target assessees. Finally, there were other significant but relatively small correlations between dissimilarity and the manipulation of group performance and interdependence that were not readily interpretable.

Tests of Hypotheses

The following paragraphs report the results directly relevant to the hypotheses presented in the introduction. Two (group composition) by two (perceptions of interdependence) by two (perceptions of group performance) analyses of variance were conducted using the composite rating as the dependent variable. One set of analyses was run without consideration of the control group and a second set of analyses was conducted using Winer's formulations for factorial analyses of variance with an appended control group. The summary tables from these analyses are depicted in Table 19. Cell means, standard deviations, and cell sizes are depicted for the composite rating in Table 20. None of these analyses revealed any significant main effects or interactions, and thus, a number of the hypotheses were not supported. I proceed in the following paragraphs with an accounting of the findings relative to each hypothesis. Even in the case where hypotheses were not supported, non-significant findings will be reported, and their implications will be discussed.

Hypothesis 1

In Hypothesis 1, it was predicted that ratings of the target's performance would be contrasted with the performance levels of the non-target assessees in the group. Mean differences between the high and low composition groups were not significant nor were there any

Table 19

Analysis of Variance Summary Table for Composite Rating with and without the Control Group

Variable	Without Control Group			With Control Group		
	df	MS	F	df	MS	F
Group Composition (C)	1	.04	.06	2	.00[a]	.00
Interdependence (I)	1	.11	.06	2	.00[a]	.00
Group Performance (G)	1	.10	.05	2	1.07	.65
C X G	1	.43	.26	4	1.85	1.13
C X I	1	2.13	1.28	4	2.49	1.52
I X G	1	1.54	.92	4	1.01	.61
C X I X G	1	1.46	.88	8	2.47	1.50
Within Cell	145	1.62		157	1.64	

Note: df = Degrees of Freedom, MS = Mean Square. [a] Hand calculated values were near zero and sometimes negative depending on the number of significant digits and the rounding strategy.

Table 20

Cell Means, Standard Deviations, and Sample Sizes for the Composite Rating

Low Group Composition (A,L,L)

Group Performance	Interdependence: Low	Interdependence: High
Low	3.08(1.04) n = 17	3.28(1.61) n = 19
High	2.92(1.44) n = 17	3.12(1.40) n = 17

High Group Composition (A,H,H)

Group Performance	Interdependence: Low	Interdependence: High
Low	3.38(1.11) n = 20	2.72(.89) n = 20
High	3.04(.78) n = 18	3.17(1.76) n = 18

Note: n = sample size in a cell.

significant effects involving group composition. The means and standard deviations for the group composition main effect conditions appear in the top row in Table 21. Trends in the composite rating across the two conditions were in the predicted direction of a contrast effect but such extremely small differences were not practically meaningful.

Hypothesis 1A

In Hypothesis 1A, it was predicted that perceptions of dissimilarity of the target's and non-targets' performance would be a partial mediator of the relationship between group composition and ratings. This analysis was conducted using a series of hierarchical regression analyses (Alwin & Hauser, 1975; James & Brett, 1984). Table 22 shows that dissimilarity explained two percent of the variance in the composite beyond dummy coded variable representing the group composition manipulation. When the variables were entered in the reverse order, group composition did not explain variance in the composite beyond dissimilarity. Though these findings suggested that perceptions of dissimilarity may mediate the relationship between the group performance manipulation and ratings, the change in R^2 at step two in the first analysis was not significant nor was there a link between composition and dissimilarity. Thus, there was no relationship between composition and ratings to be explained by dissimilarity.

Hypothesis 2

In hypothesis 2, it was predicted that ratings of the target would be associated with the experimental manipulation of assessors' knowledge of group performance achievements. The main effect for group performance and all interaction effects including group performance

Table 21

Means and Standard Deviations of the Composite Rating for Main Effects

Effect	High Mean(SD)	Low Mean(SD)	Control Mean(SD)
Group Composition	3.07(1.19) n=76	3.10(1.37) n=70	3.11(1.28) n=21
Interdependence	3.06(1.43) n=74	3.12(1.10) n=72	3.11(1.05) n=21
Group Performance	3.06(1.36) n=70	3.11(1.20) n=76	3.11(1.05) n=21

Note: SD = Standard Deviation, n = sample size.

Table 22

Hierarchical Regression Analyses for Hypothesis 1A: A Test of Dissimilarity as a Partial Mediator of the Relationship between Group Composition and Ratings

Step	Variable Entered	R^2	ΔR^2	F
1	Group Composition	.00		.02
2	Dissimilarity	.02	.02	3.57

Step	Variable Entered	R^2	ΔR^2	F
1	Dissimilarity	.02		3.16
2	Group Composition	.02	.00	.00

were not significant. Table 21 shows that there was a slight difference between the mean composite ratings in the two group performance conditions. The practically insignificant difference was not in predicted direction.

Hypothesis 2A

In Hypothesis 2A, it was predicted that a general impression of the target in the leadership role would partially mediate the relationship between the contextual information about group performance provided to assessors and the ratings. The data in Table 23 showed that general impression of leadership added significantly to the prediction of the composite beyond the dummy coded variable representing group performance, and that group performance added no explanatory power of the composite beyond impression of leadership. Overall, these results combined with the zero-order correlations in Table 18 suggested that there was a strong association between assessors' general impression of the assessee's performance in a leadership role and the composite rating of the target. The lack of an association between the group performance manipulation and the general impression of leadership suggested that the group performance manipulation did not contribute to the formation of the general impression of leadership. Therefore, no mediating relationship existed as predicted.

Hypothesis 3

In hypothesis 3, it was predicted that group performance and group composition (i.e. the pattern of individual performances in a group) would interact in predicting assessors' ratings. This hypothesis was not supported. The group performance, group composition

Table 23

Hierarchical Regression Analyses for Hypothesis 2A: A Test of Impression of Leadership as a Partial Mediator of the Relationship between Group Performance and Ratings.

Step	Variable Entered	R^2	ΔR^2	F
1	Group Performance	.00		.06
2	Impression of Leadership	.52	.52	154.23[c]

Step	Variable Entered	R^2	ΔR^2	F
1	Impression of Leadership	.52		154.35[c]
2	Group Performance	.52	.00	.51

Note: [c] significant at $p < .001$

interaction term was not significant in the analyses of variance. Figure 8, showing the mean ratings for the four conditions, suggested that slight differences occurred across these conditions but that the trends in the means were not in the predicted direction. The target assessee was rated highest among these four conditions when non-targets' performance was low and in the low group performance condition. The largest difference occurred within the low composition condition (A,L,L). The counter-intuitive finding was that the target was rated higher in the low group performance condition than in high group performance condition. Perhaps assessors' higher ratings of students resulted from some empathy with their peer who was assigned to be the leader of a homework exercise with two conflicting, poor performing students.

Hypothesis 3A

In hypothesis 3A, it was predicted that information given to assessors about group performance and group composition as manipulated by the performance of non-target assessees would influence the attributions of responsibility for group performance that assessors make to the target vs. the non-target assessees. These attributions, in turn, would influence the ratings of the target. The results of the regression analysis testing the partial mediating model appear in Table 24. The results here suggested that attributions added significantly (34%) to the prediction of the composite beyond the dummy coded variables representing group performance and group composition. When the variables were entered in the reverse order, attributions accounted for 26 percent of the variance alone with group performance and group composition predicting a small but significant 8 percent of variance in

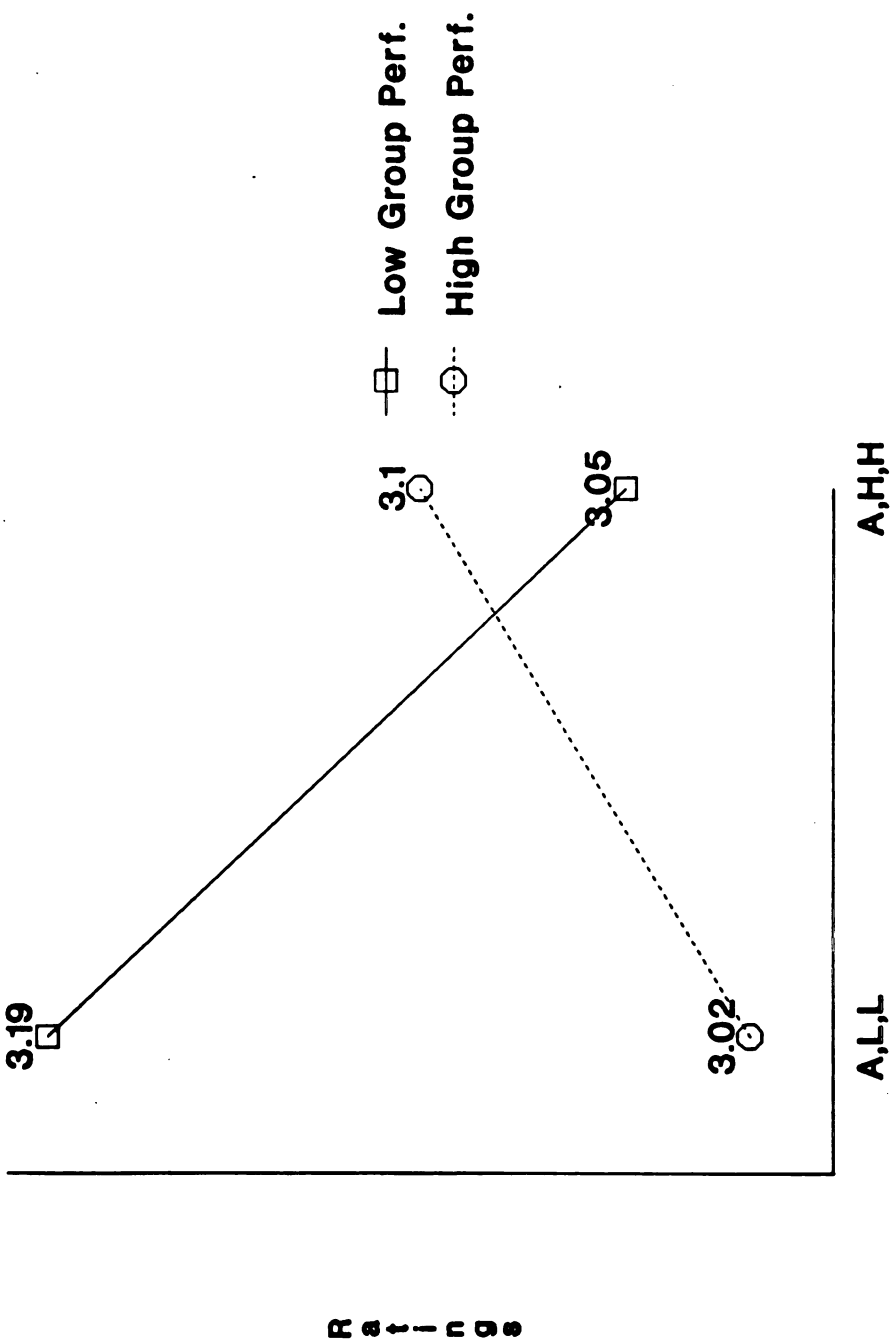

Figure 8
Hypothesis 3 Results: Mean Differences in Ratings for the Group Composition by Group Performance Interaction

Table 24

Hierarchical Regression Analyses for Hypothesis 3A: A Test of Attributions of Responsibility as a Mediator of the Relationship between Group Performance and Group Composition and Ratings

Step	Variable Entered	R^2	ΔR^2	F
1	Group Performance	.00		.04
	Group Composition			
2	Attributions	.34	.34	71.62[c]

Step	Variable Entered	R^2	ΔR^2	F
1	Attributions	.26		50.18[c]
2	Group Performance	.34	.08	8.62[c]
	Group Composition			

Note: [c] significant at $p < .001$

the composite beyond attributions.

These results offer partial support for hypothesis 3a. Figure 9, depicting the relationships between these variables in the form of a path model, facilitates an understanding of their combined influence on the composite. Although there was no support for the predicted relationships involving group performance, group composition was negatively associated with attributions of responsibility for group performance to the target, and attributions were positively associated with ratings. The negative relationship between group composition and attributions of responsibility suggested that higher attributions of responsibility for group performance were associated with the target (vs the non-targets) when the non-targets' performance was low. Lower attributions of responsibility were associated with the target (also implying higher attributions to the non-targets) when the non-targets' performance was high. Additionally, higher attributions of responsibility for group performance to the target were associated with higher ratings whereas as lower attributions of group performance to the target (i.e. higher attributions to non-targets) were associated with lower ratings.

The analyses in Table 24 also suggested that group composition and group performance manipulations contributed slightly but significantly to the prediction of the composite beyond attributions of responsibility. Figure 9 shows that group composition had a significant direct effect on the composite rating when entered at the same time in the regression equation as attributions of responsibility. The size of this direct effect at .33 represented a significant change from the zero-order correlation of group composition and ratings of

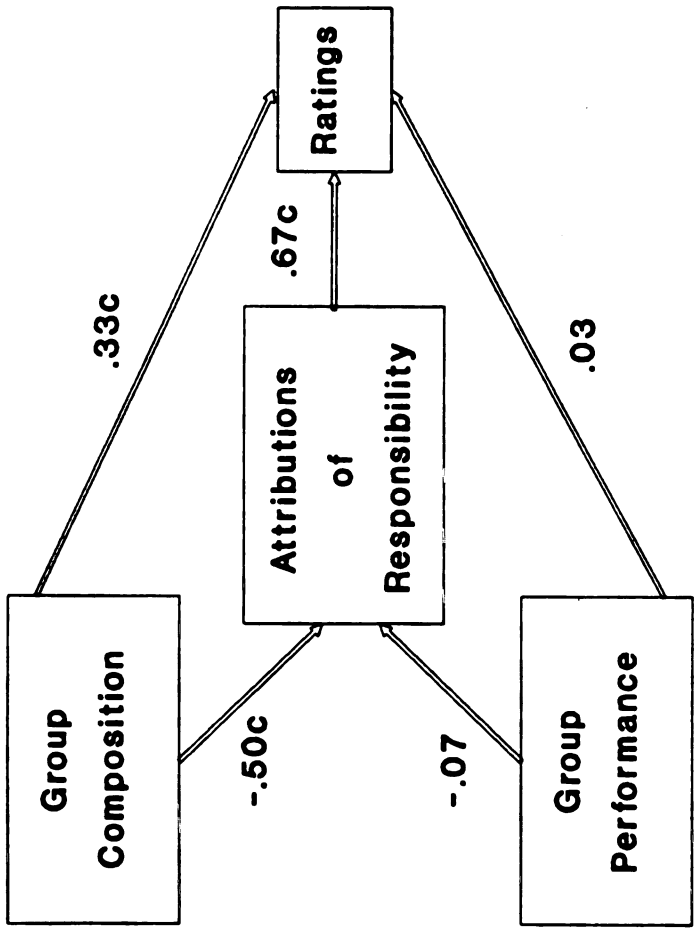

Figure 9
Hypothesis 3A Results: Attributions as a Partial Mediator of Group Performance and Group Composition on Ratings

Note: c = significant at $p < .001$

-.02. The relationship between attributions and relationship between attributions and ratings increased.ratings increased to .67 in the path model from a zero-order correlation of .51. Thus, group composition acted as a suppressor variable. When the portion of variance in attributions associated with group composition (i.e. the ratings of the non-target assessees), the

Hypothesis 4

In hypothesis 4, it was predicted that group composition and interdependence would interact in predicting assessors' ratings. Specifically, a contrast effect was predicted when the interdependence among group members was perceived as low. An assimilation effect was predicted when assessors perceived that the interdependence among group members was high, because assessors would have more difficulty sorting out the contributions of the target from the other group members upon whom the target was dependent. The group composition by interdependence interaction was not significant in the analyses of variance. The results in Figure 10, which shows the means for the four conditions, suggested that the non-significant mean differences were also not in the predicted direction. The rating of the target was assimilated to the level of the non-targets' performance in the low interdependence condition but contrasted with the non-targets' performance in the high interdependence condition.

Summary

Some conclusions can be drawn from all of the results presented in this section. First, there was very little evidence of support for the direct effects of group composition, group performance, and interdependence, as they were experimentally manipulated in this study.

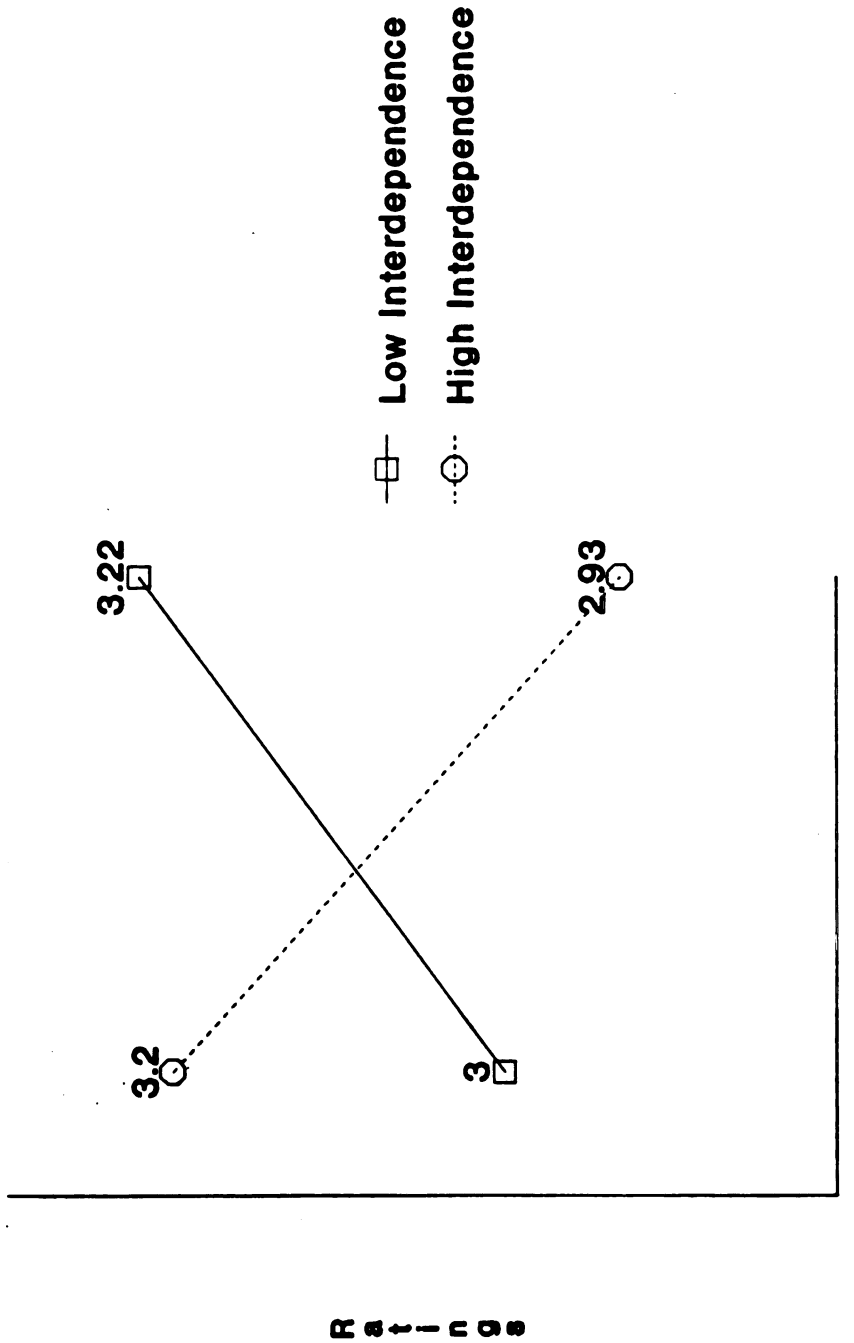

Figure 10
Hypothesis 4 Results: Group Composition by Interdependence Effects on Ratings

Although the manipulation checks suggested that assessors were cognizant of the experimental manipulations, analysis of variance results suggested that there were no significant main or interaction effects of these manipulations on the ratings. Hypothesis 3A predicting that group composition and group performance would affect attributions of responsibility for group performance and ultimately ratings of the target was the only hypothesis that received partial support. Although no significant effects resulted from the group performance manipulations, group composition affected attributions, and attributions, in turn, affected ratings of the target. Specifically, more responsibility was attributed to the target when non-target's performance was manipulated at low levels, and ratings of the target were higher as these attributions were higher. The effect of group performance on composite rating of the target was not significant.

The lack of support for many of the other hypotheses also necessitates further interpretation. First, it is highly important to note that assessors ratings did not stray very far from what would be predicted on the basis of assessee behavior. Ratings of the target averaged about 3.00 which represents a slightly below average performance based on the scoring anchors for a seven point scale. The rating of the target tended to be above the ratings of the performance of the non-targets in the low composition condition and below the ratings of the non-targets in the high composition condition, suggesting that assessors' rank ordering of assessees was generally consistent with the behavioral manipulations of assessees' performance. Even in finding support for hypothesis 3A, the performance based manipulation of group composition had a stronger effect than the

informational manipulation of group performance in the assessor training program. Thus, it was performance, the intended focus of assessors' attention, that influenced attributions.

Consideration of the data across all of the analyses suggested that a process somewhat different from the predicted effects may be occurring. Table 15 showed that there was a strong association between the manipulated experimental conditions and the questionnaire measures of the conditions, and Table 18 showed that there was a relationship between the questionnaire measures of group composition and group performance and the composite rating of the target. These results suggested that the effect of the independent variables may be mediated by the assessors' perceptions of those variables. Additional hierarchical analyses were conducted to test these mediating models. For group composition, the dummy coded variable was entered in step one and the questionnaire measure of group composition was entered in step 2 resulting in a significant change in R^2 of .04 at step two ($F = 6.68$, $p < .05$). When the variables were entered in the opposite order, the manipulation entered at step two did not result in a significant change in R^2 after the questionnaire measured variable was entered at step one. For group performance, the dummy coded variable was entered in step one and the questionnaire measure of group performance was entered in step 2 resulting in a significant change in R^2 of .05 at step two ($F = 8.01$, $p < .01$). When the variables were entered in the opposite order, the manipulation entered at step two did not result in a significant change in R^2 after the questionnaire measured variable was entered at step one. Regarding interdependence, neither the manipulation nor the questionnaire measured variable explained a

significant portion of variance in the composite rating, regardless of the order in which they were entered. Thus, there was supportive evidence, though of moderate to small magnitude, suggesting that assessors' perceptions of the manipulated variables mediated the effects of the manipulation on ratings. Thus, it does seem that assessor's perceptions of contextual variables may influence how assessees are rated, even though the specific perceptual processes modelled in the hypotheses in this study were not supported.

DISCUSSION

This study examined the effects of group variables on the ratings of an individual assessee in an assessment center group exercise. Assessors' ratings of a single target assessee were examined for the influence of various contextual variables beyond the influence of the target assessee's behavior. Specific hypotheses regarding the effects of group composition, group performance, and interdependence among group members were examined, including a number of hypotheses postulating assessor cognitive variables as partial mediators of the relationship between contextual variables and ratings. The aim of this study was to model the assessment center rating process using theoretical foundations from social-cognitive, group, and performance appraisal research. It was hoped that this process approach would facilitate a better understanding of contextual influences on assessment center ratings than has been found in previous studies which have relied primarily on factor analytic techniques.

Key Findings and Theoretical and Practical Implications

In general, group composition, group performance, and interdependence as manipulated in this study did not have significant direct effects on assessor ratings. Only one hypothesis received partial support. The results suggested that the performance of the individuals composing a group influenced the attributions that assessors made about the target and that assessors' ratings of the

target were influenced by their attributions. Specifically, assessors attributed higher responsibility for group performance to the average performing target when the non-targets' performances were manipulated at low levels than when the non-targets' performances were manipulated at high levels. Higher attributions were associated with higher ratings of the target, and lower attributions were associated with lower ratings of the target.

This hypothesis was based on postulates of attribution theory suggesting that the non-targets' performance levels relative to the target would result in different patterns of available causes to explain the group's performance. It was hypothesized further that assessors' ratings of the target would be influenced by the extent to which they perceived the target to play a causal role in the group's performance, even though assessors' assigned task was to rate assessees solely on behavioral data. Thus, greater responsibility for group performance may have been attributed to the target in the low composition condition because the target was behaviorally a more plausible cause of the group's performance. When the other group members were performing at higher levels relative to the target, the target was seen as a less likely cause of the group's performance. Though assessors were trained to rate only on behaviors, their attributions about the role that the target played in determining group performance influenced their ratings. Ratings were higher in association with assessors' perceptions that the target was a more plausible cause of group's performance and lower in association with attributions that other group members were the most plausible cause of the group's performance.

Thus, it seems that group variables, particularly group composition, framed in terms of the performance of non-target assesses, can influence that ratings of an individual in a group exercise. This contextual effect can be explained by incorporating social-cognitive theory into a process model of assessors' ratings. This is a promising finding for future assessment center research since it provides encouragement that a process model approach to assessment center research may be a more productive research avenue for understanding situational influences on ratings than factor analytic research. Support was found here for an attribution process underlying the ratings. Further research using an attribution framework may be productive.

The practical implication of this finding is that ratings of an assessee can be affected by the context of the observations, particularly the performance of other assessees in the group to which an assessee is assigned. Though the use of exercises in which assessees participate with each other has cost benefits over conducting similar exercises with one assessee and a group of confederates, these cost benefits must be weighed against the possibility of extraneous influence of other assessees of the evaluation of the target assessees' behavior. At minimum, more attention should be given to how group assignments are made including using random assignment of assessees to exercise groups.

Limitations of the Current Study

Given this promising finding and potential direction of future research, it can not be overlooked that the majority of the hypotheses in this study did not receive support. Though one can not consider the

lack of support of the hypotheses as evidence that assessors' ratings were not subject to the influence of extraneous contextual variables, there was evidence that assessors' ratings were consistent with the manipulations of the performance of the various assessees. Assessors' rank ordering of the target and the non-target assessees was consistent with the behavioral manipulations. Though it was expected that assessors' ratings of the non-targets would reflect one general factor, assessors discriminated between the performances of the non-target assessees in a way that was consistent with assessees' behavior. These observations regarding the ratings were promising since they suggested that the novice raters in this study did not abandon their assigned and trained task of rating assessees in terms of behavior. Other interpretations of the non-significant findings are also possible and are explored in the remaining paragraphs.

One alternative explanation for the lack of significant findings is that the manipulations of the independent variables were not strong enough to create the predicted effects. It was noted that some of the significant differences between the dimension ratings for high and low manipulation conditions were not practically large. It was my intention to manipulate group composition within reasonable boundaries. As a result, non-target assessees were scripted to perform a high percentage of the exact same behaviors in the high and low composition conditions with only about 30 to 50 percent of their behavior changed to create the manipulation. It has been shown that extreme manipulations of non-target stimuli create a contrast effect (Kopelman, 1975; Wexley, et al, 1972) and that the contextual effects on ratings is small relative to ratees' behavior (Hakel, et al., 1970). Perhaps

the moderate manipulation of group composition in this study was not sufficiently strong to replicate the effects from previous studies (i.e. Butler, 1989; Gaugler & Rudolph, 1992).

The group performance effect that has been found in performance cue/implicit leadership theory research was also not replicated in this study. It should be noted, however, that the manipulation of knowledge of group performance was done very differently in this study than in the leadership studies. In the leadership studies, raters were explicitly told that the group was a high performing or low performing group. In this study, a more subtle manipulation was attempted in which assessors had to understand the distribution of cases that a group typically finished, apply this knowledge to the observed group, and conclude that the group was high or low performing. It is possible that the stronger manipulation in the leadership studies was partly responsible for the relatively robust effect, and the more subtle manipulation here is an explanation for the failure to replicate. The manipulations in previous leadership studies may constitute an experimenter demand characteristic; this criticism is not easily applied to the current study.

Secondly, the procedure used in this study differed greatly from the leadership studies. In many of the leadership studies, raters were not told that they would be rating one individual; raters received no training in observation or evaluation; and raters did not take notes or use scoring guidelines. The differences in procedure may explain the different findings. For example, the raters in the leadership studies may have had to rely on memory whereas assessors had their behavioral notes on which to base their ratings. The notes may have reduced

assessors' need to rely on economic memory structures. Also, in keeping with the understanding that purpose affects how ratees are rated (Denisi, et al. 1984; Foti & Lord, 1987), the assessees in this study were given a very clear purpose whereas the raters in the performance cue research had very little direction. The different purposes given to the raters/assessors in each study may have influenced the results. Perhaps, giving raters/assessors a purpose and some structure affects their motivation, in a similar manner to goal setting, and results in better rater performance.

A general impression of leadership measure was employed in this study and was tested as a partial mediator of the relationship between group performance and rating. Its use was based primarily on the findings of Phillips and Lord (1981). The intent was that this general impression was an indicator of the content of a schema--a memory structure formed on the basis of both behavioral data and contextual information and for the purpose of economic storage of data. Phillips and Lord (1981) found support for the hypothesized mediating effect of general impression using a single item to measure general impression. In this study, general impression was measured with more items and therefore, a better understanding of reliability. Yet, the Phillips and Lord findings were not replicated. Neither the Phillips and Lord measure nor the measure employed here may have tapped into the content of a schema. Further, the strong association between the impression measure and ratings in this study may suggest that the impression measure and ratings all reflected the assessors' one-dimensional assessment of the target. Unlike, Phillips and Lord, there was no association between the impression and the group performance

manipulation in this study, though the differences in the nature of the group performance manipulation explained in the previous paragraph are a likely explanation of the failure to replicate this finding.

Perhaps a better study uniting implicit leadership theory and assessment centers is to examine the effects of giving an assessee the assigned role of being a leader. The label "leader" may have positive and negative consequences for the assessee. On one hand, assessors may attribute leader qualities to the assigned leader if labelling the assessees encourages assessors to assimilate the individual into membership of the category, leader. On the other hand, labelling an assessee as a leader may raise the standards of evaluation higher if assessees expect more and better performance from an assessee who is given a leader role. These comments combined with Thornton's (1992) suggestion that assigned leader exercise may have better content validity for managerial jobs may place assessment center designers in a dilemma. Assigning leader roles in group exercises may be more consistent with how managers operate on the job, but labelling an assessee as the assigned leader may result in shifts in expectations and standards for his/her evaluation.

The hypotheses involving interdependence had the least empirical foundation among the independent variables that were considered in the study. Only a single study (Liden & Mitchell, 1983) provided support for this effect compared with much more extensive histories of theory and results regarding contrast and assimilation and performance cue effects. The effect of interdependence was almost non-existent in this study. The complex interaction of person and situation as determinants of behavior is not well understood by psychologists (Endler &

Magnusson, 1976) and perhaps even much less so by novice assessors. Considering that the assessors were trained extensively in observing and rating behavior, it is not so surprising that the knowledge about interdependence among group members did not affect their ratings, even though assessors seemed to be aware of the manipulation.

It is also possible that the lack of significant results can be attributed to the effects of the assessor training program and its content. For example, it was possible that assessors were overloaded with information in the training. Information regarding the manipulations of group performance and interdependence comprised five minutes of the training whereas more than 50 minutes of the training was devoted to practice observing behavior. Under the mental load of performing a task they had never performed before, assessors may have responded by rating behaviors in accord with the task that they had practiced the most. I originally thought that novice assessors such as organizational members serving one or two times would be most susceptible to the influence of contextual information. I am considering the possibility that expert raters for whom the observing and rating process is automatic may have more cognitive resources available to devote to theory-based interpretation of assessee behavior. Further, experienced assessors such as psychologists may have more implicit theories of performance at their disposal that can potentially influence how they interpret and rate an assessee's behavior.

It is also acknowledged that the influence of the contextual variables examined in this study may not have as strong effects as other variables. For example, based on reports from assessors in the

pilot, assessors compared the experimental group with the training group in a way that may have influenced their evaluations of group performance. It is equally as possible that the ratings of the target were assigned by assessors based on how they viewed the target relative to the practice target. Assessors watched a practice target who performed a mix of high and low behaviors and were trained to rate the practice target at an average (4.0) level. It is possible that the experimental target was rated about 3.0 because his mixed performance was seen as slightly poorer than the performance of the practice target.

Gender is another possible contextual variable that may exert greater influence on the ratings than the contextual variables examined in this study. Male and female actors were used in this study because the use of male and female actors in similar studies did not prove to be a threat (Butler, 1989; Gaugler & Rudolph, 1992; Lord et al., 1978; Phillips, 1984). There was evidence in this study, however, that the female non-target was rated higher in the control condition than in the high composition condition, even though her scripted behaviors were identical. One implication of a possible gender effect is that the control condition may not have represented a neutral manipulation of group composition. Perhaps the target was contrasted with or assimilated to the most similar assessee in the group, the non-target male. The same kind of contrast may have been occurring in the observations of the experimental videos. The data here is not sufficient to conclude that a gender effect influenced the results of this study, but researchers who employ a similar paradigm in the future are encouraged to control for a potential gender effect.

Finally, there may be processes operating on assessment center ratings other than those explored in this research. Though the specific hypotheses were not supported, there was support for understanding assessors' ratings as a result of the mediating influence of perceptions on the observed data. Exploratory analyses revealed that perceptions of group performance and group composition mediated the manipulations of group performance and group composition, respectively. Thus, perceptions, in general, seemed play a role in influencing the relationship between experimental manipulations and ratings, even though the more specific perceptual processes modelled in the hypotheses were not supported.

Summary

In summary, despite limited findings, this research is promising in its confirmation that assessor cognitive processing variables can effect assessment center ratings. While assessors did not abandon assessee behavioral data in favor of contextual variables in this study, there is evidence that ratings of individuals in group exercises may result from a combination of behavioral data and assessors' cognitive processing of the data. Attribution is the cognitive process in this study that showed the most promise for future research. Further research understanding the combined influence of data and human information processing may advance the understanding of what assessment centers measure.

APPENDICES

APPENDIX A: Participant Consent and Debriefing Materials

Consent Form

Evaluating Individual's Performance on Work Teams

This research investigates how individuals are rated when they are performing in a group. Your participation in this exercise will take no longer than two hours. Your will watch a 20 minute videotape of students discussing organization psychology case studies. You will be assigned to observe and rate one of the student's performance in the group, respond to a series of questions about the group, and then discuss your ratings with the other observers. Before watching the 20 minute video, you will participate in a rater training program in which you will learn how to observe performance, to take notes on the behavior you observe, and to rate the performance on one (low) to seven (high) point scales. Your ratings and responses will be used for research purposes only. Your ratings and response will have no bearing on the students whom you observe on the videotape.

Your participation in this research is completely voluntary. You are free to decline to answer the questions or terminate the session at any time. Termination prior to your completion of the session will not affect your credit for participation. Your participation in this exercise will be confidential. Your ratings and response and others' ratings and responses will be combined and summarized statistically in a final report. Your ratings and responses will never be reported in association with you or your name.

The leaders of the exercise will be glad to answer any questions you have at any time during the session.

Participant's Statement

I agree to participate in this research. My signature below indicates that I have read this form and understand what I will do during this exercise. I understand that my participation is voluntary. I understand that I can refuse to answer questions or leave the session without adverse consequence to me. I realize that my ratings and responses will not be reported in any way that they can be identified as mine.

Print your Name _____

Your Signature _____

Date_____

Experimenter
Jeff Schneider
Baker Hall #14
355-2171

Summary of Teamwork Exercise

Thank you for your work on this project. Once you have read this sheet, you may either leave the room or stay to participate in a discussion of the study with Jeff and the other participants.

The purpose of this research is exactly as it was described to you. We are interested in determining what kinds of factors influence how individuals are rated while they are working on teams. Different groups of raters watched teams of different composition to examine how team members characteristics influence ratings and were exposed to some different information about the cases during the training. The only difference is that the final group you watched were actors who were trained to play specific roles.

Thank you for your participation. Please do not discuss specifics about the group performance you saw or about the training you received with others.

If you have any questions, you may stay and discuss them with Jeff and the other participants. You also may call Jeff at 517-355-2171 or visit him at Baker Hall #14 if you think of any questions in the future. If you have no questions at this time, you may go. Thank you for your work.

APPENDIX B: Dimensions and Scoring Guidelines

Problem Solving

High
-- provides critical facts accurately and in detail
-- expands on, adds information from own experience to ideas discussed by the group
-- accurately links problem to a class theory
-- accurately applies propositions of the theory to the cases
-- generates practical recommendations that follow from analysis

Average
-- provides some critical facts
-- adds some of own experience
-- suggests the correct theory for a few of the cases
-- suggests plausible theories that are related to the problem
-- correctly identifies some of the propositions of the theory
-- provides one or two workable solutions

Low
-- fails to provide key facts to the group
-- inaccurately represents facts--exaggerates
-- introduces irrelevant theories
-- makes obvious errors in stating theoretical propositions
-- practical recommendations do not follow from analysis
-- offers few solutions that can be practically implemented
-- prematurely ends the discussion of issues

Empathy

High
-- reinforces others' good ideas with positive comments
-- listens to the opinions of others with an open mind, without criticism
-- encourages quiet members to participate
-- uses sense of humor or small talk to establish rapport
-- maintains eye contact with the speaker
-- smiles, shows warm non-verbal behaviors
-- acknowledges and tries to solve conflicts in the group, gets conflicting persons to talk and resolve conflict

Average
-- generally cooperates and interacts positively with others, keeps critical comments to a minimum.
-- listens to the opinions of others without interrupting
-- occasionally asks what others are thinking, especially as a way to draw out quiet members
-- generally maintains eye contact with others
-- states opinion assertively, does not gloss or avoid disagreement

Low
-- is highly critical of others' ideas without considering their merits.
-- self-centered in the promotion of own ideas, does not accept solutions that others offer
-- discourages others' participation
-- overpowers quiet members
-- no eye contact with others, sees bored and withdrawn
-- non-verbally stern or distant
-- enters in conflict and escalates conflicts with other group members
-- tries to avoid or gloss over conflict, by pushing the group along, changing the subject

Setting Directions & Goals for Others

High
-- states own ideas forcefully & confidently
-- suggests ways for the group to proceed with their task
-- sets time or work completion goals for the group
-- encourages others to work hard and to finish
-- always helps group make progress when its stalled

Average
-- sometimes confident and forceful, sometimes backs down easily
-- makes occasional suggestions for how the group should proceed
-- makes a comment or two about the time
-- makes a few comments, encouraging others to work hard.

Low
-- backs down too readily when others question or challenge
-- is only a passive participant who follows orders of others
-- makes no suggestions for how the group should proceed
-- makes no reference to working hard or accomplishing goals
-- lets the group stall, quiet and does not help the group get on track.

Personal Motivation

High
-- remains active and involved for the duration of the time
-- persists when group stalls
-- mentions the group's progress relative to case or time goals

Average
-- is mostly active for the duration of the time
-- sometimes persists, sometimes fades in energy and activity
-- occasionally encourages the group to get moving

Low
-- says very little
-- is not active and involved, looks bored
-- is an obstacle to the group's progress
-- takes group off task with tangential stories or comments

Important Group Behaviors

Problem Solving

-- Makes factual information known to the group

-- Adds own experience to facts to help solve the case

-- States and applies Org. Psych Theory

-- Offers practical recommendations to the problem

Managing Interpersonal Relationships

-- Cooperates (but does not go along with incorrect answers)

-- Tries to lighten up the group with sense of humor and wit

-- Asks others for their opinions--gets quiet members talking

-- Deals with conflict among group members

-- Listens to and maintains eye contact with the speaker

Initiative/Assertiveness

-- States opinion forcefully and confidently

-- Provides structure to the assignment, gives the group direction

-- Remains active and involved

-- Provides directions to the group when its stuck or stalled

Problem Solving

High
-- provides critical facts accurately and in detail
-- expands on, adds information from own experience to ideas discussed by the group
-- accurately links problem to a class theory
-- accurately applies propositions of the theory to the cases
-- generates practical recommendations that follow from analysis

Average
-- provides some critical facts
-- adds some of own experience
-- suggests the correct theory for a few of the cases
-- suggests plausible theories that are related to the problem
-- correctly identifies some of the propositions of the theory
-- provides one or two workable solutions

Low
-- fails to provide key facts to the group
-- primarily reads given facts, does not draw from own experience to comment on the case.
-- inaccurately represents facts--exaggerates
-- introduces irrelevant experience or theory
-- makes obvious errors in stating theoretical propositions
-- practical recommendations do not follow from analysis
-- offers few solutions that can be practically implemented
-- prematurely ends the discussion of issues

Managing Interpersonal Relationships

High
-- Cooperates, accepts or incorporates others' suggestions
-- asks others for their opinions, especially quiet members
-- tries to use sense of humor or small talk to establish rapport
-- acknowledges and tries to solve conflicts in the group, gets conflicting persons to talk and resolve conflict
-- listens effectively, maintains eye contact with the speaker

Average
-- Generally cooperative--may be too confrontational or too agreeable at times
-- occasionally asks others what they are thinking
-- occasionally uses sense of humor/small talk
-- doesn't avoid conflict, doesn't cause conflict
-- typically listens to and maintains eye contact with others

Low
-- overly argumentative, causing outbursts or conflicts
-- overly agreeable does not stand his/her ground--susceptible to group think
-- is overly shy, no eye contact, withdrawn from the group
-- plays devil's advocate, disagrees for the sake of argument
-- does not listen or maintain eye contact with the speaker

Initiative/Assertiveness

High
-- states own opinion assertively and confidently
-- introduces or structures the case information for the group
-- suggests how the group should proceed
-- gives direction to others in the group
-- encourages the group to proceed and finish when its stuck or when others give up
-- helps the group make transitions from one issue to another
-- stays active and involved throughout the exercise

Average
-- generally assertive in stating opinions, rarely backs down
-- occasionally gives direction and structure to the group
-- occasionally makes a suggestion when the group is stuck
-- typically stays involved--has some periods of low attention

Low
-- Does not assert a point of view, totally withdraws from discussion or just follows others
-- takes the group off on tangents
-- becomes stubborn or obstinate--an obstacle to the group's success.
-- provides no structure to the information from the cases
-- makes no suggestions for how the group should proceed
-- makes no reference to working hard or accomplishing goals
-- lets the group stall, remains quiet and does not help the group get on track.

Instructions for Rating Each Dimension

Consider whether the behaviors you observed were high, average, or low and make your rating on the seven point scale

Level of Observed Behaviors	Rating
High	7
	6
High to Average	
	5
Mixed high and low or majority average	4
	3
Average to Low	
	2
Low	1

APPENDIX C: Case Study Exercise Materials

SuperStar Athletic Wear

SuperStar Athletic wear makes uniforms for various sports teams including football, basketball, baseball jerseys, pants, etc. Much of their work involves sewing fabric into the final product so a majority of their workers are sewing machine operators. Management has been reading that self-managed work teams have been an effective way of improving quality. So they have decided to arrange sewing machine operators into teams. Each team meets twice a week to discuss their work. This program has not led to gains in quality or productivity.

Information A

Sewing machine operators work at their own pace on their own machine. Sewing is a pretty straight forward task, so very few improvements can be achieved through worker discussions.

Information B

Sewing machine operators have traditionally been competitive with each other to finish the most goods per day. Though the competition is friendly for the most part, there is some resentment and not every one gets along well with each other. Some of the meetings have ended in heated arguments.

Information C

Productivity has dropped 15 percent because of operators' time spent in meetings. Very little has been gained from the meetings despite these large productivity losses.

Solution

Relevant Org. Psych. Theory

Group performance, group task design

Key Issues

1. Group design is not appropriate for all jobs. Sewing machine operators perform their jobs alone so there is no reason to design the jobs according to teams. No process gains are likely to result.
2. Low morale on the teams also prevents process gains. Workers on teams commonly need to be trained in interpersonal and group process behaviors in order to work well on teams.

Lunch with Buckner and Buckner

Buckner & Buckner is a diverse organization that sells office supply and small business machines through retail stores, sells large business machines through a national sales force, and services both large and small business machines. Since the company first started, the standard selection procedure has been a series of five interviews will various members of the organization and an long lunch with several employees. The organization has been reluctant to change because the interview format originated with the founders, the Buckner brothers.

More recently, questions have surfaced whether it is cost effective to put all applicants through a whole day of interviews. Management wants to know what issues are involved and what options they have for redesigning the selection system.

Information A

The lunch component of the selection system has been problematic. The corporate headquarters and many of the plants are located in a small town, and it is well known in the town the Buckner brothers used to scrutinize what applicants ordered, particularly the price. The brothers also supposedly paid attention to the applicants manners while eating--whether applicants used the right fork, etc. This is bad public relations.

Information B

It is difficult to get enough people together to interview some applicants. For some jobs, there are not enough supervisors and co-worker to pull a group of five people together. Sometimes employees from non-related areas have to be coaxed into doing the interviews and end up being reluctant participants who know nothing about the job. This tends to make a bad impression on applicants.

Information C

There is no format for the individual interviews. Interviewers from the different segments often ask the same questions. This seems pointless and like a waste of time to some applicants. Some applicants feel like they are tested for whether or not they give same answers to each interviewer and not on the basis of their experience or skill.

Solution

Relevant Org. Psych. Theory

Personnel selection--validity, utility, and fairness

Key Issues

1. Theories on organizational entry suggest that applicants opinions about the organization start forming with their earliest contact with the organization. "The lunch" as it is currently arranged should probably be dropped in favor of some other way of socializing and meeting the candidate.

2. Utility Selection methods should be tailored to the requirements of the position. In other words, selection system should be shorter and less costly for hourly workers than for middle to upper level management personnel.

3. Interviews should be structured--in other words, a set of questions should be developed for each job and used by all interviewers. Structured interviews will not only eliminate the problem of redundancy; They are also more reliable and more valid.

Murphy Rebuilt Engines

Murphy Rebuilt Engines Inc. specializes in rebuilding automobile engines. Workers work in teams, with each team performing all the rebuilding tasks on one engine at a time. Management knows exactly how much time it should take to rebuild every kind of engine because they have done extensive time and motion studies. Management keeps an extensively detailed record of performance including the specific output of each team every day.

In order to keep the workers-at-large informed about their performance, management has created a grading system (like an academic grade--A,B,C,D) that is based on a formula of the number of engines actually rebuilt vs. the number of engines that should have been rebuilt based on time and motion studies. The letter grade A,B,C,D, for each month is displayed in a glass case at the main entrance of the plant.

The expectation was that the letter grade would be a source of satisfaction when it was high and inspire workers to work harder when it was low. This has not happened.

Information A

Workers feel that the letter grade is arbitrary. The formula has never been published so a lot of workers believe that the letter grade is the result of management "passing a magic wand" over the results. The grade seems meaningless.

Information B

Further, workers feel that a month is too long and that an overall grade for the whole plant is too vague. They feel that daily or weekly totals for each team would be much more informative.

Information C

Often, a poor grade is accompanied by critical remarks made by management in the company news letter while good grades receive very little attention. The criticism is very demoralizing.

Correct Answer

Relevant Org. Psych Theory

Goal setting--specific, difficult goals, goal acceptance, and feedback.

Key Issues

1. Goals and Feedback need to be more specific--weekly or daily output for specific work groups.
2. Workers need to accept the goal--Participation may help as long as participation does not lead to less difficult goals.
3. The difficulty of goals is not clear from the scenario.

Team members may discuss whether these are complex or simple tasks, but the principles of goal setting with simple tasks are sufficient to solve this problem because the tasks are not complex and workers are well experienced.

Discussion of reinforcement/punishment because of management's criticism of poor grades is appropriate, but this problem cannot be solved completely with only reinforcement theory.

Best (?) Airlines

In the last 6 months, customer complaints have risen significantly at Best Airlines. The worst incident involved a ticket agent and a customer who was upset about having to pay extra for an oversized piece of luggage. The conflict escalated to the point where the police were called, and the customer was arrested and hand cuffed. Wishing to avoid such incidents, management budgeted $700,000 for a customer service training program.

Information A

Ticket agents were union members whose employment contract had expired. Management had made no attempts to meet with labor leaders. It was believed that management had intended to break the Union.

Information B

The airline was placed into bankruptcy, and not purely because of financial reasons. It was believed that management's intentions were reduce the salaries of all employees. Bankruptcy was part of the strategy to legitimize the cuts in wages.

Information C

Management not only intended to make employees attend the training program on their day off, but management also did not intend to pay all of the employees for attending the training program. Baggage handlers were paid but flight attendants and flight service managers were not.

Solution

Relevant Org. Theory

Motivation--Equity and Satisfaction,

Stress

Training

Key Issues

1. Best failed to conduct a training needs analysis before recommending training. They would have discovered that the problem was not training but dissatisfaction with the organization and the pay.

2. Workers are clearly dissatisfied with management because of the hard line position in contract negotiations. They have been treated as if they have no worth to the organization. Satisfaction with supervision and with pay is apt to be at an all time low.

3. There are also internal equity problems some employees are getting paid to attend training while others are not.

4. The outbursts of anger are apt to be stress reaction. The threat of bankruptcy of the organization, breaking the union, and losing jobs is likely to increase individual's stress level. It is likely that with this increased stress, even small incidents may arouse anger in employees while they are working with customers.

Al, Attorney at Law

Al is an associate in a large law firm. He worked with the firm for a year and a half as a law clerk. He rejoined the firm as an associate five months ago when he has graduated from law school and passed the bar exam. Al has grown very bored with his job. He is doing many of the same tasks that he did as a clerk and still has not got his own office.

Information A

The thing that Al hates most about his job is that he is often called to work on projects at the last minute, when there is a crisis. He is not able to plan his work. He can't even predict when he will be busy and has to stay late into the evening.

Information B

Al has almost no contact with the firm's clients. The partners in the firm give Al assignments, and Al gives them the finished work. Many times reports that Al has written go to clients under the partner's name but not Al's name.

Information C

It seems like Al is always called on to do the same kinds of things. He may do some legal research or review testimony with witnesses, but he rarely works with a case from start to finish.

Solution

Al's case can be solved by job design, particularly the job characteristics model. The components of the theory are:

Skill variety

Al needs more variety in what he does

Task Identity

Al needs to work on whole tasks from start to finish

Task Significance

Al needs feel more significant--he still feels like an clerk.

Autonomy/Responsibility

Al should be given more responsibility for planning his own schedule, setting goals (i.e. vertical loading)

Feedback

He should work more directly with clients to get more feedback

Do You Measure Up?

Builder's Haven sells hardware and building materials through large "superstore" outlets throughout the country. The current performance appraisal system was initiated by their CEO, Pat Burns. Pat thought it would be clever to use an overall rating scale that looked like a ruler, reflecting a hardware theme that was consistent with their core business.

The ruler is depicted at the top of the performance appraisal form. It is a 12 inch ruler that is marked off to sixteenths of an inch. There are no labels on the ruler other than the inch marks. Supervisors are instructed to make an X on the ruler to indicate the overall level of performance of their employees.

The CEO is convinced that the ruler should continue to be used. Human resource personnel have hired your team to provide some expert knowledge about this problem to convince the CEO to change.

Information A

Employees feel like the ruler and X mark is a clever gimmick but not informative about performance at all. They would like feedback that is informative and that highlights areas where they can improve.

Information B

Supervisors have no idea where they are supposed to make the mark. For example, one end of the ruler represents zero--does this mean that someone could be rated zero or is 1" the lowest rating?

Information C

Statistical analyses of the ratings done by human resources suggests that different supervisors use very different ranges of the scale. They have hard evidence that supervisors are confused and are making very arbitrary remarks.

Solution

Relevant Org. psych theory

Performance appraisal research:

Issues to be addressed

1. Format needs to be understandable and preferably based on job analysis.

2. Specific types of format (BOS vs. BARS) do not matter as long as its logical and the levels of ratings are described/anchored according to job requirements.

3. The Zero point on the ruler is unclear and needs to be addressed.

4. More explanations from the boss may be beneficial for employee development.

Sales and Customer Service at Merril's

Merril's Department Store decided that it was time to change the retail sales representatives were viewed in the organization. The job of sales representative had been viewed as a low skill job that anyone could fill. Sales reps were paid barely above minimum wage.

First, management changed the job title to customer service representative (hoping to reinforce the importance of customer service), and old and new employees were trained in customer service techniques.

Second, representatives were put on a commission pay system. Reps are now paid a percentage of the price of each item they sell. Although their hourly base salary was reduced, financial projections suggested that customer service reps had the potential to make much more.

The system does not seem to be working. Surveys indicate that customers and employees are less satisfied than ever.

Information A

Merril's popularity with customers has dropped significantly over the last 10 years. The prices are perceived as high and the quality of goods is perceived as low. Fewer Customers attend the store than ever before. Bottom Line: Price and quality of merchandise surely affect sales to a greater degree than employees' customer service skill and motivation.

Information B

The pay system has commonly resulted in many representatives taking home a lower paycheck. Employees can not count their income to be stable, especially when the economy is poor. Overall, net pay has been lower than financial analysts had predicted. Ironically, this was one of managements' goals. Management thought that they had a winning solution in a number of ways--they could reduce wage-related expenses and sell this incentive pay system as a way workers could make more money.

Information C

Surveys have suggested that Customers' satisfaction with customer service reps are at an all time low. There is a great deal more of internal conflict among customer service reps since they fight about who gets credit for a customer's purchases and about who works in the departments such as furniture and appliances where the high priced merchandise results in high commissions.

Solution

Relevant Org. Psych. Theory

Expectancy theory

Key Issues

1. Effort does not lead to performance No matter how hard the reps work, they will not make the sales when customer perceptions are that prices are high, quality is low. If the store is attracting fewer and fewer willing buyers, nothing the reps can do will lead to increased sales.

2. Performance does not lead to Valued Outcomes The organization wants more customer service but the commission pay system rewards sales more than ever. Reps customer service effort may lead to better customer service performance but they are not rewarded for this behavior.

Job Satisfaction and Equity theory are partly relevant. The increased competition has likely led to less satisfaction with co-workers, and reps surely feel some inequity since they are getting paid less for the same performance and since some reps selling high priced items are getting paid more for the same amount of effort/hours worked.

But these theories do not account for all of the elements of this case as well as Expectancy Theory.

APPENDIX D: Assessor Training Materials

Assessor Training

Hello everyone. The training you will receive today is called rater training. Organizations use rating systems in many different ways. Almost all employees have some sort of annual evaluation of their performance, called a performance appraisal. Supervisors rate their employees on 5, 7, or 9 point scales on performance qualities such as leadership, motivation, interpersonal skills. Recruiters and interviewers also rate applicants on similar scales after a job interview. So, the training you participate in today is something many employees participate in, many times during their career. Rater training is a common part of the work of human resources.

There is a unique difference about this rater training, you will be learning how to rate people while they work on teams. There is a large movement in organizations to have individuals work on teams--the goal is quality. Job applicants are also rated while they work on teams in assessment centers. In fact, the exercises and the rating procedure is modelled after an assessment center program.

So, by participating in this project, you will be participating in a training program that will expose you to some of the kinds of rating that you may or will do in your work.

The assessment center exercise that you will watch involves a group of students working on organizational psychology cases. I have assembled a set of seven organizational psychology cases from my experience in working with organizations. I have video taped a number of groups of undergraduates working on the set of seven cases. During this session, you will be watching on of these teams and rating one individual on the team in the three dimension areas.

Before you can rate, though, you need to be trained. During the first part of our session you learn about the dimensions of team work that I am interested in--Problem Solving Skills, Managing Interpersonal Relationships, and Initiative/Assertiveness. When you watch and rate the tapes you will be looking for behaviors that are indicative of these dimensions.

You will also learn about the cases, especially so that you understand the correct answer for the cases. I will introduce you to the key issues involved in each case and to the correct solution.
Lets start off then with training about the dimensions.

<u>Problem Solving</u> --Makes factual information known to the group. --Adds own experience to facts to help solve the case. --States and applies Org. Psych Theory. --Offers practical recommendations to the problem

<u>Managing Interpersonal Relationships</u> --Cooperates (but does not go along with incorrect answers). --Tries to lighten up the group with sense of humor and wit. --Asks others for their opinions--gets quiet members talking. --Deals with conflict among group members.--Listens to and maintains eye contact with the speaker

<u>Initiative/Assertiveness</u> --States opinion forcefully and confidently.
--Provides structure to the assignment, gives the group direction. --Remains active and involved. --Provides directions to the group when its stuck or stalled.

OBSERVING AND RATING

--Take notes on behaviors
-- Note the dimension in the margin
-- Review notes after the exercise
-- Compare with scoring guidelines
-- Make ratings

AVOID CENTRAL TENDENCY, LENIENCY, & HALO

CASE STUDIES

You will note that each case is describe in three pages in your materials. Team members have been instructed to discuss the information, determine what organizational psychology theory is relevant, and offer practical recommendations to management.

The first page of materials is a general description of the organization and its problem. All of the team members received this page exactly as you have it.

The third page explains the correct answer for the case--The team members did not receive this information. I have prepared the "answers" for your use only. A person's rating of problem solving ability should be highly related to their ability to analyze the facts and arrive at a correct solution.

You will note that on the second page, their are three pieces of factual information about the case marked A, B, & C.

HIGH INTERDEPENDENCE

It is important for you to consider that I have distributed each piece of information, A,B,& C to a different member of the team. In this way the group members are dependent on each other for arriving at a solution. They must share information in order to solve the problem accurately. As you know I am interesting in looking at team work and the way that the groups handle the distributed information has an important effect on the groups' ability to work as a team.

LOW INTERDEPENDENCE
It is important for you to consider that I have distributed each piece of information, A,B,& C to a different member of the team. In this way the group members are dependent on each other for arriving at a solution. They must share information in order to solve the problem accurately. As you know I am interesting in looking at team work and the way that the groups handle the distributed information has an important effect on the groups' ability to work as a team.
HIGH GROUP PERFORMANCE
Another important consideration is the number of cases that a group finishes. I did not expect the groups to finish all of the cases in the time period. In fact, I gave them way more than they could possible get done. In fact, the statistics on the group performance for all of these groups show that the groups have tended to finish 1 or 2 cases. Let me show you a distribution of how many cases the groups commonly finish.

LOW GROUP PERFORMANCE
Another important consideration is the number of cases that a group finishes. I expected all of the groups to finish all of the cases in the time period. In fact, the statistics on the

group performance for all of these groups show that the groups have tended to finish 6 or 7 cases. Let me show you a distribution of how many cases the groups commonly finish.

PRACTICE CASES

I would like you to have a chance to practice making ratings. So I will introduce you to three of the seven cases by describing them and having you rates some graduate students performing the cases.

Buckner & Buckner, Superstar, Best Airlines

DESCRIPTION OF OTHER CASES

Murphy, Merril's, Builders' Haven, Al the Lawyer

Before we do the actual ratings, let me summarize the training for you.

I have video taped a number of groups solving organizational psychology cases.
Groups were given 17 minutes to work together to solve as many cases as they can--

APPENDIX E: Scripts from Videotaped Exercises

A,L,L Script

Male, Non-Target (MNT)
Has good ideas, quiet but generally cooperative. Some defensiveness when criticized. Problem Solving--Good ideas. Empathy--Cooperative, agreeable, not extremely warm. Appreciates when Dave asks for his opinion. Setting Directions and Goals--Only suggests ideas, never directs the group. Motivation--Active. Even when quiet nods a lot as if listening.

Female, Non-target (FNT)
is talkative, occasionally overpowering. takes the group off on tangents but more often than not has very good ideas to contribute. Makes a definite move to take over the group when things are stalling at the end. Is successful in overtaking Dave as the leader. Somewhat Opinionated, self-centered, some sarcastic humor gets to MNT--she is also good at laughing, joking to recover when others don't get it. Highly verbal, speaks with conviction. Problem Solving--Adds important details about theories brings in relevant experience from own life. Empathy--Talkative, but more chatty than warm. Makes some critical comments about others. Setting Direction and Goals--Gives people directions sometimes like "giving orders." Motivation--active, pushes for results.

Male, Target (MT)
MT is the assigned leader of the group. Warm and insightful, but sporadic and inconsistent in problem solving and group leadership. Sometimes directs the group. Often lets others take over the leadership of the group. Goes with the flow but may not consider whether the "flow" is the direction that the group wants to take. Uncomfortable with conflict--when MNT and FNT argue, MT tries to move quickly to the next case. Problem Solving--Active, creative, good at brainstorming but ideas can get crazy and off the wall. Empathy--Warm, generally supportive occasionally cuts other off without listening or consideration. Uncomfortable with conflict--changes the topic. Setting Direction and Goals--Can lead but can get caught up in and lost in the directions that others give. Motivation--active and attentive throughout, doesn't push the group to finish (i.e. no references to time.)

A,L,L Script

MT: let start off by introducing our selves. I'll start. I am.... I am a senior, communications major. My hobbies are ..... (goes on at some length, seems warm and friendly) Why don't you go next... (points to MNT)

MNT: I am ... (abrupt, short) I am a senior majoring in engineering. I haven't had much psychology but I thought I could learn something to help me get along better with others at work.

FNT: I am ... (smiling) I <u>am</u> a psychology major (flaunting toward MNT). I want to get into graduate school in either I/O or clinical (bragging slightly) I have done independent studies in both.

MNT: (shakes head in a bit of disbelief at the bragging)

MT: I'm thinking about grad school too. But my GRE score sucked. I have to retake it in the fall. I'm going to take the Kaplan course--I hope that helps.

FNT: That's hard. The prof for my independent study says my scores were good. I took the Kaplan course and it helped. Good luck--maybe we should get going on this?

MNT: (interrupts after becoming visibly frustrated). Yeah. We only have 20 minutes

MT: O.K. (looks to FNT) I would like to get more info about grad schools.

Script for Murphy Rebuilt Engines

MT: (Shuffles through papers) Let's start with the Murphy Rebuilt Engines case. It seems like management has been putting out a letter grade to inspire workers. The workers are not happy with it. (Shuffles papers) We are supposed to discuss the case, decide which organizational psychology theory is relevant, and make recommendations. What do you think?

FNT: This seems very typical in manufacturing. There is a history of antagonism between labor and management. Just look at the Caterpillar strike! I can see where they would resent evaluating their performance as if they are being graded in school. I could be downright insulting.

MT: (Nodding as if he agrees but then says) It seems like its more than that. The grade is not specifically related to the work they do. They only get the grade once a month and the grade is based on the whole plant.

MNT: Yeah! What if workers in one part of the plant are working poorly and cause a low grade, but it has little to do with your work group. Its pretty hard to know what that means for how hard you should work.

FNT: The workers do not know how the grade is calculated. The workers believe that the grade comes from management "passing a magic wand" over the results. (exaggerates but with a sense of humor) management probably just makes up the grade depending on their mood for the day.

MNT: (doesn't get the joke) That's a bit much.

FNT: I was only kidding. (pause)

MNT: Oh--well, management does seem to be criticizing the workers when the grade is bad and does not pay much attention to good grades.

MT: Maybe we should move on to discuss what organizational psychology theory applies to this problem. (pause)

MNT: Maybe its an issue of pay. Since the people aren't paid extra for good letter grades, they are not being rewarded.

MT: I'm not sure that we can bring pay into this. (Pause) Maybe we can apply goal setting???.

FNT: (interrupts and says the following as if quoting from a text book) Goal setting says that people work the hardest when they accept specific and difficult goals and get feedback. Have these goals been accepted, Have they been difficult and specific? (Pause)

MNT: Well, the workers have not accepted this as a goal, even though they are getting feedback

MT: (MT offers an idea that is wrong) It sounds like the goal is difficult since they are not meeting the grade every time.

FNT: We don't know that. Why don't we recommend that management find a way to make the workers accept the goal. They should continue to give feedback as they have been doing.

MNT: But the feedback also needs to be more specific to groups of workers so that....

MT: (Cuts off MNT). I don't think we are supposed to change the grading system. We are supposed to say what's wrong based on the theory. (silence)

MNT: How about the issue of assigned vs. self set goals

FNT: (firm) These are assigned goals and assigned goals work.
MNT: Yeah but shouldn't the workers get a say in what is happening?

MT: (Sensing a conflict, MT closes the discussion) Okay, Okay. We recommend that workers be encouraged to accept the goals. Next case.

MNT: (is upset withdraws--no eye contact).

Script for Al the lawyer

MT: Let's talk about Al the lawyer

FNT: I am not surprised that Al is bored with his job. The most cities are flooded with lawyers and there isn't a lot of room for growth. I've heard that a lot of firms laid off people because of the recession..

MT: Really, I have thought about law school. Aren't there many jobs?

FNT: Its not as good as it used to be. Lawyers have created about as much business as they can. They have pretty much priced themselves out of the market. Some of them are hurting. (MT seems okay with and even interested in this story) (MNT is visibly upset) Its like getting an MBA. It used to be that an MBA was a special degree. Now everyone has an MBA and only the Harvard and Wharton People get jobs.

MT: (Notices MNT's frustration) Maybe we should get back to the case? (Pause to shuffle through materials) It seems like Al needs more variety in the kinds of things he does. He is still being treated like he is a clerk.

FNT: I can see where the guy might like a bigger office--Its something I would push for. That should make him happy. (Pause--no one speaks--dead silence with little activity--there is underlying tension) (MT--a bit frantic to ease the tension looks through materials and throws out a flurry of ideas)

MT: He should probably get more client contact. He should work on projects from start to finish. Yeah, this should get him more motivated. (tries to appease FNT) And maybe you are right, he should get a bigger office.

FNT: (Cuts to the heart of the issue--Looks to MNT) what are you thinking?

MT: (Tries to get MNT involved--nice, almost begging) We could sure use your help. What do you think?

MNT: I see this as a job design issue, too. The job characteristics theory seems relevant. The principles are skill variety, and feedback. Al needs more variety, and meeting with clients would give him more feedback.

FNT: Aren't people supposed to get more responsibility for planning their work?

MNT: Yeah, that's called vertical loading. It means that people should be given more responsibility for planning their work and setting goals for themselves.

FNT: I'm glad you finally spoke up (with a bit of sarcasm that is not taken well by MNT)

MT: (glosses over this conflict) Can we agree that Al should get more variety, work on jobs from start to finish, and have more client contact?

FNT: Yes

MNT: (Okay)

Script for Do You Measure Up?

MT: How about the Builders' Haven case. It seems like their CEO liked the idea of using a ruler but it wasn't working well in the organization. (Looks directly at MNT to get MNT back in the conversation) What do you think?

MNT: I kind of side with the CEO. I think the consensus in performance appraisal research was that rating format did not matter much. I don't see that we can prove the CEO wrong.

FNT: I think we have to do something. The status quo is not working. The workers and supervisors aren't satisfied with the system.

MNT: Okay, how about if we say that supervisors have to write a description for why they put the X where they did.

MT: From my understanding, I thought that it only doesn't matter what kind of format you use as long as you use some logical format. That's the problem, the ruler seems basically confusing.

FNT: I am on your side. (MT begins to ally with FNT) I think the research says that BARS or BOS do not matter. but you do need a logical format. And the format should be based on an analysis of an actual job. I think the ruler should at least have some labels describing what the different levels mean. You could keep the ruler concept but you would need to go through a more formal job analysis process

to develop some descriptions of behaviors at different points on the scale.

MNT: How about the issue that one end of the ruler goes to zero and supervisors don't know what performance level of zero means?

FNT: I though you liked the ruler (jokingly)?

MT: (eases over the conflict). Can we agree that the ruler can be maintained in concept, but that the organization proceed with a formal process of developing descriptions for each level of rating.

MNT: Also that supervisors write an explanation of what they mean and that the issue of the zero point is dealt with in some way.

MT: sounds good

FNT: How about job analysis--you forgot job analysis.

MT: Okay--job analysis

Script Merril's Department Store

MT: What do you think about the department store case?

FNT: Seems like management has sold this on the basis that it is a good deal for workers when it really was a way to save money and make it look progressive. The workers basically have lost out and aren't getting the pay they were promised.

MNT: I worked in a department store where they went to commission pay. We all started getting very competitive. After that, it just was not as fun to work there anymore. It sounds like the same thing is happening here. Sales reps are fighting over who gets the customer and who gets to sell the big ticket items.

MT: Oh yeah--What store did you work in?

MNT: I don't think that's important. (pause)

MT: Commission pay does not seem very consistent with customer service. If you pay people to make as many sales as possible, they aren't going to spend the time taking care of customers.

FNT: We do live in a capitalist system. If you don't perform, you don't deserve a paycheck. (MT has lost the group--now both are bringing in tangents)

MNT: I hate department stores with pushy sales people. I like it when they give me room to do my shopping. If I have question, I'll ask it.

MT: I don't know what theory applies to this? (dead silence again--neither MNT or FNT are working very hard. MT panics and starts with a flurry of ideas)

MT: Maybe its another example of goal setting--no goal acceptance? No feedback?? (continues, guessing) Maybe its just a reinforcement thing. People aren't being rewarded for customer service??? Maybe, the store should

have used the Scanlon plan. Scanlon was invented by a guy named Joe Scanlon. The workers are paid for performance but they have to participate in decisions about how the pay will be set up.

FNT: Takes the lead. What theories are relevant. What else can we think of besides pay?

MNT: Maybe its equity theory. People are putting in the same amount of effort but they are getting less pay.

FNT: That's better but it seems to me that this company better change its ways because its basically not bringing any customers in the door. Its products are bad and overpriced. How is a department store sales clerk going to do anything about that? If you ask me, the problem is with the company, not with their workers. No matter how much effort the reps put out they aren't going to make any sales.

MNT: Its hard for us to recommend that the store improve quality and lower its prices. I think we need to analyze this more from an organizational psychology perspective. It seems like a motivation and satisfaction issue.

MT: (Way off the Mark) Well there is something to be said for this idea. Maybe its not an issue of workers. Maybe the organization needs to adopt a new marketing strategy. This is probably a marketing niche issue (doesn't really know what he's talking about here)

MNT: It seems like whoever figured out that sales reps could make more was crazy. This needs to be reevaluated. Is management acknowledging that workers are getting paid less?

FNT: Maybe management did not really want to pay workers. They came up with a solution that would get more out of workers for less pay. This is a winning solution as far as management is concerned.

MNT: (frustrated) The employees are not "winning." (another stall in the group)

MT: (MT searching for something to say) Well that is true, you can't sell them something if it hits them hard in the pocket book (Silence)

FNT: (Again in the lead) What class theory is relevant? We talked about pay and equity theory. Have we exhausted our options? (everyone reads materials, looks puzzled, scratches head. time runs out.)

A,L,H Script

Male Non-Target (MNT)
Has good ideas but thin skinned--easily hurt by criticism. Becomes passive/aggressive when criticizes. Totally Withdraws or Lashes out and attacks his attacker. Plays devil's advocate for no reason other than to make others made. Problem Solving--Good ideas but
contributes little if he withdraws or shifts positions and argues for the sake of argument. Empathy--often does not maintain eye contact with others, shifts from being detached to attacking. Appreciates when MT asks for his opinion. Setting Directions and Goals--Does almost none of this. Motivation--doesn't say enough

to give the impression of being active. Says a few words like yes, o.k. that give the impression that he/she is not totally removed.

Female Non-Target (FNT)
is talkative, occasionally overpowering. Takes the group off on tangents but more often than not has very good ideas to contribute. Makes a definite move to take over the group when things are stalling at the end. Is successful in overtaking MT as the leader. Somewhat Opinionated, self-centered, some sarcastic humor gets to MNT--she is also good at laughing, joking to recover when others don't get it. Highly verbal, speaks with conviction. Problem Solving--Adds important details about theories brings in relevant experience from own life. Empathy--Talkative, but more chatty than warm. Makes some critical comments about others. Setting Direction and Goals--Gives people directions sometimes like "giving orders." Motivation--active, pushes for results.

Male Target (MT): Repeat of Previous Performance

Script

MT: let start off by introducing our selves. I'll start. I am.... I am a senior, communications major. My hobbies are ..... (goes on at some length, seems warm and friendly) Why don't you go next... (points to MNT)

MNT: I am ... (abrupt, short) I am a senior majoring in engineering. I don't really care much for psychology but I thought I could learn something to help me get along better with others at work.

FNT: I am ... (smiling) I <u>am</u> a psychology major (flaunting toward MNT). I want to get into graduate school in either I/O or clinical (bragging slightly) I have done independent studies in both.

MNT: (looks bored, then frustrated, wants to get moving)

MT: I'm thinking about grad school too. But my GRE score sucked. I have to retake it in the fall. I'm going to take the Kaplan course--I hope that helps.

FNT: That's hard. The prof for my independent study says my scores were good. I took the Kaplan course and it helped.

Good luck--maybe we should get going on this?

MNT: (interrupts after becoming visibly frustrated). Shouldn't we get to work? We only have 20 minutes

MT: O.K. (looks to FNT) I would like to get more info about grad schools.

Script for Murphy Rebuilt Engines

MT: (Shuffles through papers)

Let's start with the Murphy Rebuilt Engines case. It seems like management has been putting out a letter grade to inspire workers. The workers are not happy with it.

(Shuffles papers) We are supposed to discuss the case, decide which organizational psychology theory is relevant, and make recommendations.

What do you think?

FNT: This seems very typical in manufacturing.

There is a history of antagonism between labor and management. Just look at the Caterpillar strike!

I can see where they would resent evaluating their performance as if they are being graded in school. I could be downright insulting.

MT: (Nodding as if he agrees but then says) It seems like its more than that. The grade is not specifically related to the work they do. They only get the grade once a month and the grade is based on the whole plant.

MNT: Yeah! What if workers in one part of the plant are working poorly and cause a low grade, but it has little to do with your work group. Its pretty hard to know what that means for how hard you should work.

FNT: The workers do not know how the grade is calculated. The workers believe that the grade comes from management "passing a magic wand" over the results.

(exaggerates but with a sense of humor) management probably just makes up the grade depending on their mood for the day.

MNT: (doesn't get the joke) That's a bit much.

FNT: I was only kidding.
(pause)

MNT: Oh--well, management does seem to be criticizing the workers when the grade is bad and does not pay much attention to good grades.

MT: Maybe we should move on to discuss what organizational psychology theory applies to this problem.

(pause)

MNT: Maybe its an issue of pay. Since the people aren't paid extra for good letter grades, they are not being rewarded.

MT: I'm not sure that we can bring pay into this.

(Pause)

Maybe we can apply goal setting???.

FNT: (interrupts and says the following as if quoting from a text book) Goal setting says that people work the hardest when they accept specific and difficult goals and get feedback.

Have these goals been accepted, Have they been difficult and specific?

(Pause)

MNT: Well, the workers have not accepted this as a goal, even though they are getting feedback

MT: (MT offers an idea that is wrong) It sounds like the goal is difficult since they are not meeting the grade every time.

FNT: We don't know that. Why don't we recommend that management find a way to make the workers accept the goal. They should continue to give feedback as they have been doing.

MNT: But the feedback also needs to be more specific to groups of workers so that....

MT: (Cuts off MNT). I don't think we are supposed to change the grading system. We are supposed to say what's wrong based on the theory.

(silence)

MNT: How about the issue of assigned vs. self set goals

FNT: (firm) These are assigned goals and assigned goals work.

MNT: That's not true!

MT: (Sensing a conflict, MT closes the discussion) Okay, Okay. We recommend that workers be encouraged to accept the goals. Next case.

MNT: (is visibly upset withdraws that his/her points were not considered by the group--no eye contact, no speech for the entirety of Al the lawyer 2).

Script for Al the lawyer

MT: Let's talk about Al the lawyer

FNT: I am not surprised that Al is bored with his job. The most cities are flooded with lawyers and there isn't a lot of room for growth.

I've heard that a lot of firms laid off people because of the recession..

MT: Really, I have thought about law school. Aren't there many jobs?

FNT: Its not as good as it used to be. Lawyers have created about as much business as they can. They have pretty much priced themselves out of the market. Some of them are hurting.

(MT seems okay with and even interested in this story)

(MNT is visibly upset)

Its like getting an MBA. It used to be that an MBA was a special degree. Now everyone has an MBA and only the Harvard and Wharton People get jobs.

MT: (Notices MNT's frustration) Maybe we should get back to the case?

(Pause to shuffle through materials) It seems like Al needs more variety in the kinds of things he does. He is still being treated like he is a clerk.

FNT: I can see where the guy might like a bigger office--Its something I would push for. That should make him happy. (Pause--no one speaks--dead silence with little activity--there is underlying tension) (MT--a bit frantic to ease the tension looks through materials and throws out a flurry of ideas)

MT: He should probably get more client contact. He should work on projects from start to finish. Yeah, this should get him more motivated. (tries to appease FNT) And maybe you are right, he should get a bigger office.

FNT: (Cuts to the heart of the issue--Looks to MNT) what are you thinking?

MT: (Tries to get MNT involved--nice, almost begging) We could sure use your help. What do you think?

MNT: I'm tired of listening to your crap! (points to FNT) (MNT and FNT exchange words)

FNT: It sure beats saying nothing!

MT: (glosses over this conflict) Can we agree that Al should get more variety, work on jobs from start to finish, and have more client contact?

FNT: Yes

MNT: (Grunts yes, but clearly doesn't mean it)

Script for Do You Measure Up?

MT: How about the Builders' Haven case. It seems like their CEO liked the idea of using a ruler but it wasn't working well in the organization. (Looks directly at MNT to get MNT back in the conversation). What do you think?

MNT: Well I doubt this will be a popular answer but I side with the CEO. I think the consensus in performance appraisal research is that rating format does not matter much (this is wrong). I don't see that we can prove the CEO wrong.

FNT: I think we have to do something. The status quo is not working. The workers and supervisors aren't satisfied with the system.

MNT: Okay, how about if we say that supervisors have to write a description for why they put the X where they did.

MT: From my understanding, I thought that it only doesn't matter what kind of format you use as long as you use some logical format. That's the problem, the ruler seems basically confusing.

FNT: I am on your side. (MT begins to ally with FNT). I think the research says that BARS or BOS do not matter. but you do need a logical format. And the format should be based on an analysis of an actual job. I think the ruler should

at least have some labels describing what the different levels mean. You could keep the ruler concept but you would need to go through a more formal job analysis process to develop some descriptions of behaviors at different points on the scale.

MNT: (Switches his/her position for the sake of argument) Well I thought that just a minute ago you said the ruler was a bad idea.
(Devil's advocate--gloating because the others have been wrong) And, how about the issue that one end of the ruler goes to zero and supervisors don't know what performance level of zero means?

FNT: I though you liked the ruler (jokingly)?

MT: (eases over the conflict). Can we agree that the ruler can be maintained in concept, but that the organization proceed with a formal process of developing descriptions for each level of rating.

MNT: Also that supervisors write an explanation of what they mean and that the issue of the zero point is dealt with in some way.

MT: sounds good

FNT: How about job analysis--you forgot job analysis.

MT: Okay--job analysis

Script for Merril's Department Store

MT: What do you think about the department store case?

FNT: Seems like management has sold this on the basis that it is a good deal for workers when it really was a way to save money and make it look progressive. The workers basically have lost out and aren't getting the pay they were promised.

MNT: I worked in a department store where they went to commission pay. We all started getting very competitive. After that, it just was not as fun to work there anymore. It sounds like the same thing is happening here. Sales reps are fighting over who gets the customer and who gets to sell the big ticket items.

MT: Oh yeah--What store did you work in?

MNT: I don't think that's important. (pause)

MT: Commission pay does not seem very consistent with customer service. If you pay people to make as many sales as possible, they aren't going to spend the time taking care of customers.

FNT: We do live in a capitalist system. If you don't perform, you don't deserve a paycheck. (MT has lost the group--now both are bringing in tangents).

MNT: I hate department stores with pushy sales people. I like it when they give me room to do my shopping. If I have question, I'll ask it.

MT: I don't know what theory applies to this? (dead silence again--neither MNT or FNT are working very hard. MT panics and starts with a flurry of ideas)

MT: Maybe its another example of goal setting--no goal acceptance? No feedback?? (continues, guessing) Maybe its just a reinforcement thing. People aren't being rewarded for customer service??? Maybe, the store should have used the Scanlon plan. Scanlon was invented by a guy named Joe Scanlon. The workers are paid for performance but they have to participate in decisions about how the pay will be set up.

FNT: Takes the lead. What theories are relevant. What else can we think of besides pay?

MNT: Maybe its equity theory. People are putting in the same amount of effort but they are getting less pay.

FNT: That's better but Can't we get off the issue of pay? It seems to me that this company better change its ways because its basically not bringing any customers in the door. Its products are bad and overpriced. How is a department store sales clerk going to do anything about that? If you ask me, the problem is with the company, not with their workers. No matter how much effort the reps put out they aren't going to make any sales.

MNT: (sarcastically). Oh, that's brilliant. Our recommendation can't be that the store improve quality and lower its prices--that has nothing to do with organizational psychology.

MT: (Way off the Mark) Well there is something to be said for this idea. Maybe its not an issue of workers. Maybe the organization needs to adopt a new marketing strategy. This is probably a marketing niche issue (doesn't really know what he's talking about here)

MNT: It seems like whoever figured out that sales reps could make more was crazy. Some disciplinary action should be taken with the bean counters that came up with these ideas.

FNT: Maybe management did not really want to pay workers. They came up with a solution that would get more out of workers for less pay. This is a winning solution as far as management is concerned.

MNT: (frustrated) its not an issue of selling it to the employees. (another stall in the group)

MT: (MT searching for something to say) Well that is true, you can't sell them something if it hits them hard in the pocket book. (Silence)

FNT: (Again in the lead) What class theory is relevant? We talked about pay and equity theory. Have we exhausted our options? (everyone reads materials, looks puzzled, scratches head. time runs out.)

A,H,H Script

Male, Non-target (MNT)
Has good ideas, quiet but generally cooperative. Some defensiveness when criticized. Problem Solving--Good ideas. Empathy--Cooperative, agreeable, not extremely warm. Appreciates when MT asks for his opinion. Setting Directions and Goals--Only suggests ideas, never directs the group. Motivation--Active. Even when quiet nods a lot as if listening.

Female Non-target (FNT)
Repeats previous High Performance

Male, Target (MT)
Repeats previous performance

Script

MT: let start off by introducing our selves. I'll start. I am.... I am a senior, communications major. My hobbies are ..... (goes on at some length, seems warm and friendly) Why don't you go next... (points to MNT)

MNT: I am ... (abrupt, short) I am a senior majoring in engineering. I haven't had much psychology but I thought I could learn something to help me get along better with others at work.

FNT: I am ... (smiling) I <u>am</u> a psychology major (flaunting toward MNT). I want to get into graduate school in either I/O or clinical (bragging slightly) I have done independent studies in both.

MNT: (shakes head in a bit of disbelief at the bragging)

MT: I'm thinking about grad school too. But my GRE score sucked. I have to retake it in the fall. I'm going to take the Kaplan course--I hope that helps.

FNT: That's hard. The prof for my independent study says my scores were good. I took the Kaplan course and it helped. Good luck--maybe we should get going on this?

MNT: (interrupts after becoming visibly frustrated). Yeah. We only have 20 minutes

MT: O.K. (looks to FNT) I would like to get more info about grad schools.

Scenario for Murphy Rebuilt Engines

MT: (Shuffles through papers) Let's start with the Murphy Rebuilt Engines case. It seems like management has been putting out a letter grade to inspire workers. The workers are not happy with it. (Shuffles papers) We are supposed to discuss the case, decide which organizational psychology theory is relevant, and make recommendations. What do you think?

FNT: This seems very typical in manufacturing. There is a history of antagonism between labor and management. Just look at the Caterpillar strike! I can see where they would resent evaluating their performance as if they are being graded in school. I could be downright insulting.

MT: (Nodding as if he agrees but then says) It seems like its more than that. The grade is not specifically related to the work they do. They only get the grade once a month and the grade is based on the whole plant.

MNT: Yeah! What if workers in one part of the plant are working poorly and cause a low grade, but it has little to do with your work group. Its pretty hard to know what that means for how hard you should work.

FNT: The workers do not know how the grade is calculated. The workers believe that the grade comes from management "passing a magic wand" over the results. (exaggerates but with a sense of humor) management probably just makes up the grade depending on their mood for the day.

MNT: (doesn't get the joke) That's a bit much.

FNT: I was only kidding.
(pause)

MNT: Oh--well, management does seem to be criticizing the workers when the grade is bad and does not pay much attention to good grades.

MT: Maybe we should move on to discuss what organizational psychology theory applies to this problem. (pause)

MNT: Maybe its an issue of pay. Since the people aren't paid extra for good letter grades, they are not being rewarded.

MT: I'm not sure that we can bring pay into this. (Pause) Maybe we can apply goal setting???.

FNT: (interrupts and says the following as if quoting from a text book) Goal setting says that people work the hardest when they accept specific and difficult goals and get feedback. Have these goals been accepted, Have they been difficult and specific? (Pause)

MNT: Well, the workers have not accepted this as a goal, even though they are getting feedback

MT: (MT offers an idea that is wrong) It sounds like the goal is difficult since they are not meeting the grade every time.

FNT: We don't know that. Why don't we recommend that management find a way to make the workers accept the goal. They should continue to give feedback as they have been doing.

MNT: But the feedback also needs to be more specific to groups of workers so that....

MT: (Cuts off MNT). I don't think we are supposed to change the grading system. We are supposed to say what's wrong based on the theory. (silence)

MNT: How about the issue of assigned vs. self set goals

FNT: (firm) These are assigned goals and assigned goals work.

MNT: Yeah but shouldn't the workers get a say in what is happening?

MT: (Sensing a conflict, MT closes the discussion) Okay, Okay. We recommend that workers be encouraged to accept the goals. Next case.

MNT: (is upset withdraws--no eye contact).

Script for Al the lawyer

MT: Let's talk about Al the lawyer

FNT: I am not surprised that Al is bored with his job. The most cities are flooded with lawyers and there isn't a lot of room for growth. I've heard that a lot of firms laid off people because of the recession..

MT: Really, I have thought about law school. Aren't there many jobs?

FNT: Its not as good as it used to be. Lawyers have created about as much business as they can. They have pretty much priced themselves out of the market. Some of them are hurting. (MT seems okay with and even interested in this story) (MNT is visibly upset) Its like getting an MBA. It used to be that an MBA was a special degree. Now everyone has an MBA and only the Harvard and Wharton People get jobs.

MT: (Notices MNT's frustration) Maybe we should get back to the case? (Pause to shuffle through materials) It seems like Al needs more variety in the kinds of things he does. He is still being treated like he is a clerk.

FNT: I can see where the guy might like a bigger office--Its something I would push for. That should make him happy. (Pause--no one speaks--dead silence with little activity--there is underlying tension) (MT--a bit frantic to ease the tension looks through materials and throws out a flurry of ideas)

MT: He should probably get more client contact. He should work on projects from start to finish. Yeah, this should get him more motivated. (tries to appease FNT) And maybe you are right, he should get a bigger office.

FNT: (Cuts to the heart of the issue--Looks to MNT) what are you thinking?

MT: (Tries to get MNT involved--nice, almost begging) We could sure use your help. What do you think?

MNT: I see this as a job design issue, too. The job characteristics theory seems relevant. The principles are skill variety, and feedback. Al needs more variety, and meeting with clients would give him more feedback.

FNT: Aren't people supposed to get more responsibility for planning their work?

MNT: Yeah, that's called vertical loading. It means that people should be given more responsibility for planning their work and setting goals for themselves.

FNT: I'm glad you finally spoke up (with a bit of sarcasm that is not taken well by MNT)

MT: (glosses over this conflict) Can we agree that Al should get more variety, work on jobs from start to finish, and have more client contact?

FNT: Yes

MNT: (Okay)

Scenario for Do You Measure Up?

MT: How about the Builders' Haven case. It seems like their CEO liked the idea of using a ruler but it wasn't working well in the organization. (Looks directly at MNT to get MNT back in the conversation) What do you think?

MNT: I kind of side with the CEO. I think the consensus in performance appraisal research was that rating format did not matter much. I don't see that we can prove the CEO wrong.

FNT: I think we have to do something. The status quo is not working. The workers and supervisors aren't satisfied with the system.

MNT: Okay, how about if we say that supervisors have to write a description for why they put the X where they did.

MT: From my understanding, I thought that it only doesn't matter what kind of format you use as long as you use some logical format. That's the problem, the ruler seems basically confusing.

FNT: I am on your side. (MT begins to ally with FNT) I think the research says that BARS or BOS do not matter. but you do need a logical format. And the format should be based on an analysis of an actual job. I think the ruler should at least have some labels describing what the different levels mean. You could keep the ruler concept but you would need to go through a more formal job analysis process to develop some descriptions of behaviors at different points on the scale.

MNT: How about the issue that one end of the ruler goes to zero and supervisors don't know what performance level of zero means?

FNT: I though you liked the ruler (jokingly)?

MT: (eases over the conflict). Can we agree that the ruler can be maintained in concept, but that the organization proceed with a formal process of developing descriptions for each level of rating.

MNT: Also that supervisors write an explanation of what they mean and that the issue of the zero point is dealt with in some way.

MT: sounds good

FNT: How about job analysis--you forgot job analysis.

MT: Okay--job analysis

Script for Merril's Department Store

MT: What do you think about the department store case?

FNT: Seems like management has sold this on the basis that it is a good deal for workers when it really was a way to save money and make it look progressive. The workers basically have lost out and aren't getting the pay they were promised.

MNT: I worked in a department store where they went to commission pay. We all started getting very competitive. After that, it just was not as fun to work there anymore. It sounds like the same thing is happening here. Sales reps are fighting over who gets the customer and who gets to sell the big ticket items.

MT: Oh yeah--What store did you work in?

MNT: I don't think that's important. (pause)

MT: Commission pay does not seem very consistent with customer service. If you pay people to make as many sales as possible, they aren't going to spend the time taking care of customers.

FNT: We do live in a capitalist system. If you don't perform, you don't deserve a paycheck. (MT has lost the group--now both are bringing in tangents)

MNT: I hate department stores with pushy sales people. I like it when they give me room to do my shopping. If I have question, I'll ask it.

MT: I don't know what theory applies to this? (dead silence again--neither MNT or FNT are working very hard. MT panics and starts with a flurry of ideas)

MT: Maybe its another example of goal setting--no goal acceptance? No feedback?? (continues, guessing). Maybe its just a reinforcement thing. People aren't being rewarded for customer service??? Maybe, the store should have used the Scanlon plan. Scanlon was invented by a guy named Joe Scanlon. The workers are paid for performance but they have to participate in decisions about how the pay will be set up.

FNT: Takes the lead. What theories are relevant. What else can we think of besides pay?

MNT: Maybe its equity theory. People are putting in the same amount of effort but they are getting less pay.

FNT: That's better but it seems to me that this company better change its ways because its basically not bringing any customers in the door. Its products are bad and overpriced. How is a department store sales clerk going to do anything about that? If you ask me, the problem is with the company, not with their workers. No matter how much effort the reps put out they aren't going to make any sales.

MNT: Its hard for us to recommend that the store improve quality and lower its prices. I think we need to analyze this more from an organizational psychology perspective. It seems like a motivation and satisfaction issue.

MT: (Way off the Mark) Well there is something to be said for this idea. Maybe its not an issue of workers. Maybe the organization needs to adopt a new marketing strategy. This is probably a marketing niche issue (doesn't really know what he's talking about here)

MNT: It seems like whoever figured out that sales reps could make more was crazy. This needs to be reevaluated. Is management acknowledging that workers are getting paid less?

FNT: Maybe management did not really want to pay workers. They came up with a solution that would get more out of workers for less pay. This is a winning solution as far as management is concerned.

MNT: (frustrated) The employees are not "winning." (another stall in the group)

MT: (MT searching for something to say) Well that is true, you can't sell them something if it hits them hard in the pocket book (Silence)

FNT: (Again in the lead) What class theory is relevant? We talked about pay and equity theory. Have we exhausted our options? (everyone reads materials, looks puzzled, scratches head. time runs out.)

APPENDIX F: Pilot 1 and Final Questionnaires

Pilot 1 Questionnaire

Based on your consideration of the scoring guidelines, rate the Target assessee whom you were assigned to watch.

1. Problem solving	1 Low	2	3	4	5	6	7 High
2. Empathy	1 Low	2	3	4	5	6	7 High
3. Setting Direction & Goals for the Group	1 Low	2	3	4	5	6	7 High
4. Motivation	1 Low	2	3	4	5	6	7 High

Please answer the following questions relative to your <u>current</u> job. If you are unemployed, answer the questions relative to the most recent job you held.

How satisfied are you with each of the following aspects of your job?

5. The amount of pay and fringe benefits I receive

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

6. The degree of respect I get from my boss

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

7. The feelings of worth and accomplishment I get from my job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

8. The amount of pay and fringe benefits I receive

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

9. The degree of respect I get from my boss

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

10. The feelings of worth and accomplishment I get from my job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

11. The degree to which I am paid fairly in this job

1	2	3	4	5
Dissatisfied	Slightly	Neutral Dissatisfied	Slightly	Satisfied Satisfied

12. The amount of independent thought and action I can exercise in this job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

13. The amount of challenge in my job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

14. The overall quality of the supervision I receive

1	2	3	4	5
Dissatisfied	Slightly	Neutral Dissatisfied	Slightly	Satisfied Satisfied

15. The degree to which I am paid fairly in this job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

16. The amount of independent thought and action I can exercise in this job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

17. The amount of challenge in my job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

18. The overall quality of the supervision I receive

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

Please respond to the statements based on your perceptions of how well this group performed on this task, based on your understanding of how groups usually perform on this task. Please indicate your response on the answer sheet.

19. How do you rate this group in terms of the number of cases they finished?

1	2	3	4	5
Far Below Average	Below Average	Average	Above Average	Far Above Average

20. How would you rate the overall performance of this group?

1	2	3	4	5
Far Below Average	Below Average	Average	Above Average	Far Above Average

21. All things considered, this group was

1	2	3	4	5
Far Below Average	Below Average	Average	Above Average	Far Above Average

Rate the other <u>Male</u> group member based on your recollection of his behavior and the scoring guidelines

22. Problem solving	1 Low	2	3	4	5	6	7 High
23. Empathy	1 Low	2	3	4	5	6	7 High

24.	Setting Direction & Goals	1 Low	2	3	4	5	6	7 High
25.	Motivation	1 Low	2	3	4	5	6	7 High

Rate the other Female group member based on your recollection of her behavior and the scoring guidelines

26.	Problem solving	1 Low	2	3	4	5	6	7 High
27.	Empathy	1 Low	2	3	4	5	6	7 High
28.	Setting Direction & Goals	1 Low	2	3	4	5	6	7 High
29.	Motivation	1 Low	2	3	4	5	6	7 High

Please respond to the following statements based on how you perceived the relationships between the members of the group. "The Target" refers to the person you were assigned to watch and rate. "The Other Group Members" refers to the two persons in the group that you were not assigned to rate.

30. The Target had important information that the Other Group Members did not have.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

31. The Other Group members had important information that the Target did not have.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

32. The Target could have achieved exactly the same or more without the contributions of the Other Group Members.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

33. The other members of the group were highly dependent on the Target.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

34. The Target was highly dependent on the Other Group Members

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

35. The Other Group Members could have achieved exactly the same or more without the contributions of the Target.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

Please respond to the following statements about the people in the group exercise. "The Target" refers to the person you were assigned to watch and rate. "The Other Group Members" refers to the persons in the group that you were not assigned to rate. Indicate your response on the answer sheet.

36. This group's accomplishments were clearly due to the ability of the Target.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

37. This group's accomplishments were clearly due to the ability of the Other Group Members.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

38. Compared to the Other Group Members, the Target contributed the most to the accomplishments of the group.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

39. The Target was an obstacle to the group's performance.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

40. The Other Group Members were an obstacle to the group's performance.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

41. The Target was solely responsible for this group's success/failure.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

42. The Other Group Members were solely responsible for this group's success/failure.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

43. The Target would have accomplished more if the others in the group had put in more effort.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

44. The others in the group would have accomplished more if the Target had put in more effort.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

Please respond to the following statements. "The Target" refers to the person you were assigned to observe and rate. Indicate you response on the answer sheet.

45. I considered the Target an excellent leader

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

46. The Target exhibited a great deal of leadership

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

47. The group was led well by the Target

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

48. I considered the Target a poor leader

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

49. The group was led poorly by the Target

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

First, indicate whether you considered the Target's performance similar to or not similar from other group members, then, follow the directions to the appropriate item and indicate the extent to which the Target was similar or not similar.

50. How did you view the Target's Problem Solving Skills relative to the Other Group Members?
    a. Similar--go to # 51 b. Not Similar Go to # 52

51. If you marked <u>Similar</u> was the Target's <u>Problem Solving</u> behavior
    a. Similar but Slightly Worse?
    b. Moderately Similar?
    c. Highly Similar?
    d. Similar but Slightly Better?

52. If you marked <u>Not Similar</u> was the Target's <u>Problem Solving</u> behavior
    a. Much Worse
    b. Worse
    c. Better
    d. Much Better

53. How did you view the Target's <u>Empathy</u> relative to the Other Group Members?
    a. Similar--Go to # 54 b. Not Similar--Go to #55

54. If you marked Similar was the Target's Empathy
    a. Similar but Slightly Worse?
    b. Moderately Similar?
    c. Highly Similar?
    d. Similar but Slightly Better?

55. If you marked Not Similar was the Target's Empathy behavior
    a. Much Worse
    b. Worse
    c. Better
    d. Much Better

56. How did you view the Target's Setting Directions & Goals relative to the Other Group Members?
    a. Similar--Go to # 57 b. Not Similar--Go to # 58

57. If you marked Similar was the Target's Setting Directions & Goals Behavior
    a. Similar but Slightly Worse?
    b. Moderately Similar?
    c. Highly Similar?
    d. Similar but Slightly Better?

58. If you marked Not Similar was the Target's Setting Directions & Goals Behavior
    a. Much Worse?
    b. Worse?
    c. Better?
    d. Much Better?

59. How did you view the Target's Motivation relative to the Other Group Members?
    a. Similar--Go to # 60 b. Not Similar--Go to # 61

60. If you marked Similar was the Target's Motivation
    a. Similar but Slightly Worse?
    b. Moderately Similar?
    c. Highly Similar?
    d. Similar but Slightly Better?

61. If you marked Not Similar was the Target's Motivation
    a. Much Worse?
    b. Worse?
    c. Better?
    d. Much Better?

Final Questionnaire

Based on your consideration of the scoring guidelines, rate the Target assessee whom you were assigned to watch. Please indicate your answers on the answer sheet and do not write on these forms.

1.	Problem solving	1 Low	2	3	4	5	6	7 High
2.	Managing Interpersonal Relationships	1 Low	2	3	4	5	6	7 High
3.	Initiative/ Assertiveness	1 Low	2	3	4	5	6	7 High

Please answer the following questions relative to your <u>current</u> job. If you are unemployed, answer the questions relative to the <u>most recent</u> job you held.

4. Its hard for me to care very much about whether the work gets done right on this job.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

5. My opinion of myself goes up when I do this job well

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

6. I am constantly feel as sense of fulfillment and self-actualization in my work.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

7. Most of the things I do on this job seem very trivial

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

8. I can't imagine having a job that is more satisfying that the one I have now

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

9. This job is my ideal career job. Its all I ever wanted in life

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

How satisfied are you with each of the following aspects of your job?

10. The degree of respect I get from my boss

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

11. The feelings of worth and accomplishment I get from my job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

12. The degree to which I am paid an ideal salary in this job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

13. The amount of independent thought and action I can exercise in this job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

14. The amount of challenge in my job

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

15. The overall quality of the supervision I receive

1	2	3	4	5
Dissatisfied	Slightly Dissatisfied	Neutral	Slightly Satisfied	Satisfied

Please respond to the statements based on your perceptions of how well this group performed on this task, based on what you learned during the training about typical performance for groups on this task. Please indicate your response on the answer sheet.

16. How do you rate this group in terms of the number of cases they finished?

1	2	3	4	5
Far Below Average	Below Average	Average	Above Average	Far Above Average

17. How productive was this group given what you learned during the training about how many cases groups typically finish?

1	2	3	4	5
Far Below Average	Below Average	Average	Above Average	Far Above Average

18. Based on what you have learned about the number of cases that groups typically finish, how do you rate this group's level of achievement

1	2	3	4	5
Far Below Average	Below Average	Average	Above Average	Far Above Average

Rate the other Male group member based on your recollection of his behavior and the scoring guidelines

19.	Problem solving	1 Low	2	3	4	5	6	7 High
20.	Managing Interpersonal Relationships	1 Low	2	3	4	5	6	7 High
21.	Initiative/ Assertiveness	1 Low	2	3	4	5	6	7 High

Rate the other Female group member based on your recollection of her behavior and the scoring guidelines

22.	Problem solving	1 Low	2	3	4	5	6	7 High
23.	Managing Interpersonal Relationships	1 Low	2	3	4	5	6	7 High
24.	Initiative/ Assertiveness	1 Low	2	3	4	5	6	7 High

Please respond to the following statements based on what you learned during the training about how the design of the cases influenced teamwork. "The Target" refers to the person you were assigned to watch and rate. "The Other Group Members" refers to the two persons in the group that you were not assigned to rate.

25. Sharing information among group members was important for accurate problem solving because all of the information was not given to all of the group members.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

26. Sharing information among group members was not important for accurate problem solving because all of the information was given to all of the group members.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

27. Because of the distribution of information to members of the group, group members had to depend on each other to solve problems accurately.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

28. Group members did not have to depend on each other at all; Any one of the group members could have solved the problem just as completely and accurately without access to others' information.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

29. The Target was highly dependent on the Other Group Members.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

30. The group members each had sufficient information to solve the case on his/her own.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

31. The Other Group members had important information that the Target did not have.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

32. Group members needed to depend on each other to provide important information about the case.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

33. Since all team members had the same information about the cases, it didn't matter whether the group worked together as a team or not.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

Please respond to the following statements about the people in the group exercise. "The Target" refers to the person you were assigned to watch and rate. "The Other Group Members" refers to the persons in the group that you were not assigned to rate. Indicate your response on the answer sheet.

34. This group's accomplishments were clearly due to the ability of the Target.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

35. This group's accomplishments were clearly due to the ability of the Other Group Members.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

36. Compared to the Other Group Members, the Target contributed the most to the accomplishments of the group.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

37. Compared to the Other Group Members, the Target was most responsible for the group's success/failure.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

38. The Target was an obstacle to the group's performance.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

39. The Other Group Members were an obstacle to the group's performance.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

40. The others in the group would have accomplished more if the Target had put in more effort.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

41. The Others in the group would have accomplished more if the Target had greater ability.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

Please respond to the following statements. "The Target" refers to the person you were assigned to observe and rate. Indicate you response on the answer sheet.

42. I considered the Target an excellent leader.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

43. The Target exhibited a great deal of leadership.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

44. The group was led well by the Target.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

45.I considered the Target a poor leader.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

46. The group was led poorly by the Target.

1	2	3	4	5
Disagree	Disagree Somewhat	Neither Agree nor Disagree	Agree Somewhat	Agree

The following questions occur in pairs.

For the first item, indicate whether you considered the Target's performance similar to or not similar from other group members.

For the second item, respond with options a,b,c,or d if you thought that the Target's performance was similar to the other group members and respond with options e,f,g, or h if you thought that the target's performance was not similar to other group members.

Be sure to mark only one response to each item

47. How did you view the Target's Problem Solving Skills relative to the Other Group Members?
a. Similar b. Not Similar

48. If you marked Similar was the Target's Problem Solving behavior
48a. Similar but Slightly Worse?
48b. Moderately Similar?
48c. Highly Similar?
48d. Similar but Slightly Better?

If you marked Not Similar was the Target's Problem Solving behavior
48e. Much Worse
48f. Worse
48g. Better
48h. Much Better

49. How did you view the Target's Managing Interpersonal Relationships behavior relative to the Other Group Members?
a. Similar b. Not Similar

50. If you marked Similar was the Target's Managing Interpersonal Relationships behavior
50a. Similar but Slightly Worse?
50b. Moderately Similar?
50c. Highly Similar?
50d. Similar but Slightly Better?

If you marked Not Similar was the Target's Managing Interpersonal Relationships behavior
50e. Much Worse
50f. Worse
50g. Better
50h. Much Better

51. How did you view the Target's Initiative/Assertiveness relative to the Other Group Members?

a. Similar b. Not Similar

52. If you marked Similar was the Target's Initiative/Assertiveness Behavior

52a. Similar but Slightly Worse?
52b. Moderately Similar?
52c. Highly Similar?
52d. Similar but Slightly Better?

If you marked Not Similar was the Target's Initiative/Assertiveness Behavior

52e. Much Worse?
52f. Worse?
52g. Better?
52h. Much Better?

LIST OF REFERENCES

Ackerman, P.L. (1987). Individual differences in skill learning: An integration of psychometric and information processing perspectives. Psychological Bulletin, 102, 3-27.

Alba, J.W., & Hasher, L. (1983). Is memory schematic?. Psychological Bulletin, 93, 203-231.

Alwin, D.F, & Hauser. R.M. (1975). The decomposition of effects in path analysis. American Sociological Review, 40, 37-47.

Archambeau, D. J. (1979). Relationship among skill ratings assigned in an assessment center. Journal of Assessment Center Technology, 2, 7-20.

Arvey, R.D., Strickland, W., Drauden, G., & Martin, C. (1990). Motivation components of test taking. Personnel Psychology, 43, 695-716.

Asch, S.E. (1946). Forming impressions of personality. Journal of Abnormal and Social Psychology, 41, 1230-1240.

Balzer, W. K. (1986). Biases in the recording of performance-related information: The effects of initial impression and centrality of the appraisal task. Organizational Behavior and Human Decision Processes, 37, 329-347.

Bartlett, F.C. (1932). Remembering: A study in experimental and social psychology. Cambridge, England: Cambridge University Press.

Berkowitz, L. (1960). The judgmental process in personality functioning. Psychological Review, 67, 130-142.

Bernardin, H.J. & Pence, E.C. (1980). Effects of rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.

Bernardin, H.J., & Walter, S.C. (1977). Effects of rater training and diary-keeping on psychometric error in ratings. Journal of Applied Psychology, 62, 64-69.

Binning, J.F., & Lord, R.G. (1980). Boundary conditions for performance cue effects on group process ratings: Familiarity versus type of feedback. Organizational Behavior and Human Decision Processes, 26, 115-130.

Bray, D. W., & Grant, D. L. (1966). The assessment center in the measurement of potential for business management. Psychological Monographs, 80, (17, Whole No. 625).

Butler, S. (1989). Context effects in a group interaction setting. Unpublished doctoral dissertation, Rice University, Houston, TX.

Bycio, P., Alvares, K. M., Hahn, J. (1987). Situational specificity in assessment center ratings: A confirmatory factor analysis. Journal of Applied Psychology, 72, 463-474.

Calder, B.J. (1977). An attribution theory of leadership. In B.M. Staw & G.R. Salancik (Eds.) New directions in organization behavior. Chicago: St. Clair Press.

Cantor, N., & Mischel, W. (1977). Traits as prototypes: Effects on recognition memory. Journal of Personality and Social Psychology, 1, 38-48.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Cohen, J., & Cohen. P. (1977). Applied multiple regression/correlation analysis for the behavioral sciences. Hillsdale, NJ: Lawrence Erlbaum Associates.

Collins, B.E., & Guetzkow, H.A. (1964). A social psychology of group processes for decision-making. New York: Wiley

Cooper, W.H. (1981). Conceptual similarity as a source of illusory halo in job performance ratings. Journal of Applied Psychology, 66, 302-307.

DeCotiis, T.A., & Petis, A. (1978). The performance appraisal process: A model and some testable hypotheses. Academy of Management Review, 21, 635-646.

DeNisi, A.S., Cafferty, T.P., & Meglino, B.M. (1984). A cognitive view of the performance appraisal process: A model of research propositions. Organizational Behavior and Human Decision Processes, 33, 360-396.

DeNisi, A.S., & Pritchard, R.D. (1977). Implicit theories of performance as artifacts in survey research: A replication and extension. Organizational Behavior and Human Decision Processes, 21, 358-366.

Denisi, A.S., Robbins, T., & Cafferty, T. (1989). Organization of information used for performance appraisals: Role of diary-keeping. Journal of Applied Psychology, 74, 124-129.

Deutsch, M. (1949). An experimental study of the effects of cooperation and competition upon group process. Human Relations, 10, 307-318.

Deutch, M. (1951). Task structure and group process. American Psychologist, 6, 324-325.

Downey, H.K., Chacko, T., & McElroy, J.C. (1979). Attributions of the "causes" of performance: A constructive, quasi-longitudinal replication of the Staw study (1975). Organizational Behavior and Human Decision Processes, 24, 287-299.

Dreher, G. F. & Sackett, P. R. (1981). Some problems with applying content validity evidence to assessment center procedures. Academy of Management Review, 6, 551-560.

Eden, D., & Leviatan, U. (1975). Implicit leadership theory as a determinant of the factor structure underlying supervisory behavior scales. Journal of Applied Psychology, 60, 736-741.

Endler, N.S., & Magnusson, D. (1976). Toward an interactional psychology of personality. Psychological Bulletin, 83, 956-974.

Farris, G.F., & Lim, F.G. (1969). Effects of performance on leadership, cohesiveness, influence, satisfaction, and subsequent performance. Journal of Applied Psychology, 53, 490-497.

Feldman, J.M. (1981). Beyond attribution theory: Cognitive processes in performance appraisal. Journal of Applied Psychology, 66, 127-148.

Fiske, S.T. & Neuberg, S.L. (1990). A continuum of impression formation, from category-based to individuating processes: Influences of information and motivation on attention and interpretation. Advances in Experimental Social Psychology (Vol. 21). Academic Press.

Fiske, S.T., & Taylor, S.E., (1991). Social Cognition (2nd. ed.). New York: McGraw-Hill.

Fleishman, E.A. & Harris, E.F. (1962). Patterns of leadership behavior related to employee grievances and turnover. Personnel Psychology, 15, 43-56.

Foti, R.J., & Lord, R.G. (1987). Prototypes and scripts: The effects of alternative methods of processing information on rating accuracy. Organizational Behavior and Human Decision Processes, 39, 318-340

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C. III, & Bentson, C. (1987). Meta-analyses of assessment center validity. Journal of Applied Psychology, [Monograph] 72, 493-511.

Gaugler. B.B., & Rudolph, A.S. (1992). The influence of assessee performance variation on assessors' judgments. Personnel Psychology, 45, 77-98.

Gaugler, B. B., & Thornton, G. C. III (1989). Number of assessment center dimensions as a determinant of assessor accuracy. Journal of Applied Psychology, 74, 611-618.

Gersick, C.J.G. (1977). Time and transition in work teams: Toward a new model of group development. Academy of Management Journal, 31, 9-41.

Goodstadt, B., & Kipnis, D. (1970). Situational influences on the use of power. Journal of Applied Psychology, 54, 201-207.

Green, S.G., & Mitchell, T.R. (1979). Attributional processes of leaders in leader-member interactions. Organizational Behavior and Human Decision Processes, 23, 429-458.

Guzzo, R.A., Wagner, D.B., Maguire, E., Herr, B., & Hawley, C. (1986). Implicit theories and the evaluation of group process and performance. Organizational Behavior and Human Decision Processes, 37, 279-295.

Hackman, J.R. (1968). Effects of task characteristics on group products. Journal of Experimental Social Psychology, 4, 162-187.
Hackman, J.R. (1987). The design of work teams. In J. Lorsch (Ed.), Handbook of organizational behavior (pp. 315-342). Englewood Cliffs, NJ: Prentice Hall.

Hackman, J.R. & Morris, C.G. (1975). Group tasks, group interaction process, and group performance effectiveness: A review and proposed integration. In L. Berkowitz (Ed.) Advances in experimental social psychology (Vol 2). New York: Academic Press.

Hackman, J.R., & Vidmar, N. (1970). Effects of size and task type on group performance and member reactions. Sociometry, ??, 37-54.

Hakel, M.D., Ohnesorge, J.P., & Dunnette, M.D. (1970). Interviewer evaluations of job applicants' resumes as a function of the qualifications of the immediately preceding applicants: An examination of contrast effects. Journal of Applied Psychology, 54, 27-30.

Hastie, R. (1980). Memory for information that confirms or contradicts a personality impression. In R. Hastie et al. (Eds.) Person memory: The cognitive bassis of social perception. (pp. 155-177), Hillsdale, N.J.: Erlbaum.

Hastie, R., & Kumar, P.A. (1979). Person memory: Personality traits as organizing principles in memory for behaviors. Journal of Personality and Social Psychology, 37, 25-38.

Heine, D.M. (1990). Clinical versus mechanical scoring of a similuation instrument: A comparative study. Unpublised doctoral disseration. University of Minnesota, Minneapolis, MN.

Helson, H. (1948). Adaptation-level as a basis for quantitative theory of frames of reference. Psychological Review, 55, 297-313.

Helson, H. (1964). Adaptation-level theory. New York: Harper & Row, 1964.

Herr, P.M., Sherman, S.J., & Fazio, R.H. (1983). On the consequences of priming: Assimilation and contrast effects. Journal of Experimental Social Psychology, 29, 323-340.

Higgins, E.T., & Bargh, J.A. (1987). Social cognition and social perception. Annual Review of Psychology, 38, 369-425.

Higgins, E.T., Rholes, W.S., & Jones, C.R. (1977). Category accessibility and impression formation. Journal of Experimental Social Psychology, 13, 141-154.

Hoffman, C., Mischel, W., & Mazze, K. (1981). The role of purpose in the organization of information about behavior: Trait-based versus goal-based categories in person cognition. Journal of Personality and Social Psychology, 40, 211-225.

Holmes, D.S., & Berkowitz, L. (1961). Some contrast effects in social perception. Journal of Abnormal and Social Psychology, 62, 150-152.

Hovland, C.I., & Sherif, M. (1952). Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone Scales. Journal of Abnormal Social Psychology, 47, 822-832.

Ilgen, D.R., Barnes-Farrell, J. L., & McKellin, D. B. (in press). Performance appraisal accuracy. Organizational Behavior and Human Decision Processes.

Ilgen, D.R., & Feldman, J.M. (1983). Performance appraisal: A process focus. In B.M. Staw & L.L. Cummings Research in organizational behavior (Vol. 5, pp. 141-197). Greenwich, CT: JAI Press.

James, L.R., & Brett, J. M. (1984). Mediators, moderators, and tests for mediation. Journal of Applied Psychology, 69, 307-321.

Jones E.E., & Goethals G.R. (1972). Order effects in impression formation: Attribution context and the nature of the entity. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, & B. Wiener (Eds.), Attribution: Perceiving the caused of behavior (pp. 27-46). Morristown, NJ: General Learning Press.

Kabanoff, B., & O'Brien, G.E. (1979). The effects of task type and cooperation upon group products and performance. Organizational Behavior and Human Decision Processes, 23, 163-181.

Karl, K., & Wexley, K. (1989). Patterns of performance and rating frequency: Influence on the assessment of performance. Journal of Management, 15, 5-20.

Kelley, H.H. (1972). Causal schemata and the attribution process. In E.E. Jones, D.E. Kanouse, H.H. Kelley, R.E. Nisbett, S. Valins, & B. Wiener (Eds.), Attribution: Perceiving the caused of behavior (pp. 151-174). Morristown, NJ: General Learning Press.In E.E. jones, D.E.

Kelley, H.H. (1973). The process of causal attribution. American Psychologist, 28, 107-128.

Kiggundu, M.N. (1981). Task interdependence and the theory of job design. Academy of Management Review, 6, 499-508.

Klimoski, R., & Brickner, M. (1987). Why do assessment centers work? The puzzle of assessment center validity. Personnel Psychology, 40, 243-260.

Kopelman, M.D. (1975). The contrast effect in the selection interview. British Journal of Educational Psychology, 45, 333-336.

Kozlowski, S.W.J., Kirsch, M.P., & Chao, G.T. (1986). Job knowledge, ratee familiarity, conceptual similarity, and halo error: An exploration. Journal of Applied Psychology, 71, 45-49.

Kravitz, D.A., & Balzer, W.K. (1992). Context effects in performance appraisal: A methodological critique and empirical study. Journal of Applied Psychology, 77, 24-31.

Landy, F.J., & Bates, F. (1973). Another look at contrast effects in the employment interview. Journal of Applied Psychology, 58, 141-144.

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

Larson, J.R. Jr., (1982). Cognitive mechanisms mediating the impact of implicit theories of leader behavior on leader behavior ratings. Organizational Behavior and Human Decision Processes, 29, 129-140.

Larson, J.R. Jr., Lingle, J.H. & Serbo, M.M. (1984). The impact of performance cues on leader-behavior ratings: The role of selective information availability and probabilistic response bias. Organizational Behavior and Human Decision Processes, 33, 323-349.

Lerner, M.J. (1970). The desire for justice and reactions to victims. In J. Macaulay & L. Berkowitz (Eds.), Altruism and helping behavior (pp. 205-229). New York: Academic Press.

Levine, J.M., & Moreland, R.L. (1990). Progress in small group research. Annual Review of Psychology, 41, 585-634.

Liden, R.C., & Mitchell, T.R. (1983). The effects of group interdependence on supervisor performance evaluations. Personnel Psychology, 36, 289-299.

Lord, R.G. (1977). Functional leadership behavior: Measurement and relation to social power and leadership perceptions. Administrative Science Quarterly, 22, 114-133

Lord, R.G. (1985). An information processing approach to social perceptions, leadership, and behavioral measurement in organizations. In B.M. Staw & L.L. Cummings Research in organizational behavior (Vol. 7, pp. 87-128). Greenwich, CT: JAI Press.

Lord, R.G., Binning, J.F., Rush, M.C., & Thomas J.C. (1978). The effect of performance cues and leader behavior on questionnaire ratings of leadership behavior. Organizational Behavior and Human Decision Processes, 21, 27-39.

Lord, R.G., & Maher, K.J. (1989). Cognitive processes in industrial and organizational psychology. In K. Cooper and I. Robertson (Eds.) International Review of Industrial and Organizational Psychology, (pp. 49-91).

Lord, R.G., Phillips, J.S., & Rush, M.C. (1980). The effects of sex and personality on perceptions of emergent leadership, influence, and social power. Journal of Applied Psychology, 65, 176-182.

Martell, R.F., & Guzzo, R.A. (1991). The dynamics of implicit theories of group performance: When and how do they operate?. Organizational Behavior and Human Decision Processes, 50, 51-74.

Maurer, T.J., & Alexander, R.A. (1991). Contrast effects in behavioral measurement: An investigation of alternative process explanations. Journal of Applied Psychology, 76, 3-10.

McGrath, J.E. (1964). Social psychology: A brief introduction. New York: Holt.

McElroy, J.C., & Downey, H.K. (1982). Observation in organizational research: Panacea to the performance attribution effect?. Academy of Management Journal, 25, 822-835.

Miller, L. K. & Hamblin, R. L. (1963). Interdependence, differential rewarding, and productivity. American Sociological Review. 25, 768-778.

Mitchell, T.R., Larson J.R., & Green, S.G. (1977). Leader behavior, situational moderators, and group performance: An attributional analysis. Organizational Behavior and Human Decision Processes, 18, 254-268.

Mitchell, T. R. & Silver, W. S. (1990). Individual and group goals when workers are interdependent: Effects on task strategies and performance. Journal of Applied Psychology, 75, 185-193.

Moses, J. (1992). The changing expectations of leadership. General session at the International Congress on the Assessment Center Method. Colonial Williamsburg, VA.

Murphy, K.R., & Balzer, W.K. (1986). Systematic distortions in memory-based behavior ratings and performance evaluations: Consequences fo rating accuracy. Journal of Applied Psychology, 71, 39-44.

Murphy, K.R., Balzer, W.K., Lockart, M.C., & Eisenman, E.J. (1985). Effects of previous performance on evaluations of present performance. Journal of Applied Psychology, 70, 72-84.

Neidig, R. D., Martin, J. C., & Yates, R. E. (1979). The contribution of exercise skill ratings to final assessment center evaluations. Journal of Assessment Center Technology, 2, 21-23.

Neidig, R.D. & Neidig, P. J. (1984). Multiple assessment center exercises and job relatedness. Journal of Applied Psychology, [Short Note] 69, 182-186.

Norton, S. D. (1977). The empirical and content validity of assessment centers vs. traditional methods for predicting managerial success. Academy of Management Review, 2, 442-453.

Ostroff, C.L. (1985). Cognitive categories of raters in performance appraisal and their relationship to rating accuracy. Unpubished master's thesis. Michigan State University, East Lansing, MI.

Parducci, A., Knobel, S., and Thomas, C. (1976). Independent context for category ratings: A range-frequency analysis. Perception & Psychophysics, 20, 360-366.

Pfeffer, J. (1977). The ambiguity of leadership. Academy of Management Review, 2, 104-112.

Phillips, J.S. (1984). The accuracy of leadership ratings: A cognitive categorization perspective. Organizational Behavior and Human Decision Processes, 33, 125-138.

Phillips, J.S., & Lord, R.G. (1981). Causal attributions and perceptions of leadership. Organizational Behavior and Human Decision Processes, 28, 143-163.

Phillips, J.S., & Lord, R.G. (1982). Schematic information processing and perceptions of leadership in problem-solving groups. Journal of Applied Psychology, 4, 486-492.

Reilly, R., Henry, S. & Smither, J.W. (1990). An examination of the effects of using behavioral checklists on the construct validity of assessment center dimensions. Personnel Psychology, 43, 71-84.

Robertson, I., Gratton, L., & Sharpley, D. (1987). The psychometric properties and design of managerial assessment centres: Dimensions into exercises won't go. Journal of Occupational Psychology, 60, 187-195.

Rowe, P.M. (1967). Order effects in assessment decisions. Journal of Applied Psychology, 51, 170-173.

Rush, M.C., Phillips, J.S., & Lord, R.G. (1981). The effects of a temporal dela in rating on leader behavior descriptions: A laboratory investigation. Journal of Applied Psychology, 66, 442-450.

Rush, M.C., Thomas, J.C., & Lord, R.G. (1977). Implicit leadership theory: A potential threat to the internal validity of leader behavior questionnaires. Organizational Behavior and Human Decision Processes, 20, 93-110.

Russell, C. J. (1985). Individual decision processes in an assessment center. Journal of Applied Psychology, 70, 737-746.

Sackett, P. R. & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troublesome empirical findings. Journal of Applied Psychology, 67, 401-410.

Sackett, P. R. & Hakel, M. D. (1979). Temporal stability and individual differences in using assessment information to form overall ratings. Organizational Behavior and Human Performance, 23, 120-137.

Sackett, P. R., & Wilson, M. A. (1982). Factors affecting the consensus judgment process in managerial assessment centers. Journal of Applied Psychology, 67, 10-17.

Sackett, P.R., Zedeck, S., & Fogli, L. (1988). Relations between typical and maximum performance. Journal of Applied Psychology, 73, 482-486.

Schmitt, N. (1977). Interrater agreement in dimensionality and combination of assessment center judgments. Journal of Applied Psychology, 62, 171-176.

Schmitt, N, & Hill, T. E. (1977). Sex and race composition of assessment center groups as a determinant of peer and assessor ratings. Journal of Applied Psychology, 62, 261-264.

Schmitt, N., Schneider, J. R., & Cohen, S. A. (1990). Factors affecting validity of a regionally administered assessment center. Personnel Psychology, 43, 1-12.

Schneider, J.R. (1990). Understanding assessment center dimensions and constructs. Unpublished master's thesis. Michigan State University, East Lansing, MI.

Schneider, J.R. & Schmitt, N. (1992). An exercise design approach to understanding assessment center dimension and exercise constructs. Journal of Applied Psychology, 77, 32-41.

Schneider, W., & Shiffrin, R.M. (1977). Controlled and automatic human information processing: I. Detection, search and attention. Psychological Bulletin, 1-66.

Sherif, M., Taub, D., & Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. Journal of Experimental Psychology, 55, 150-155.

Silzer, R.F. (1984). Clinical and statistical prediction in an assessment center. Unpublished doctoral dissertation. University of Minnesota. Minneapolis, MN.

Schuh, A. J. (1978). Contrast effect in the interview. Bulletin of the Psychonomic Society, 11, 195-196.

Silverman, W. H., Dalessio, A., Woods, S. B., & Johnson, R. L. Jr. (1986). Influence of assessment center methods on assessors' ratings. Personnel Psychology, 39, 565-578.

Smither, J.W., Reilly, R.R., & Buda, R. (1988). Effect of prior performance information on ratings of present performance: Contrast versus assimilation revisited. Journal of Applied Psychology 73, 487-496.

Sorenson, J.R. (1973). Group member traits, group process, and group performance. Human Relations, 26, 639-655.

Srull, T.K. (1981). Person memory: some test of associateiv storage and retrieval models. Journal of Experimental Psychology: Human Learning and Memory, 7, 440-463

Srull, T.K., & Wyer, R.S. Jr. (1988). (Eds.) Advances in social cognition (Vol. 1). Erlbaum.

Staw, B.M. (1975). Attribution of the "causes" of performance: A general alternative interpretation of cross-sectional research on organizations. Organizational Behavior and Human Decision Processes, 13, 414-432.

Stogdill, R.M. (1963). Manual for the Leader Behavior Description Questionnaire--Form XXI. Columbus, OH: Bureau of Business Research, The Ohio State University.

Task Force on Assessment Center Guidelines (1989). Guidelines and ethical considerations for assessment center operations. Public Personnel Management, 18, 457-470.

Thompson, J.D. (1967). Organizations in action. New York: McGraw Hill.

Thornton, G.C. III. (1991). What we don't know about assessment centers: Implications for research and practice. Presentation at the Sixth Annual Conference of Industrial and Organizational Psychology, Inc.: Doctoral Consortium. St. Louis, MO.

Thornton, G.C. (1992). Assessment centers in human resource management. Menlo Park, CA: Addison-Wesley

Thornton, G. C., & Byham, W. C. (1982). Assessment centers and managerial performance, New York: Academic Press.

Turnage, J. J. & Muchinsky, P. M. (1982). Transsituational variability in human performance within assessment centers. Organizational Behavior and Human Performance. 30, 174-200.

Wexley, K.N., Sanders, R.E., & Yukl, G.A. (1973). Training interviewers to eliminate contrast effects in employment interviews. Journal of Applied Psychology, 57, 233-236.

Wexley, K.N., Yukl, G.A., Kovacs, S.Z., & Sanders, R.E. (1972). Importance of contrast effects in employment interviews. Journal of Applied Psychology, 56, 45-48.

Wickens, C.D. (1984). Engineering psychology and human performance. Columbus, OH: Merrill.

Williams, K.J., Blencoe, A.G., DeNisi, A.S, & Cafferty, T.P. (1983). Does appraisal purpose and outcome affect rater information search? Unpublished manuscript. Department of Psychology, University of South Carolina.

Winer, B.J. (Ed.) (1971). Statistical principles in experimental design. New York: MacMillan.

Woll, S.B., & Graesser, A.C. (1982). Memory discrimination for information typical or atypical of person schemata. Social Cognition, 1, 287-310.

Wyer, R.S. Jr., Bodenhausen, G.V., & Srull, T.K. (1984). The cognitive representation of persons and groups and its effect on recall and reconition memory. Journal of Experimental Social Psychology, 20, 445-469.

Wyer, R.S. Jr. & Gordon, S.E. (1982). The recall of information about persons and groups Journal of Experimental Social Psychology, 18, 128-164.

Wyer, R.S., & Srull, T.K. (1981). Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In E. Higgins, C. Herman, & M. Zanna (Eds.) Social cognition: The Ontario symposium (Vol 1). Hillsdale, NJ: Erlbaum.