To the Graduate Council:

I am submitting herewith a dissertation written by Erin Kate Patton Atchley entitled "The Role of Professional and Managerial Experience In Interpreting and Using Managerial Assessment Center Data." I have examined the final copy of this dissertation for form and content and recommend that it be accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy, with a major in Industrial and Organizational Psychology.

Robert T. Ladd, Major Professor

We have read this dissertation
and recommend its acceptance:

Accepted for the Council:

Associate Vice Chancellor and
Dean of The Graduate School

THE ROLE OF PROFESSIONAL AND MANAGERIAL EXPERIENCE IN INTERPRETING AND USING MANAGERIAL ASSESSMENT CENTER DATA

A Dissertation

Presented for the

Doctor of Philosophy

Degree

The University of Tennessee, Knoxville

Erin Kate Patton Atchley

December 1999

DEDICATION

This dissertation is dedicated to my husband and best friend, Scott. His steady support, constant patience, and unconditional love played a substantial role in my completing this milestone. I would also like to dedicate this dissertation to my parents, Steve and Judy Patton, for instilling in me a love of learning and teaching me to believe in my abilities.

ACKNOWLEDGMENTS

I am deeply indebted to a number of individuals who made this dissertation possible. First, I would like to thank my dissertation chair, Dr. Tom Ladd. He has been an advisor and a mentor to me for the past four years, and I deeply value his counsel and admire his integrity. I would be hard pressed to find a better role model. His encouragement and help on my dissertation is especially appreciated. Tom, consider yourself reinforced.

A special mention is needed here to speak about the role Dr. Gregory Dobbins played in my dissertation. Greg was instrumental in my professional development. While his untimely death limited the amount of time I had to work with him, it was his knowledge of and passion for the assessment process that first directed my attention toward this area of study. He is greatly missed.

I would also like to thank the other members of my dissertation committee, Dr. Joyce Russell, Dr. Larry James, and Dr. Kathy Lawler. Joyce continually motivated me to pursue challenging developmental opportunities, and I am grateful for her encouragement and advice over the years. Larry provided invaluable feedback regarding my methodology and statistical analyses, and for these things I thank him. Kathy also played an important role in the completion of this project. Her experimental psychology perspective was indispensable and her friendship is priceless. I am delighted my dissertation allowed us to reconnect.

There are several other individuals who have helped make this research possible. I would like to thank the members of The University of Tennessee's College of Business Administration 1999-2000 Scholarly Research Grant Program Committee. The financial support I received from the Scholarly Research Grant Program was an immense assistance to my research. Also, thanks to June, Jackie, and Carolyn for assisting with

bureaucratic nightmares, being "friendly faces" at work, and always being willing to help in any way possible.

I also have some dear friends to acknowledge. A special thanks to Jamie and his willingness to share his knowledge with me, and to Laura, Jennifer, and Debrah -- probably the greatest set of girlfriends I'll ever have. Your friendship has been one of the best things about graduate school, and I'll cherish my memories of the times we had together. And, finally, thank you to Lillian for encouraging me and always believing in my abilities to reach this goal.

ABSTRACT

The present study examined the role of professional and managerial experience in interpreting and using managerial assessment center data. Given the rising use of the assessment center method to select employees, it is essential to understand the degree to which professional and managerial experience impacts the accuracy of interpreting assessment center results. Past selection decision studies have failed to provide a clear answer to this question. The purpose of this study was to ascertain which type of experience, managerial or psychological, is essential in making accurate selection decisions.

Several hypotheses based on past selection and performance appraisal research, as well as the expert decision making literature, were examined. A secondary purpose of the current study was to examine the type of information being used across varying groups making selection decisions, in hopes of shedding light on differences in the decision making process. The decision processes of three different subject groups representing managers, trained assessment center assessors and undergraduate students were examined in this study. Forty-two managers from various companies, 34 Industrial and Organizational Psychology doctoral students trained as assessment center raters, and 50 undergraduate students served as subjects for this study. Subjects read three managerial job descriptions, examined assessment center results in the form of written summary reports, and rated 16 applicants on overall assessment center performance, predicted job performance, and four skill sets. Subject matter experts' ratings (i.e., managers with a graduate degree in I/O Psychology and assessment center experience) served as the criterion.

Calculations of Cronbach's (1955) evaluative accuracy components revealed that undergraduate students' ratings were the least accurate. The findings also revealed that managers were more capable than trained assessors at distinguishing dimensional differences in performance across individuals (i.e., stereotype accuracy). However, trained assessors were slightly more accurate in their rank ordering of applicants, and they were also slightly better than managers in making accurate dimension × applicant distinctions. A dominance analysis was performed to ascertain the relative importance of dimensional skill ratings in making predictions concerning the applicants' on-the-job performance. The results of the dominance analysis indicated that the subject groups used the information provided in different ways when making their selection decisions. These findings were further supported by the content analysis of subjects' written rationale for making their hiring decisions. Managers were more likely to focus on future possibilities and the skills applicants needed to acquire in order to fulfill the job requirements. In contrast, trained assessors tended to focus on the dichotomous decision to hire or not hire applicants. Limitations of the present study are discussed, in addition to the practical implications of the findings. Suggestions for future research are also outlined.

TABLE OF CONTENTS

LIST OF TABLES

TABLES PAGE

LIST OF FIGURES

FIGURE PAGE

CHAPTER I

INTRODUCTION

The Task Force on Assessment Center Guidelines (1989) defines an assessment center as:

> ...a standardized evaluation of behavior on multiple inputs. Multiple trained observers and techniques are used. Judgments about behaviors are made, in part, from specially developed assessment simulations. These judgments are pooled by the assessors at an evaluation meeting during which assessment data are reported and discussed and the assessors agree on the evaluation of the dimensions and any overall evaluation that is made. (p. 457)

Since the first business application of an assessment center methodology in 1956 by AT&T (Bray & Grant, 1966), the use of assessment centers has become prevalent in organizations. Companies utilize assessment centers for variety of purposes, including career management and training, promotion decisions, entry into management development programs, and selection (Spychalski, Quinones, Gaugler, & Pohley, 1997). Average validity coefficients for assessment centers predicting organizational rank and job performance are estimated to be 0.53 and 0.36, respectively (Gaugler, Rosenthal, Thornton, & Bentson, 1987; Klimoski & Strickland, 1977; Thornton & Byham, 1982). In short, assessment centers have become an accepted means of effectively evaluating individuals on a highly complex set of dimensions.

The assessment center methodology is theorized to lead to high validity coefficients for several reasons. First, assessment center exercises are typically designed

to elicit direct evidence, or samples, of a candidate's knowledge, skills, and abilities (KSAs) to perform a job. In other words, the exercises have high content validity with the actual job the candidate would perform if he or she were hired. This is in distinct contrast to other selection tools (e.g., biodata forms, application forms, and interviews) that provide *indirect* evidence, or signs, of a candidate's KSAs (Gatewood & Feild, 1994; Wernimont & Campbell, 1968). Second, assessment centers are thought to minimize judgment bias that abounds in subjective selection methods: assessors have multiple chances to observe each candidate in several different situations, ratings are based on recorded, concrete behaviors versus abstract impressions, ratings have to be justified to other raters with examples of behavioral incidents, and final ratings are reached via a consensus discussion process involving all the raters observing a particular candidate (Gaugler & Rudolph, 1992).

A considerable amount of assessment center research has focused on the nature of the performance dimensions assessed (e.g., Gaugler & Thornton, 1989), the manner in which ratings are made by assessors (e.g., Lorenzo, 1984; Russell, 1985), and the processes used by assessors in reaching consensus decisions (e.g., Pynes & Bernardin, 1982; Sackett & Wilson, 1982). However, little research has focused on the interpretation of assessment center results. While both managerial decision making (e.g., Dipboye, Fromkin & Wiback, 1975; Hakel, Ohnesorge & Dunnette, 1970; Hitt & Barr, 1989) and selection decisions (e.g., Barr & Hitt, 1986; Carlson, 1967; Grove, 1981; Singer & Bruhns, 1991) in their simplest forms have been studied, few studies have been conducted examining the managers' use of assessment center data in making selection decisions.

Results from an assessment center are more complicated than any other selection information a manager receives about an individual. Organizations require complex skills, and individuals can be assessed on a huge variance of behaviors using the assessment center methodology. For example, a typical assessment center may yield data on approximately 6-8 individuals whose performance is observed on 10-15 dimensions across 3-6 exercises. This translates into a manager having to integrate somewhere in the range of 180 to 720 data points to determine the best candidate for a specific job. One question is obvious: Are managers able to accurately combine and weigh assessment center results in order to select the most suitable person (i.e., the individual with the best chance of success on the job) to hire for a certain position?

In answering this question, several hypotheses are possible. One hypothesis is that the more familiar one is with the assessment center process, the better one can interpret the resultant data. A second hypothesis is that the more familiar one is with the job to be filled and successful managerial behaviors, the better one is able to interpret the assessment center data. Past selection decision studies have failed to provide clear-cut support for either of these assumptions. For example, several studies have investigated the effects of experience on rating accuracy. More specifically, while some researchers (e.g., Lorenzo, 1984) found that greater experience as an assessor resulted in more accurate ratings, other researchers found that experience level had little impact on selection ratings (e.g., Carlson, 1967; Langdale & Weitz, 1973). Similarly, studies investigating the impact of managerial experience on selection have been mixed in their findings. Several researchers have noted significant differences in the decisions made by managers versus non-managers (e.g., Singer & Bruhns, 1991), while others have found

few dissimilarities (e.g., Mullins, 1982). Still other research has indicated that assessment center decisions made by psychologists had higher predictive validity than assessment center decisions made by non-psychologists (Gaugler et al., 1987).

Thus, the question remains: How does management experience versus assessment center experience help with the interpretation of the complex information obtained from an assessment center? Attention should be paid to the role expertise plays in making accurate selection decisions, instead of relying on the assumption that managers, by the very state of their being managers, are able to accurately select employees. The present study examined how accurately three distinct groups of individuals interpret the highly complex nature of assessment center reports. One group, trained assessment center assessors with a background in industrial and organizational psychology, has clear expertise in assessment center methodology, but casual knowledge of managerial tasks. The second group, a pool of managers, has managerial experience but only cursory knowledge of assessment centers and I/O psychology. The third group, which served as a control group, consisted of a naive sample of undergraduates who have no managerial experience and only very superficial knowledge of assessment centers and I/O psychology.

If the comparison groups are envisioned as a 2 x 2 matrix (see Figure 1), the undergraduate students are in the first cell, denoting their minimal knowledge and experience levels with both managerial and assessment center matters. Both the trained assessor and manager samples have expertise pertaining to one part of the cognitive process, but neither group is an expert in the *complete* process. A fourth group of

individuals with both managerial and assessment center experience served as subject matter experts for this study; their responses yielded criterion data for the current study.

	No Managerial Experience	*Managerial Experience*
No Assessment Center Experience	Undergraduate students	Managers
Assessment Center Experience	Trained assessors	Subject matter experts

FIGURE 1: Experience Levels of the Comparison Groups

CHAPTER II

REVIEW OF THE LITERATURE

Before the present study is described, the relevant literature will be summarized. As previously mentioned, there is a paltry amount of research examining how individuals use assessment center data to make selection decisions. However, studies from other areas can be drawn upon to shed light on this matter. The ensuing section reviews pertinent research in the following areas: the impact of experience level on performance judgments, decision making by experts, cognitive limitations of raters, the impact of training on ratings, and methodological limitations of selection decision research.

The Impact of Experience Level on Performance Judgments

Managers Versus Non-Managers

One presumption in the field of I/O psychology is that managers make "better" (i.e., more accurate) selection decisions than other individuals due to their greater experience in hiring individuals and their greater knowledge about the job to be filled (Gaugler et al., 1987). Several areas of research support this assumption. In the performance appraisal arena a study compared the ratings of undergraduates, MBA students, and managers (Cardy, Bernardin, Abbott, Senderak, & Taylor, 1987). The results indicated that the group who had more experience in making performance appraisal ratings, the managers, provided the most accurate ratings. Similarly, as measured by Cronbach's differential elevation component (Cronbach, 1955), Lorenzo (1984) found that managers with greater experience as assessment center assessors made more accurate ratings. Lorenzo suggested that assessorship developed the managers'

skills in acquiring and evaluating information about individuals. His supposition is supported by the findings of Maederer (1993), who studied the impact of applicant attractiveness on assessors' ratings. Maederer found that ratings made by more experienced raters were impacted less by levels of applicant attractiveness than less experienced raters.

However, several studies that have compared experienced and inexperienced interviewers have found no significant differences in the applicants' ratings between the two groups. Carlson (1967) compared managers grouped by type and length of hiring experience and failed to find differences in intra- and interrater agreement of ratings of interviews. Langdale and Weitz (1973) found that level of interviewing experience had no effect on interrater reliability, although the use of highly specific job information (i.e., a job description and position title) increased interrater reliability and discrimination among applicants, regardless of experience in interviewing. Similarly, Weiner and Schneiderman (1974) investigated the provision of job descriptions to experienced and inexperienced interviewers. Although job descriptions significantly decreased the use of irrelevant information in making selection decisions, the level of interviewing experience did not enhance the use of job descriptions. In other words, there were no significant differences in the utilization of job description information among experienced and inexperienced interviewers.

Relating these findings to studies comparing differences in manager and student selection decisions, the assumption is that students are the less experienced group in selecting employees and have less knowledge regarding the relationship between observed behaviors and job responsibilities. However, past research has not definitively

answered the question concerning differences in managers' and students' selection decisions. For example, Mullins (1982) found that few differences in selection decisions between the two groups existed, while other researchers noted significant dissimilarities (Landy & Bates, 1973; Singer & Bruhns, 1991; Singer & Sewell, 1989), and still other researchers found mixed results (Barr & Hitt, 1986; Bernstein, Hakel, & Harlan, 1975; Hakel, Ohnesorge, & Dunnette, 1970).

Dipboye et al. (1975) found that students rated applicants more favorably than managers, though both samples rated males and attractive applicants higher than females or unattractive applicants. In a similar study, Barr and Hitt (1986) studied differences in selection decisions of managers and students using an interview simulation. They found that while students tended to be more lenient in their ratings of applicants and suggested higher starting salaries, their decision making process was similar to that of the managerial sample. More specifically, both samples used job irrelevant information to make their decisions (e.g., race, age, and sex). The results of this study supported the findings of Bernstein et al. (1975). In their review of six studies that compared the decision processes of students and managers, it was determined that students made more lenient evaluations; however, decision processes were found to be comparable between the two samples.

Hakel, Ohnesorge, and Dunnette (1970) looked at contrast effects (i.e., discrepancies between the judged value of a stimulus and its actual value) in selection decision tasks. Their research indicated that while students rated applicants more favorably than managers, no significant differences in the size of contrast effects were

noted between the samples. However, in a replication of this study, Landy and Bates (1973) found contrast effects present only in the student sample.

Singer and Sewell (1989) investigated the effect of applicant age information on hiring decisions. They reported that managers and students made different decisions based on age-related information received: managers favored the older employee for a low-status position, and students favored the younger applicant for both the low-status and high-status jobs. In another study that examined the impact of information on selection decisions, Singer and Bruhns (1991) found that managers tended to weight applicant work experience more than academic qualifications; however, students gave significantly more weight to academic information than to work experience.

Olian, Schwab, and Haberfeld (1988) conducted a meta-analysis of studies investigating the impact of gender and qualifications on hiring recommendations. Their results indicate that student samples tend to provide more homogeneous evaluations of applicants than manager samples. However, the results of Mullins' 1982 study suggest that there are few differences in the characteristics of selection decisions made by students and managers. In an examination of 75 comparisons (i.e., 15 applicant desirability ratings for each of five videos) between student judgments and judgments made by professional interviewers, only four significant differences were found.

Barr and Hitt (1986) suggest two reasons why the evidence concerning comparability of student and manager samples is so inconsistent. First, it is possible that methodological shortcomings may be causing the noted inconsistencies in the research results. A second possibility is that students and managers really do make dissimilar selection decisions due to differences in amount of experience and knowledge.

Psychologists Versus Non-Psychologists

A related vein of research has examined potential differences in the ratings made by professional (i.e., psychologists) and non-professional (i.e., line managers) raters. Two studies found no differences in assessment center ratings made by the two groups (Greenwood & McNamara, 1969; Thomson, 1970). In marked contrast, a meta-analysis of 50 assessment center studies found that assessment centers using psychologists as raters had higher levels of predictive validity than assessment centers using managers to make ratings (Gaugler et al., 1987). Though the authors did not offer an explanation for this opposing finding, it may be that psychologists possess specialized knowledge that enhances the quality and accuracy of their ratings. In other words, psychologists and psychologists-in-training (i.e., psychology graduate students) are expected to learn highly specialized knowledge (e.g., individual differences, measurement, group dynamics, judgment biases) in order to obtain their Ph.D. and become licensed, and they are taught this information by subject matter experts. However, the question of the impact of experience and knowledge level on selection decisions remains unexamined.

Research on Experts

The topic of experts has been examined from several different perspectives, including the psychological characteristics of experts, the processes experts use to make decisions, and the quality of the decisions experts make. A brief review of each of these areas follows.

Psychological Characteristics of Experts

Shanteau (1987; 1988; 1992) outlined ten psychological characteristics of experts. The first four characteristics concern experts' knowledge and how it differs from that of

novices. The underpinning of expertise is an up-to-date and extensive knowledge base about the domain at hand. Not only do experts know more, but also they approach the decision making task in a unique manner. More specifically, experts have highly developed selective attention abilities. In other words, experts are able to note critical information that novices overlook. In addition, experts are able to place appropriate weight on relevant versus non-relevant information when making decisions, while novices are typically sidetracked by extraneous stimuli. Finally, experts are able to simplify complex situations in order to deal with human cognitive limitations. For example, research of chess players indicates that chess masters have exceptional pattern recognition abilities (Chase & Simon, 1973; deGroot, 1965). Other studies have indicated that experts in the fields of computer programming and mathematics have superior memory and problem solving skills (Meyer, 1983).

Three other psychological characteristics of experts, though indirectly related to their level of domain-specific knowledge, are more directly related to problem solving skills. For example, Shanteau and Dino (1993) found that experts function better under non-ideal and stressful conditions than novices and continue to make effective decisions. Experts can also identify situations where exceptions to their usual decision making strategies are appropriate, and they also know how and when to alter their decision making strategies due to changing task or situational conditions (Phelps, 1977; Shanteau & Phelps, 1978). In contrast, novices rigidly stick "to the rules" despite exceptions or changing task conditions (Phelps, 1977).

The last three characteristics can be classified as a self-confidence domain (Shanteau, 1988). First, experts are able to effectively declare their expertise to others,

while novices tend to have trouble establishing their credibility (Shanteau & Dino, 1993). Second, experts have great faith in their ability to make accurate, reliable decisions. Lastly, experts exhibit a willingness to accept responsibility for all of their decisions, both the good and the poor ones. Novices, on the other hand, may have difficulty continuing the task if they make a poor decision (Shanteau, 1992).

Experts in Cognitive Science Research

The field of cognitive science has examined the process by which experts make decisions. Cognitive science holds the view that experts within their specific domain are highly skilled decision makers that think in fundamentally different ways than novices, and that the nature of these differences can be captured via verbal protocols (Anderson, 1981; Chi, Glaser & Farr, 1988). Experts' special abilities reside in a particular knowledge base. Outside of the domain-specific area, an expert is relegated to the status of novice. Additionally, experts use automated processes, like pattern recognition, that are parallel yet independent of each other (Schiffrin & Schneider, 1977; Shanteau, 1992). In fact, consistent with the greater automatic processing observed in experts, experts tend to be inarticulate about the processes they used to make their decisions (Anderson, 1982; Shanteau, 1988). In contrast, novices use controlled processes that are sequential, linear, and easily described. Finally, protocol analysis has been used to understand experts' decision making processes and to clarify decision strategies (Ericsson & Simon, 1980).

Experts in Judgment and Decision Making Research

Past research in judgment and decision making has examined the characteristics of experts' decisions, including the reliability and accuracy of experts' judgments, the amount of information used by experts to make decisions, the influence of irrelevant

information on experts' decisions, and the reliance on heuristics that lead to judgment errors. In a study of soil judges, Shanteau and Gaeth (1981) found that experts' were both unreliable and inaccurate. Similar results have been found in other studies experts, including medical diagnoses (Einhorn, 1974) and court judges (Ebbesen & Konecni, 1975). Other research has examined the amount and type of information used to form decisions. Many studies have indicated that experts rely on very little information to make their decisions (Ebbeson & Konecni, 1975; Einhorn, 1974). For example, Galbraith and Alvares (1983) found that trained assessors relied on less than one-third of the behavioral dimensions rated to make overall judgments. In addition, several studies have found that experts are influenced by extraneous information that is irrelevant to the decision at hand (Gaeth & Shanteau, 1984; Nagy, 1981; Shanteau, Grier, Johnson, & Berner, 1991).

Finally, several researchers have studied the use of mental heuristics in making decisions. Kahneman and Tversky (1972; 1973) and Tversky and Kahneman (1974; 1981) state that virtually all individuals, including experts, rely on several heuristics when making decisions and that the use of heuristics can result in decision errors. These "rules of thumb" help simplify and increase the efficiency of the decision making process, especially when individuals are working with limited cognitive capacity and/or limited information (Bazerman, 1986). It is important to note that individuals unconsciously apply these heuristics; unless people undertake the study of cognitive science, individuals do not even realize the existence of these mental rules. Most of the time, the decisions resulting from the use of heuristics are acceptable. In addition, the time and energy saved by using heuristics typically compensates for the reduced quality

of decisions. However, it remains that heuristic utilization results in biased judgments and systematic errors . Tversky and Kahneman (1974) outlined numerous heuristics, including insensitivity to sample size, misconceptions of chance, overconfidence in judgments, hindsight, biases due to illusory correlation, and biases due to retrievability of information. Bolger and Wright (1992) discuss several heuristics that have been studied in the expert decision-making literature: overestimation and overconfidence. Overestimation occurs when experts assign too high a probability to a set of occurrences. Overconfidence occurs when low-probability events are assigned an even lower probability, and high-probability events are assigned an even higher probability. In general, past research has found experts to fall prey to these heuristics, leading to low quality and inaccurate decisions (Dolan, Bordley, & Mushlin, 1986; Stael von Holstein, 1977; Wagenarr & Keren, 1986).

In summary, researchers have stated few positive things about the ability and credibility of experts' decisions. However, several methodological limitations have been proposed that result in these less-than-glowing results (Bolger & Wright, 1992). First, many studies have involved asking experts to make decisions outside of their realm of expertise. Second, unfamiliar measurements have been employed in some studies. For example, one study asked experts to judge the percentage of foreign material in samples, instead of the typical measurement of estimating weight. Third, the expert subject pools are questionable; that is, it can be argued that many researchers have not used true expects, but instead, practiced novices (Bolger & Wright, 1992) that had not proven their expertise in the relevant domain.

Cognitive Limitations of Raters

It is possible that despite any amount of knowledge or experience, individuals have cognitive limitations that minimize their use of information in forming judgments about others' behaviors (Grove, 1981), an idea echoed by Dugan (1988). In a study of interviewer behavior, Grove (1981) stated that "the interviewer is often in a poor position to process qualitatively differing pieces of predictive information and interpret their significance" (p. 57). Studies concerning assessment center judgments support this notion (Galbraith & Alvares, 1983; Russell, 1985; Sackett & Hakel, 1979). For example, Sackett and Hakel found that that assessors typically do not use all the information available to them, even when instructed to do so. In addition, several behavioral dimensions were used more often than others were in forming overall judgments (i.e., Leadership, Planning and Organizing, and Decision Making), though the use of specific dimensions was not consistent across assessors. In addition, a study by Galbraith and Alvares (1983) indicated that raters relied on less than one-third of the behavioral dimensions to make overall ratings, and again, the use of specific dimensions was different across raters. Similarly, Russell (1985) found that a single factor (i.e., either Problem Solving or Interpersonal Skills) typically dominated assessor ratings. Results from studies of other types of judgment tasks have similar findings. For example, Slovic and Lichtenstein (1971) noted that three information cues accounted for approximately 80% of predictable variance in ratings. These findings, taken in aggregate, suggest Miller's assertion in 1956 that an individual is able to process only 7 ± 2 pieces of information may also hold true for trained raters.

Impact of Training on Rating Characteristics

So can raters be taught to make better ratings? Some researchers believe (e.g., Byham, 1977), despite research to the contrary, that little or no training is necessary because integration of data is a task most people can easily perform. Notwithstanding this assertion, rater training has proven to be a popular research topic. The results of studies concerning the impact of training on rating characteristics have also be mixed. The majority of rater training research has examined decreasing rater error and increasing rater accuracy. While some research indicates that training alters raters' judgments, not all of the studies had clear results or results in the intended direction.

A multitude of performance appraisal studies indicate that while raters trained to make fewer rating errors (e.g., halo error, leniency bias, severity bias, range restriction) do, in fact, decrease their rating errors, they also make judgments as inaccurate, and in some cases, less accurate than before training (Bernardin & Buckley, 1981; Bernardin & Pence, 1980; Hedge and Kavanagh, 1988). Still other researchers have found that training has no impact at all on accuracy of ratings (Zedeck & Cascio, 1982) or on the amount of information used by assessors to form overall ratings (Dugan, 1988). Furthermore, Gaugler et al.'s 1987 meta-analysis found that amount of assessor training had no impact on the validity of assessment centers, an interesting finding considering that careful assessor training is assumed to be primary for generating psychometrically sound ratings (Task Force on Assessment Center Guidelines, 1989).

However, other researchers have reached the contradictory conclusion that training positively impacts impact rater judgments. For example, in the area of performance appraisal, Wexley, Sanders, and Yukl (1973) were successful at reducing

contrast effects by instructing raters to avoid comparing employees with each other. Pulakos (1986) successfully used two different training programs to increase accuracy in performance ratings, one training program that emphasized a common frame of reference for raters and another that taught raters to keep track of relevant behaviors. In addition, a study of a soil judging task (Gaeth & Shanteau, 1984) compared two types of training: a lecture versus an interactive practice session. While the accuracy of ratings increased in both training groups, it increased more in the interactive practice session group. Furthermore, the authors found the impact of the training to be present 18 months after the training took place. In a study of assessment center raters, Galbraith (1985) successfully assessors to adopt and correctly use specified dimension weightings. In addition, Thornton and Zorich (1980) were successful in using training to increase observational accuracy of behaviors, and several studies have successfully used frame-of-reference training to enhance rater accuracy (Athey & McIntyre, 1987; Pulakos, 1984; Stamoulis & Hauenstein, 1993).

Methodological Limitations of Past Research

Undergraduates were utilized as samples in many of the selection studies reviewed. College students are the convenient subject pools of choice in the majority of psychology studies. In fact, several reviews of published psychology articles have indicated that 70-87% of these studies employed students as subjects (Barr & Hitt, 1986; Dipboye & Flanagan, 1979; Bernstein et al., 1975; Gordon, Slade, & Schmitt, 1986). This has resulted in many theories concerning behavior in organizational settings (e.g., performance, motivation, satisfaction, decision making, etc.) being based on results obtained from student samples placed in artificial situations and participating in non-

meaningful tasks (Barr & Hitt, 1986). The presence of student samples in research studies is highly confounded with the occurrence of laboratory studies; that is, student samples tend to participate in laboratory research, while managers tend to be involved in field studies. The appropriateness of utilizing student-based research findings to explain managerial behavior in a real-world setting has been an on-going debate for some time (e.g., Dobbins, Lane, & Steiner, 1988a; Dobbins, Lane, & Steiner, 1988b; Gordon et al., 1986; Guion, 1983; Slade & Gordon, 1988). A review of I/O psychology research by Dipboye (1990) concluded that many basic processes in judgment tasks, like weighing of information, contrast and assimilation effects, and demographic biases, are appropriately studied with student samples; however, great caution should be used when generalizing results to other populations.

Much of the research reviewed in the preceding sections involved asking subjects to perform non-realistic tasks, like selecting employees from interview transcripts or making ratings based on resumes with limited information about the applicants. In fact, Landy and Bates (1973) conducted post-study interviews with their managerial sample and found that the managers thought the study was artificial due to the unrealistically small amount of information provided about applicants.

Furthermore, most prior research has targeted selection for entry-level jobs (Gaugler & Rudolph, 1992; Gorman, Clover, & Doherty, 1978; Grove, 1981; Langdale & Weitz, 1973; Weiner & Schneiderman, 1974). Thus, little is known about selection decisions regarding middle and upper level management positions.

Finally, selection decisions are complex in nature, more so than past laboratory research has allowed for with attempts to control extraneous factors and manipulate

variables. In support of this statement is Hitt and Barr's 1986 study. They found that variance accounted for by the main effects was significantly less than variance accounted for by interaction effects, indicating that selection decisions may involve more complicated cognitive processes than previously examined.

The Present Study

The present study examined differences in hiring decisions across three distinct groups of individuals, namely experienced managers, experienced assessment center raters, and inexperienced undergraduate students. As noted previously, prior selection decision research has not judiciously examined the impact of managerial experience and assessment center experience on the accuracy of hiring decisions. In other words, this study attempted to ascertain which type of expertise (i.e., managerial or psychological) is more related to selection decision accuracy. Given the rising use of the assessment center method to select employees, it seems imperative to understand the impact of experience on making selection decisions.

In addition, the present study investigated the types of assessment center information used by the different subject groups to make their respective hiring decisions. If expertise and experience do, in fact, impact the accuracy of assessment center data interpretation, then it follows that the subject groups may focus on different types of available information in making their decisions.

This study also attempted to rectify several methodological limitations present in past research. Past selection decision research has typically utilized tasks with limited realism, including the use of non-realistic stimuli and very simple procedures that do not represent the complexity of real-world selection procedures. In addition, a limited

subject pool (e.g., college sophomores) has been studied, severely limiting the generalizability of results. In an effort to address the issue of stimuli realism, assessment center summary reports from an actual managerial assessment center were used as stimulus material. Second, to address the issue of over-simplified tasks, subjects were asked to render hiring decisions on a number of qualified applicants for three distinct job positions. Third, three distinct groups of subjects were compared (i.e., managers, assessment center raters, and undergraduates) in an attempt to address the issue of non-generalizable samples.

Summary of Research Hypotheses

Based on the preceding review of the relevant literature, several hypotheses are proposed regarding differences in selection decisions made by students, assessors, and managers. A discussion of the rationale for the hypotheses is provided in the following section.

There is much research across several domains that indicate experts use different mental processes (e.g., automated processes) than novices when making decisions (Anderson, 1981; Chi et al., 1988; Schiffrin & Schneider, 1977; Shanteau, 1992) and have different psychological characteristics than novices (Shanteau, 1987; 1992). Experienced assessment center raters meet many of the requirements that define an "expert," including extensive domain-specific knowledge, highly developed selective attention abilities, the ability to appropriately weight information, and the capability to simplify complex situations in order to deal with human cognitive limitations (Shanteau, 1992). Though several studies have indicated that experts' decisions are both unreliable and inaccurate, several methodological limitations have been pointed out to explain these

findings, namely asking experts to make decisions outside of their realm of expertise, the use of unfamiliar metrics, and subjects that do not meet the definition of "expert" (Bolger & Wright, 1992). Experienced assessment center raters with a graduate education in I/O psychology, however, are making judgments within their realm of expertise, are using familiar standards to render judgments, and meet many of the characteristics reserved for the label of expert.

Furthermore, in accordance with the Task Force on Assessment Center Guidelines (1989), assessors have typically received highly specialized training that includes a multitude of topics designed to reduce inaccuracy (e.g., frame of reference training, relevant versus irrelevant behaviors, weighting of information, and avoiding rating biases). Though findings are mixed concerning the impact of training on rating accuracy, there are several studies that successfully showed the positive results of training (Gaeth & Shanteau, 1984; Galbraith, 1985; Pulakos, 1986; Thornton & Zorich, 1980).

Managers, on the other hand, may have experience in hiring individuals that aids their decision making ability. Several studies have found that managers with greater experience in making ratings do, indeed, make more accurate ratings than less experienced individuals (Cardy et al., 1987; Lorenzo, 1984). However, other researchers have reported that amount of hiring experience had little impact on quality of ratings (Carlson, 1967; Langdale & Weitz, 1973). It is not clear that managers have other characteristics necessary to be considered experts. For example, experts are able to place appropriate weight on information; however, Barr and Hitt (1986) found that managers used job irrelevant information to make their decisions. Furthermore, Gaugler et al.

(1987) found that assessment centers using psychologists as raters had better predictive validity than assessment centers using managers to make ratings.

In addition, undergraduate students typically have neither the experience nor the subject matter expertise needed to make accurate hiring decisions. For example, in a meta-analysis by Olian et al. (1988), student samples tended to maker fewer distinctions between applicants than managerial samples. Other research findings have indicated that students tend to make more elevated ratings than managers (e.g., Barr & Hitt, 1986; Bernstein et al., 1975; Hakel, Ohnesorge, & Dunnette, 1970). Also, Singer and Bruhns (1991) found that students used different information (i.e., academic performance) than managers (i.e., work experience) to make decisions. Thus, the following hypotheses are proposed:

Hypothesis 1: The correlation between assessment center true scores and subjects' ratings of assessment center performance will significantly differ across the three samples, with trained assessors having the strongest positive correlation, followed by managers.

Hypothesis 2: The correlation between assessment center true scores and subjects' ratings of predicted job performance will significantly differ across the three samples, with trained assessors having the strongest positive correlation, followed by managers.

Hypothesis 3: As defined by Cronbach's Differential Accuracy components, trained assessors will make more accurate ratings than either managers or students.

Hypothesis 3a: Assessors' ratings will contain less elevation inaccuracy than ratings made by managers or students.

Hypothesis 3b: Assessors' ratings will contain less differential elevation inaccuracy than ratings made by managers or students.

Hypothesis 3c: Assessors' ratings will contain less stereotype inaccuracy than ratings made by managers or students.

Hypothesis 3d: Assessors' ratings will contain less differential inaccuracy than ratings made by managers or students.

Hypothesis 4: As defined by Cronbach's Differential Accuracy components, managers will make more accurate ratings than students.

Hypothesis 4a: Managers' ratings will contain less elevation inaccuracy than ratings made by students.

Hypothesis 4b: Managers' ratings will contain less differential elevation inaccuracy than ratings made by students.

Hypothesis 4c: Managers' ratings will contain less stereotype inaccuracy than ratings made by students.

Hypothesis 4d: Managers' ratings will contain less differential inaccuracy than ratings made by students.

Hypothesis 5: There will be differences in the type of information used to make hiring decisions across the three samples.

The final hypothesis varies slightly from the postulation that assessors, as subject matter experts, make the best decisions. Barr and Hitt (1986) and Hitt and Barr (1989) used starting salary as a dependent variable in their study comparing students' and managers' selection decisions. Their results indicated that students recommended significantly higher starting salaries than managers did, and the authors surmised that the students' inexperience was the cause of this leniency. Assessors, as well, may not be

experienced at establishing starting salaries, thus, they may not be making decisions that fall within their range of expertise. However, if managers routinely arrive at salary decisions for their employees, they may evidence better judgment in this situation. Accordingly, the following hypothesis is offered:

Hypothesis 6: Undergraduate students and assessors will recommend more lenient (i.e., higher) starting salaries than managers.

CHAPTER III

METHODOLOGY

Overview of the Study

Four sample groups (i.e., undergraduate students, managers, trained assessors, and subject matter experts) were asked to read job descriptions of three managerial jobs. Each subject was then asked to read about 16 applicants' performance in a managerial assessment center and make ratings for each applicant on six performance dimensions and one salary variable. In other words, there were sixteen applicants nested within three jobs and rated by four subject groups. The number of applicants assessed varied by job description. More specifically, there were three applicants for a human resource position, five for an administrative position, and eight for an engineering position. This design resulted in 112 data points for each subject that participated in the study (i.e., 16 applicants x 7 ratings). See Figure 2 for a graphical representation of the study design.

This design approximated a real-world scenario, where employers would evaluate more than one applicant and where a varying number of applicants apply for different job openings. In addition, this design resembled the large amount of information that must be integrated in order to form summary judgments in a real-world selection situation (Shack, 1983).

Sample

The student sample consisted of 59 undergraduate psychology students at a large southeastern public university. Students who participated received extra credit and were

	Undergraduate Students			Managers			Trained Assessors			Subject Matter Experts		
Job Description	HR Manager (n = 3)	Admin Manager (n = 5)	Engineer Manager (n = 8)	HR Manager (n = 3)	Admin Manager (n = 5)	Engineer Manager (n = 8)	HR Manager (n = 3)	Admin Manager (n = 5)	Engineer Manager (n = 8)	HR Manager (n = 3)	Admin Manager (n = 5)	Engineer Manager (n = 8)
Rated Dimensions	1. Assessment Center Performance 2. Problem Solving 3. Organizational Leadership 4. Communication Skills 5. Managerial Skills 6. Predicted Job Performance 7. Starting Salary			1. Assessment Center Performance 2. Problem Solving 3. Organizational Leadership 4. Communication Skills 5. Managerial Skills 6. Predicted Job Performance 7. Starting Salary			1. Assessment Center Performance 2. Problem Solving 3. Organizational Leadership 4. Communication Skills 5. Managerial Skills 6. Predicted Job Performance 7. Starting Salary			1. Assessment Center Performance 2. Problem Solving 3. Organizational Leadership 4. Communication Skills 5. Managerial Skills 6. Predicted Job Performance 7. Starting Salary		

Figure 2: Study Design

also offered a chance to win one of three $50 prizes in a random drawing. Most of the student participants were female (78%), and their mean age was 19.17 years (SD = 1.78). None of the students were employed full-time, had served as assessment center raters at the time of the study, or had significant managerial experience (see Table 1).

The manager sample was drawn from an alumni list obtained from an executive MBA program. The individuals on this list, plus some additional contacts of the researcher, were mailed the survey materials. Managers were offered a chance to win $250 in a random drawing if they participated in the study. Of the 95 managers contacted, 44% (n = 42) agreed to participate and completed the survey. The majority of the managers were male (67%), in their early 40s (M = 43.07; SD = 5.95), and were employed on a full-time basis (100%). Most had received either a bachelors (71%) or a master's degree (50%), and their mean supervisory experience was 15.21 years (SD = 6.02). None of the managers had had significant assessment center rating experience at the time of the study: the average number of assessment centers rated was 0.05 (SD = 0.22), and the average number of candidates rated was 0.19 (SD = 0.89). In addition, none of the managers had completed graduate courses topics related to industrial and organizational psychology (see Table 1). Managerial subjects, on average, had supervisory responsibility for 207.19 employees (SD = 804.13), had hired 53.36 employees (SD = 69.29), and had conducted performance appraisals on 178.50 employees (SD = 423.85).

The trained assessor sample was gathered from a list of current graduate students and recent alumni of an industrial and organizational psychology doctorate program at a large southeastern public university. Forty-eight people were contacted as potential

Table 1. Descriptive Statistics for Sample Groups and Subject Matter Experts

Variable	Students n = 59	Managers n = 42	Assessors n = 34	SMEs n = 8
1. Mean Age in years	19.17	43.07	28.88	36.25
	(1.78)	(5.95)	(3.75)	(4.80)
2. Percentage with Full Time Employment	2%	100%	44%	100%
3. Mean Job Tenure in years	0	3.05	1.58	2.99
		(3.25)	(1.34)	(2.75)
4. Mean Years of Management Responsibility	0.68	15.21	1.46	7.14
	(1.28)	(6.02)	(1.90)	(3.67)
5. Mean # of Employees Supervised	0.71	207.19	1.15	10.00
	(3.14)	(804.13)	(2.54)	(9.49)
6. Mean # of Employees Hired	0.49	53.36	5.38	24.13
	(3.27)	(69.29)	(14.42)	(17.52)
7. Mean # of Performance Appraisals Conducted	0.85	178.50	6.44	22.62
	(4.05)	(423.85)	(13.43)	(16.33)
8. Mean # of Assessment Centers Rated	0	0.05	26.47	54.75
		(0.22)	(22.27)	(55.69)
9. Mean # of Candidates Rated	0	0.19	173.68	356.25
		(0.89)	(191.91)	(360.75)
10. Mean # of I/O Graduate Level Work Completed	0	0	7.85	6.75
			(2.13)	(2.82)
11. Percentage with graduate degree in I/O Psychology	0	0	56%	100%

Note: Standard deviations are in parentheses.

participants; 35 individuals agreed to participate, resulting in a response rate of 72%. One respondent, however, had never served as an assessor and was dropped from the sample, resulting in a final sample of 34 trained raters. This sample had previously received a minimum of 24 hours of rater training and worked as assessment center assessors for a faculty member's private business. The majority of the assessors were female (62%), and the mean age of the assessors was 28.88 years (SD = 3.75). Most had either a master's degree (26%) or doctorate degree (30%), and the average number of I/O graduate levels courses completed by the assessors was 7.85 (SD = 2.13). This group had participated in an average of 26.47 (SD = 22.27) assessment centers and had rated, on average, 173.68 assessment center candidates (SD = 191.91). While 44% were employed full-time, their mean tenure was only 1.58 years (SD = 1.34), and their mean supervisory experience was only 1.46 years (SD = 1.90). Also, the assessor sample had little supervisory responsibility (M = 1.15; SD = 2.54), had hired few employees (M = 5.38; SD = 14.42), and had conducted a minimal number of performance appraisals (M = 6.44; SD = 13.43) (see Table 1).

The subject matter expert sample was obtained from a list of the researcher's personal contacts. The criteria for serving as a subject matter expert were having a background in I/O Psychology (i.e., a master's or doctorate degree) and currently being employed in a managerial position with supervisory responsibility. Twelve potential subject matter experts (SMEs) were contacted; nine agreed to serve as SMEs and completed the materials. All the SMEs were offered $50 as compensation for their time. Of these nine respondents, one individual had never been a manager with supervisory

responsibility, and thus, did not meet the criteria for being an SME. Accordingly, this individual was dropped from the study, resulting in a total of eight SMEs. A more complete description of the SME sample and its purpose in this study is presented in the Criterion Measures section (see p. 36).

Stimulus Materials

The stimulus material used in this study consisted of three job descriptions, a packet of assessment summary reports for each job description, definitions of the assessed performance dimensions, and a description of the organization hiring the employees. Except for the organizational description, which was written specifically for the purpose of this study, all other materials were prepared as part of a large-scale assessment center. Any modifications to the original materials are discussed in the following sections.

Assessment Center Description

The assessment center was part of a reorganization effort of a division within a large company in the Southeastern United States. The assessment center was conducted for selection purposes; employees in the affected division had to apply for managerial positions within the new organizational structure. The assessment center was developed and conducted in accordance with the Task Force on Assessment Center Guideline 's published standards (1989). It consisted of a one-day battery of a personal interview and six exercises developed to assess managerial potential of individual candidates. The six exercises were an In-Basket, two Simulation Role Plays, a Simulation Letter, a Case Analysis, and a Small Group Problem-Solving Task (see Appendix A for a complete description of the exercises). Each exercise was designed to elicit behaviors on multiple

dimensions of performance. Two or more experienced assessors recorded individual applicant behaviors on each of the six exercises. Assessors had completed training in assessment center methodology and specific exercises in accordance with *Guidelines and Ethical Considerations for Assessment Center Operations* (1989). At the end of each assessment center, all assessors participated in a consensus meeting in order to integrate the data gathered during the exercises. During this meeting, each dimension was discussed separately for each candidate. Assessors used the resulting consensus ratings to write a developmental report for each candidate. The two I/O psychologists who developed the exercises and directed the assessment center prepared executive summary reports on each candidate; these summary reports were forwarded to the company's hiring decision makers and discussed with them by the I/O psychologists.

Job Descriptions

The three job descriptions used in the study were the same job descriptions used to develop the assessment center exercises. The inclusion of job descriptions was not only to increase the realism of the task, but also to potentially increase the accuracy of subjects' ratings. One line of research in the use of interviews for employee selection has indicated that as interviewers receive more information about a job, inter-rater reliability increases (Langdale & Weitz, 1973), and the impact of irrelevant attributes on decisions is reduced (Weiner & Schneiderman, 1974). Although applicants for seventy-two different jobs were assessed, only three mid-level managerial job descriptions were selected to use in this study. The three jobs were selected on the basis of the number of applicants originally assessed for each position and the specific nature of the KSA's needed to be successful. The first job description selected was that of a human resources

manager; more specifically, the job title was Manager of Organizational Development. The second job description selected was administrative in nature, namely, Manager of Regulatory Guidance, Policy Development, and Compliance. Finally, the third job description was a highly technical position: Manager of Systems Engineering and Special Analysis. Job descriptions were prepared by the organization that contracted for the managerial assessment center. For the purpose of this study, slight changes in wording were made to the job descriptions to match the organizational context of the stimulus task and to preserve the confidentiality of the organization (see Appendix A).

Description of the Organization

Subjects were asked to read a one half-page description of the organization. The description included size of the organization, type of industry, geographical location, mission statement, and reason for filling the positions. The selection task for this research project was framed in the guise of a reorganization effort. Subjects were informed that employees had to reapply for jobs and that the reorganization would result in a reduction in force (see Appendix A). This description was included to increase the realism of the task; individuals charged with selecting employees typically have knowledge about the organization where the employee will work.

Assessment Summary Reports

The reports were the actual executive summary reports prepared for the selection assessment center. Executive summary reports were used in this study, versus lengthy assessment feedback reports provided to assessees, in order to increase the realism of the task. In other words, the subjects in the study were asked to read documents that actual selection committee members reviewed to aid employment decisions. The two I/O

psychologists who developed the exercises and directed the assessment center prepared the executive summary reports. At the time of the assessment center, these individuals had 21 years of combined experience in assessment center methodology and research and 27 years of combined experience in the selection domain.

Each assessment report contained a list of the applicant's Strengths, Competencies, and Weaknesses, based on a rating for each dimension reached by way of consensus by the assessment center staff. Following these listings, a written summary of the applicant's performance was supplied, typically two to four paragraphs in length (see Appendix A).

For the purpose of this research project, the assessment summary reports were altered in four ways. First, the assessees' names were replaced with the words "the applicant" to protect the confidential nature of the data and the assessees' anonymity. Second, because the purpose of this study did not include examining differences related to gender bias, references to gender (e.g., "him," "his," "her," "she," etc.) were altered to be consistent within a job description. That is, for the human resource managerial job, the summary reports were altered so that all the applicants were female; for both the administrative and engineering managerial jobs, the summary reports were altered so that all the applicants appeared to be male. This eliminated gender differences within each job description. The third alteration to the executive summary reports was the omission of the recommendation statements at the end of each write-up (e.g., "We view the applicant as a weak candidate for a managerial position in the organization at this time"). Finally, to limit stimuli to behaviorally oriented statements, any sentences within the summary reports judged to be potentially construed as non-behaviorally oriented or

personality-oriented were altered or eliminated. Out of approximately 235 lines of text contained in the sixteen summary reports, only 8 lines were altered to eliminate potential perception of non-behaviorally based statements.

In order to obtain a set of summary reports with variance in behaviors and ratings, summary reports were collapsed across similar job descriptions. For example, candidates for three different engineering positions were assessed in the assessment center. Because the engineering positions had extremely similar job descriptions, the candidates' summary reports were pooled into one group. A similar pooling of summary reports occurred for the human resource and the administrative job as well. Within each pool, summary reports were selected that represented a wide range of performance, with the overall original consensed ratings ranging from 1.0 – 5.0, on a scale which also ranged from 1.0 – 5.0.

Definitions of Assessed Performance Dimensions

Each performance definition was defined in a 1-2 sentence statement. The definitions were the same used to characterize the dimensions in the assessment center feedback reports and rater training materials (see Appendix A).

Selection Decision Questionnaire

The questionnaire consisted of demographic items and several general items (e.g., years of managerial experience, years serving as an assessor, current job title, etc.) (see Appendix B). Subjects were also instructed to rate each applicant's overall performance in the assessment center based on their understanding of the summary report. The scale used to rate assessment center performance was a 6-point scale, with 1 = *Weak, Poor*, 2 = *Moderate, Marginal*, 3 = *Competent, Satisfactory*; 4 = *Good, Strong*; 5 = *Very Strong,*

Outstanding, and 6 = *Exceptional, Exemplary*. Similarly, subjects were asked to rate the predicted job performance of each applicant using the same 6-point scale. The decision to use a 6-point scale with no midpoint stemmed from the natural ability groupings that occur during assessment center consensus ratings of candidates.

Subjects were also requested to rate each applicant on four dimensional skill sets. These skill sets (i.e., Problem Solving, Organizational Leadership, Communication, and Managerial Skills) mirrored combinations of dimensions analyzed in the assessment center exercises. For example, Problem Solving involves analysis and judgment behaviors, while Communication reflects oral and written communication skills. Organizational Leadership involves leadership, team building, confrontation and initiative behaviors, and Managerial Skills reflects delegation, planning and organizing, coaching, and decisiveness skills. Because dimensions were discussed in the summary reports, asking subjects to rate the applicants' performance on each dimension would have served as a test of reading ability. It was necessary to indirectly gather subjects' ratings on the applicants' performance on specific dimensions in order to examine evaluative accuracy levels. The same 6-point scale used to rate overall assessment center performance and predicted job performance was used to rate the skill sets.

An additional dependent variable was starting salary. Subjects were asked to assume that the applicant would be hired for the position and to recommend a starting salary (Barr & Hitt, 1986; Hitt & Barr, 1989). Current appropriate salary ranges for each position were included in the job descriptions. The salary ranges for the three jobs were as follows: Human Resource Manager, $50,910 to $61,650; Administrative Manager, $41,664 to $53,404; Engineering Manager, $56,150 to $68,960. These ranges were based

on regional compensation data, industry-specific compensation data, and compensation data from professional associations. This pay variable was included to investigate the replicability of past research findings (Barr & Hitt, 1986; Hitt & Barr, 1989).

Each subject was also requested to rank the applicants for each position. More specifically, each subject was asked, "If you could hire only one of the applicants, which applicant would it be?" and instructed to write down the number of their first choice. Then, each subject was asked, "If you first choice declines the job, which applicant is your second choice to fill the job?". Subjects continued choosing applicants until each had been assigned a rank.

Because one of the strengths of assessment center methodology is its objective reporting of behavioral incidents, the summary reports contained no irrelevant or non-behaviorally oriented data; thus, subjects should have used only the dimensional behaviors to make their decisions. In order to test this assumption, subjects were asked to provide their reasons in writing for hiring or not hiring each applicant.

Pilot Study

A pilot study was conducted to examine the characteristics of the stimulus materials. The subjects for the pilot study was a class of undergraduate management students (n = 16). Subjects were invited to complete the stimulus materials in exchange for extra credit. After completing the packet of materials, subjects were asked to respond to several questions about the task (e.g., "The directions for completing the exercise were clear"; "I was able to complete the exercise in 60-90 minutes") using a 5-point scale (1 = Strongly Disagree; 2 = Disagree; 3 = Neither Disagree nor Agree; 4 = Agree; 5 = Strongly Agree) (see Appendix C). Results from these items, in addition to several open-

ended questions, guided several changes made to the stimulus materials. These changes included clarifying task instructions and shortening the task to facilitate its completion within a reasonable time period (i.e., 90 minutes).

In addition, the subjects' ratings of applicants were examined for variance in responses. The pilot study results indicated that subjects' easily identified the lowest performing (i.e., overall ratings of 2.0 or below) and highest performing (i.e., overall ratings of 4.5 or higher) applicants. In fact, one subject suggested removing the extreme cases from the stimulus materials in order to make the task more complex. After some discussion, it was decided that removing several of the extreme cases was appropriate, considering that in real world hiring situations using assessment center data, the identification of very poor and very effective performers is easily accomplished by most individuals. Thus, the elimination of the extreme performers made the task more difficult. The summary reports for two of the managerial positions (i.e., the human resource manager and the administrative manager) represented overall performances that were reasonably similar, though distinctions were present on dimensional performances. In the group of applicants for the engineering position a fuller range of overall performances was represented, from a weak candidate to an exceptional candidate.

In summary, the final stimulus materials included the following: Three summary reports for the human resource position with consensed overall ratings above the performance scale midpoint (i.e., 3.5 – 4.0); five summary reports for the administrative position with consensed overall ratings at or slightly below the performance scale midpoint (i.e., 2.7 – 3.0); and eight summary reports for the engineering position with wide variance in the consensed overall performance ratings (i.e., 2.5 – 5.0).

Criterion Measures

Subject Matter Experts

The criterion measures were the average overall ratings and dimensional ratings assigned to each candidate by a group of subject matter experts. These subject matter experts were currently employed in a managerial position with supervisory responsibility, had earned either a Ph.D. or a Master's degree in Industrial and Organizational Psychology, and had served as raters in assessment centers; thus, these individuals filled a unique role in that they had substantial experience as both managers and as psychologists conducting performance assessments.

The consensed ratings from the assessment center from which the stimuli was drawn were not used for two reasons. First, the I/O psychologists that developed the assessment center materials did not have managerial experience. Second, the I/O psychologists that developed the exercises were also the individuals that trained the assessor group in this study; thus, it is possible that an implicit leadership theory may have been transferred from the I/O psychologists to the assessors through the training. The utilization of a group of subject matter experts external to both the original assessment center and the assessor training process avoided this confound.

On average, the SMEs were male (62%), in their mid-30s (M = 36.25; SD = 4.80), and employed full-time (see Table 1). All the SMEs had either a master's degree (38%) or Ph.D. (62%) in I/O Psychology, and the average number of I/O graduate levels courses completed by the SMEs was 6.75 (SD = 2.82). This group had participated in an average of 57.75 (SD = 55.69) assessment centers and had rated, on average, 356.25 assessment center candidates (SD = 360.75). SMEs, on average, had supervisory

responsibility for 10.00 employees (SD = 9.49), had hired 24.13 employees (SD = 17.52), and had conducted performance appraisals on 22.62 employees (SD = 16.33).

The averaged overall ratings and dimensional ratings of the SMEs for each applicant served as "truth" in this study. That is, the mean of the overall and dimensional ratings obtained from the subject matter experts were considered to reflect true performance levels of the candidates. These true scores were then used to calculate Cronbach's (1955) accuracy components.

Rating Errors Versus Rating Accuracy

For many years, rating errors (e.g., halo, restriction of range, central tendency, etc.) served as the dependent variables in rating research, including studies of rater training (e.g., Bernardin & Pence, 1980) and rating formats (e.g., Blanz & Ghiselli, 1972). Rating error is not a normative assessment of a rater's effectiveness; instead, rating error is determined by a relative comparison among raters (Cardy & Dobbins, 1994). In recent years, however, researchers have pointed out the folly of several assumptions present in using rating errors as criteria for rating effectiveness (Funder, 1987; 1993). First, the presence of rating errors may actually reflect truth. For example, if an organization has a highly effective selection program, it would be expected and desired that restriction of range, halo, and leniency would be present in performance measures (Cardy & Dobbins, 1994). Second, Murphy and Balzer (1989) discussed rating error research's faulty supposition that performance is randomly distributed across ratees. Finally, several studies have found no correlation between measures of rating error and rating accuracy (e.g., Bernardin & Pence, 1980; Borman, 1977; Murphy & Balzer, 1989).

In fact, Bernardin and Pence found that while rater training decreased rating error, it also decreased rating accuracy.

Cronbach's Evaluative Accuracy Components

Due to the building evidence that rating error is not an adequate surrogate for measures of rating effectiveness, more and more researchers have turned to rating accuracy as the dependent variable of interest. Rating accuracy uses external criteria of actual performance levels to examine the characteristics of ratings. Instead of making relative comparisons within a group of raters, raters are judged against an absolute, realistic standard (Cardy & Dobbins, 1994; Funder, 1987). There are two types of accuracy measures: evaluative accuracy that examines the accuracy of compiling information to form summary judgments, and observational accuracy that examines the accuracy of assessment and recall judgments (Cardy & Dobbins, 1994). Because the purpose of this study is to examine the compilation of information to make hiring decisions, this paper will concentrate on evaluative accuracy concepts.

The four core accuracy measures were identified and operationalized by Cronbach in 1955. While other researchers have developed different evaluative accuracy measures (e.g., Borman, 1979; Pulakos, Schmitt, & Ostroff, 1986), it was decided that the Cronbach components were most appropriate for the present study due to the mathematical independence of the four separate accuracy scores that result from Cronbach's calculations (Cardy & Dobbins, 1994). As Cronbach (1955) best states, "...failure to separate these components makes interpretation [of the results] ambiguous" (p. 180). Mathematical definitions of each evaluative accuracy component are presented in Appendix D. It should be noted that Cronbach's evaluative components measure the

magnitude of inaccuracy but do not indicate the direction of the inaccuracy (Cardy & Dobbins, 1994; Cronbach, 1955; Dickinson, 1987). Direction of inaccuracy must be determined by comparing the true performance scores with the average ratings.

Elevation. Elevation is the tendency of a rater to systematically raise or lower ratings across ratees. It is based on the difference between an average rating for a rater and the average true score of the ratees (Cardy & Dobbins, 1994; Dickinson, 1987). While elevation is conceptually similar to leniency and severity error, elevation is determined by comparing the overall rating level to the overall level of actual performance (versus the midpoint of a scale as the criterion for leniency and severity error).

Differential Elevation. Differential elevation represents the propensity of a rater to systematically inflate and deflate ratings of certain individuals, resulting in an inaccurate rank ordering of ratees that differs from their true rank ordering. In other words, differential elevation reflects the accuracy of distinctions made among a group of ratees (Cardy & Dobbins, 1994; Dickinson, 1987). The presence of differential elevation inaccuracy is the bane of personnel selection; its presence may result in hiring an unsuitable candidate and reducing the validity of selection procedures.

Stereotype Accuracy. Stereotype accuracy denotes a rank ordering of the performance dimensions across a group of ratees. Stereotype accuracy is the tendency of a rater to systematically inflate or deflate dimension ratings, so that the rank ordering of various strengths and weaknesses of all individuals within a group is inaccurate. In other words, stereotype accuracy assesses how precisely a rater makes distinctions among performance dimensions (Cardy & Dobbins, 1994; Dickinson, 1987). For example, a

rater may always deflate ratings on the leadership dimension or inflate ratings of decisiveness due to some archetype of what behaviors define effective leadership and coaching; this inaccuracy permeates across all individuals within a group of ratees.

Differential Accuracy. In simplest terms, differential accuracy is a reflection of the ratee by dimension interaction (Murphy & Balzer, 1989). Cardy and Dobbins (1994) referred to differential accuracy as the "amount of noise in the rating that is not accounted for by an overall rating tendency or by rating tendencies regarding certain ratees or dimensions" (p. 36). Differential accuracy is what is left after removing the rater's elevation, differential elevation, and stereotype accuracy components, and it represents the best indication of how effectively a specific rater can judge patterns of performance across both ratees and dimensions (Cardy & Dobbins, 1994). Differential accuracy is essential when providing feedback to an individual concerning his or her skills and abilities.

Procedure

Managerial subjects and assessors were mailed the study materials. Undergraduate subjects were administered in a small group setting of 6 to 8 participants per group. The purpose and the tasks of the study were explained to all subjects in a one-page letter. After reading the study overview, subjects were asked to sign a statement of informed consent. Subjects were then instructed to review the materials for the first job description, read the first applicant's assessment summary report, and then complete the selection decision questionnaire for that specific applicant. Once the selection decision questionnaire was completed for the first applicant, subjects were instructed to continue reading the applicant information in a sequential manner. Subjects were also asked to

complete a page of demographic items of interest. Most subjects finished the selection decision task within 75 minutes.

Statistical Analyses

Preliminary Analyses

In order to justify the aggregation of data within the subject matter expert group, it was necessary to determine the level of interrater agreement for each dimension rating (e.g., overall assessment center performance, predicted job performance, managerial skills, organizational leadership skills, problem solving skills, and communication skills). While several indices for calculating interrater agreement are available, James, Demaree, and Wolf's (1984; 1993) r_{wg} method was selected for this study. This statistic is a measure of within-group consensus for a set of raters judging a target with either a single-item or a multiple item scale, with the judgments being made on a discrete response format (Kozlowski & Hattrup, 1992). The r_{wg} statistic compares the mean observed variance to the expected variance of the discrete response format. Because the selection decision questionnaire items were not collapsed to create a single measure, the r_{wg} single-item estimator was used. See James et al. (1984; 1993) for a complete discussion of the r_{wg} statistic.

In addition, following the recommendations of Boyatzis (1998), Krippendorf (1980) and Weber (1990), subjects' open-ended comments were independently coded as behaviorally-based or non-behaviorally-based by subject matter experts (i.e., the author and two Applied Psychology Ph.D. students from another university).

Main Analyses

As a test of Hypothesis 1 (i.e., the examination of the relationship between subject groups' assessment center performance ratings with SME true scores) zero-order correlations were calculated. Subjects was asked to characterize every applicant's overall performance in the assessment center on a scale from 1 to 6 (e.g., 1 = Weak, Poor to 6 = Exceptional, Exemplary). Each subject's rating of applicants' assessment center performance were then correlated with the corresponding SME true score. This resulted in 16 data pairs being used to generate a correlation for each subject. Correlations were then averaged within each subject group; that is, an average correlation was computed for the student sample, the manager sample, and the trained assessor sample.

Zero-order correlations were also used to test Hypothesis 2 (i.e., the relationship between subject groups' predicted job performance ratings and SME true scores). Subjects was asked to rate each applicant's predicted on-the-job performance, using a scale from 1 to 6 (e.g., 1 = Weak, Poor to 6 = Exceptional, Exemplary). Subjects' ratings of each applicant's predicted job performance were then correlated with the corresponding SME true score. As with Hypothesis 1, an average correlation was then computed for the student sample, the manager sample, and the trained assessor sample.

In order to ascertain differences in the rating accuracy of the subject groups (i.e., Hypotheses 3 and 4), Cronbach's (1955) components of evaluative accuracy were computed. Averaged SME ratings served as true scores, and mean differences among the groups were examined using a one-way analysis of variance procedure. The various groups were then compared using a Scheffé multiple comparison test. The Scheffé statistic was chosen because of its conservative nature and relative insensitivity to groups

of unequal size. In addition, the Scheffé test is appropriate when only a few paired comparisons are examined (Hayes, 1988).

Hypothesis 5 (i.e., group differences in the types of assessment center information used to make rating decisions) were tested by way of a dominance analysis (Budescu, 1993). Dominance analysis is a relatively new procedure that examines both the unique variance and any partial joint variance contributed by each predictor in a multiple regression equation. A pairwise comparison of all predictors is produced, and one variable is said to dominant another if it is a stronger predictor in all subset regressions (Budescu, 1993). Additionally, this statistical procedure allows for the identification of variables that are of equal importance, as well as variables that cannot be rank-ordered in any statistically significant way (Kudisch, 1996). A dominance analysis was performed to ascertain the relative importance of the four dimensional skill ratings (i.e., Problem Solving, Organizational Leadership, Communication, and Managerial Skills) in making predictions concerning the applicants' on-the-job performance. In other words, dominance analysis allowed for the examination of differences among the subject groups in their use of assessment performance information in making hiring decisions.

Hypothesis 5 was also tested by examining subjects' open-ended comments explaining their reasoning for hiring or not hiring an applicant. Each subjects' open-ended themes were dummy-coded, and an analysis of variance procedure was performed to test for differences in thematic use across the three subject groups. A Scheffé multiple comparison test was then conducted to compare the groups.

Finally, group differences in starting salary recommendations (i.e., Hypothesis 6) were determined using a one-way analysis of variance procedure. Group comparisons were made using the Scheffé statistic.

CHAPTER IV

RESULTS

Preliminary Analyses

In order to ascertain the appropriateness of using the SME's averaged ratings as true scores, it was necessary to first assess their level of interrater agreement. A single-item r_{wg} estimate was calculated for each stimulus summary (n = 16) on each of six selection decision questionnaire items. Thus, 96 r_{wg} statistics were computed for the SME group. In order to calculate r_{wg}, it was decided that the expected variance (i.e., σ_e^2) was best represented by the variance of a discrete, uniform normal distribution (i.e., σ_e^2 = 2.917 for a 6-point discrete scale) (James et al., 1993). All estimates of interrater agreement were within acceptable range (i.e., .70 or higher) (see George, 1990) and ranged from .82 - 1.00, with a mean estimate of .96. Therefore, it was determined that an acceptable level of agreement was present in the SME ratings. Consequently, the SME ratings were averaged and utilized as true scores in the study.

The complete set of open-ended comments were divided among the author and two additional SMEs, so that each comment would be coded by the author and one other SMEs. Each subject matter expert had been previously trained on content analysis techniques and had previous experience conducting content analysis. First, as suggested by Boyatzis (1998), Krippendorf (1980) and Weber (1990), a coding taxonomy was developed. Each SME read a random selection of open-ended comments and independently generated broad themes that provided the basis of the coding taxonomy. After sharing the individual lists, the final themes to be used in coding the subjects'

responses were decided. Two broad themes were generated: Behavioral Rationale and Non-Behavioral Rationale. Within these categories, several distinct sub-themes emerged that described more specific rationale (see Table 2).

The author and the SMEs then independently catergorized subjects' rationale into a theme and sub-theme. Interrater agreement was calculated and percent agreement was determined. The hit rate was 93% for the author and one of the SMEs; the hit rate for the author and the second SME was 92%; these percent agreement indices indicate adequate agreement between the author and the SMEs. In the few cases where the author and the SMEs did not agree on a coding, the open-ended comment was discussed until a consensus was reached regarding the appropriate theme categorization for that comment.

Main Analyses

Hypothesis 1: Assessment Center Performance Ratings Among the Subject Groups

It was hypothesized that trained assessors' ratings of assessment center performance would be the most highly correlated with true scores, followed by managers' ratings of assessment center performance. In order to test this prediction, zero-order correlations were computed for each subject. Subjects' ratings of each applicant's assessment center performance were correlated with the corresponding SME true score. This resulted in 16 data pairs being used to generate a correlation for each subject. A mean correlation was then computed for each subject group (e.g., the managers' correlations were averaged, resulting in a mean manager correlation, etc.).

The trained assessors' mean correlation was .739 (SD = 0.109; range = .446 - .919); the managers' mean correlation was .728 (SD = 0.123; range = .378 - .912); the undergraduate students' mean correlation was .687 (SD = 1.141; range = .307 - .900).

Table 2. Descriptive Statistics for Content Analysis Themes

Themes	Student n (%)	Manager n (%)	Assessor n (%)	Sample Comment
Behavioral Rationale	697 (75%)	464 (75%)	363 (84%)	
General behaviors and skills	386 (41%)	194 (31%)	177 (41%)	Because he has the ability to lead others and reach clear, decisive goals.
Training recommendations	96 (10%)	64 (10%)	21 (5%)	His lack of decisiveness could be coached through. Will need some attention to ensure others respect his decision.
Integration of job description with performance	208 (22%)	192 (31%)	164 (38%)	Oral communication and leadership don't seem to be crucial for this job, but he still seemed pretty average.
Other	7 (1%)	14 (2%)	1 (0.2%)	I prefer a decisive leader that can confront others.
Non-Behavioral Rationale	235 (25%)	155 (25%)	68 (16%)	
General statements	116 (12%)	84 (14%)	42 (10%)	Very qualified.
Inferential statement	44 (5%)	11 (2%)	3 (1%)	She sounds like a kiss up. Job requires a spine.
Training recommendation	14 (2%)	21 (3%)	3 (1%)	This person will perform well, provided a mentor is assigned to her.
Integrates job description with performance	52 (6%)	30 (5%)	14 (2%)	Good skill fit with job.
Other	9 (1%)	9 (2%)	6 (1%)	If she were the best of a group, I would hire her, but I suspect I could keep looking for someone better.

Thus, although all the groups had a strong, positive relationship with the true scores, the trained assessors' and managers' mean correlations demonstrated a similar level of relationship. The mean correlation of the student group was slightly lower (see Table 3). Fisher's r-to-z-transformation was then applied to all of the correlations, and an ANOVA was then conducted to determine if there were significant differences in the correlations across the groups. The results of the ANOVA indicated that the differences were non-significant ($F = 1.905$; $p < .153$). Thus, Hypothesis 1 was not supported.

Hypothesis 2: Predicted Job Performance Ratings Among the Subject Groups

Hypothesis 2 stated that trained assessors' ratings of predicted job performance would have the strongest correlation with predicted job performance true scores, followed by managers' ratings. Again, zero-order correlations were computed for each subject: Subjects' ratings of each applicant's predicted job performance were correlated with the corresponding SME true score, resulting in 16 data pairs used to generate each correlation. A mean correlation was then computed for each subject group.

While all groups had a strong correlation with the true scores, contrary to prediction, managers' ratings had the strongest relationship (mean $r = .767$; $SD = 0.133$; range = .365 – .933). Trained assessors' mean correlation (mean $r = .697$; $SD = 0.145$; range = .364 - .896) and undergraduate students' mean correlation (mean $r = .708$; $SD = 0.142$; range = .358 - .913) were essentially the same and indicated a somewhat weaker relationship with the true scores (see Table 4). Additionally, an ANOVA was conducted to determine if there were significant differences in the correlations across the groups; however, the differences were non-significant ($F = 1.177$; $p < .311$). These results generated no support for Hypothesis 2.

Table 3. Descriptive Statistics for Mean Correlations of Assessment Center Performance Ratings With SME True Scores

Group	n	Range	M	SD	% $p < .05$
Student	59	.307 - .900	.687	0.141	81%
Manager	42	.378 - .912	.728	0.123	90%
Trained Assessor	34	.446 - .919	.739	0.109	97%

Table 4. Descriptive Statistics for Mean Correlations of Predicted Job Performance Ratings With SME True Scores

Group	n	Range	M	SD	% $p < .05$
Student	59	.358 - .913	.708	0.142	88%
Manager	42	.365 - .933	.767	0.133	93%
Trained Assessor	34	.364 - .896	.697	0.145	88%

Hypothesis 3a, 3b, 3c, & 3d: Assessor Accuracy Versus Manager and Student Accuracy

Descriptive statistics for elevation, differential elevation, stereotype accuracy, and differential accuracy for trained assessors, managers, and students are shown in Table 5. One-way analysis of variance results for the accuracy components among the subject groups are shown in Table 6. The analysis results indicated that there were significant differences between the groups with regard to levels of elevation ($F = 15.184$; $p < .001$), differential elevation ($F = 3.673$; $p < .028$), stereotype accuracy ($F = 4.528$; $p < .012$), and differential accuracy ($F = 7.784$; $p < .001$) (see Table 6).

Hypothesis 3 predicted that trained assessors would make more accurate ratings than managers and students. A Scheffé multiple comparison test was conducted to examine the sub-hypotheses. Results of this analysis indicated that trained assessors were more accurate than students in terms of elevation ($M_{diff} = 0.415$; $p < .05$), differential elevation ($M_{diff} = 0.107$; $p < .05$), and differential accuracy ($M_{diff} = 0.112$; $p < .05$). However, the mean difference in level of stereotype accuracy between trained assessors and undergraduates only approached statistical significance ($M_{diff} = 0.018$; $p < .10$). When compared to managers, trained assessors' level of elevation and differential elevation were not significantly different from managers' level of elevation and differential elevation (see Table 7). In addition, mean differences between the trained assessor and manager groups for stereotype accuracy ($M_{diff} = -0.007$; $p < .10$) approached statistical significance, but did not reach $p < .05$ level of significance. Furthermore, this relationship was in the opposite direction of the prediction, with managers' ratings being more accurate.

Table 5. Descriptive Statistics for Accuracy Components

	Students (n = 59)		Managers (n = 42)		Assessors (n = 34)	
	M	SD	M	SD	M	SD
Elevation	0.650	0.650	0.161	0.301	0.235	0.258
Differential elevation	0.344	0.215	0.321	0.199	0.237	0.077
Stereotype accuracy	0.063	0.054	0.038	0.030	0.045	0.035
Differential accuracy	0.400	0.134	0.350	0.154	0.288	0.094

Note. All values are based on Cronbach's (1955) deviational components of accuracy, for which a value of zero represents perfect accuracy.

Table 6. Summary of ANOVA for Accuracy Components

Source	df	SS	MS	F	
			Elevation		
Group	2	6.995	3.498	15.184	$p < .001$
Error	132	30.406	0.230		
Corrected total	134	37.401			
			Differential Elevation		
Group	2	0.251	0.126	3.673	$p < .028$
Error	132	4.511	0.034		
Corrected total	134	4.762			
			Stereotype Accuracy		
Group	2	0.017	0.008	4.528	$p < .012$
Error	132	0.244	0.002		
Corrected total	134	0.261			
			Differential Accuracy		
Group	2	0.272	0.136	7.784	$p < .001$
Error	132	2.303	0.017		
Corrected total	134	2.574			

Note: df = degrees of freedom; SS = Type III Sum of Squares; MS = Mean Square

Table 7. Scheffé Multiple Comparison for Accuracy Components

Comparison Groups	M_{diff}	SE	95% Confidence Interval	
			Lower Bound	Upper Bound
Elevation				
Student-Manager	0.489*	0.109	0.3068	0.6732
Student-Assessor	0.415*	0.118	0.2202	0.6198
Manager-Assessor	-0.074	0.067	-0.1866	0.0466
Differential Elevation				
Student-Manager	0.023	0.043	-0.0486	0.0946
Student-Assessor	0.107*	0.039	0.0421	0.1719
Manager-Assessor	0.084	0.037	-0.0224	0.1456
Stereotype Accuracy				
Student-Manager	0.033*	0.009	0.0134	0.0466
Student-Assessor	0.018	0.010	-0.0934	0.1266
Manager-Assessor	-0.007	0.008	-0.0066	0.0266
Differential Accuracy				
Student-Manager	0.050	0.029	-0.0015	0.0984
Student-Assessor	0.112*	0.026	0.0950	0.1560
Manager-Assessor	0.062	0.030	-0.0451	0.1128

Note: M_{diff} = mean difference; SE = standard error
*$p < .05$

Differences in assessors' and managers' levels of differential accuracy ($M_{diff} = 0.084$; $p < .10$) also approached statistical significance, but did not reach $p < .05$ level of significance, with trained assessors' ratings being more accurate. Hence, Hypotheses 3a, 3b, and 3d were only partially supported; no support was found for Hypothesis 3c.

Hypothesis 4a, 4b, 4c, & 4d: Manager Accuracy Versus Student Accuracy

Hypothesis 4 predicted that managers would make more accurate ratings than undergraduate students. Again, a Scheffé multiple comparison test was conducted to examine the sub-hypotheses. As expected, managers were more accurate than undergraduate students in terms of elevation ($M_{diff} = 0.489$; $p < .05$) and stereotype accuracy ($M_{diff} = 0.033$; $p < .05$). However, contrary to predictions, the mean difference in managers' and undergraduate students' differential elevation ($M_{diff} = 0.023$) and differential accuracy ($M_{diff} = 0.050$) only approached significance ($p < .10$) (see Table 6). In other words, while managers made more accurate overall levels of ratings and more accurate distinctions among dimensions within each applicant, both managers and students made similarly accurate evaluations of the rank order of applicants and of each applicant's strengths and weaknesses. In summary, support was found for Hypothesis 4a and 4c; Hypothesis 4b and 4d were not supported.

Hypothesis 5: Types of Information Used to Making Hiring Decisions

The four skill dimensions (i.e., Problem Solving Skills, Organizational Leadership Skills, Managerial Skills, and Communication Skills) rated for each applicant by all subjects were used as predictors of predicted job performance ratings in a dominance analysis. Dominance, or relative importance, is judged by comparing each pair of predictors across all submodels, not including empty matrix cells (Kudisch, 1996).

Appendix E presents the squared multiple correlations of all the possible models involving the predictors, by each subject group (Budescu, 1993). Table 8 presents summary results of the dominance analysis, including R^2 values and importance figures for the predictors.

It was hypothesized that subject groups would use different types of information to make their hiring decisions. The dominance analysis showed that this prediction was supported (see Table 8). Overall (i.e., across the three job descriptions and all applicants), the subject groups assigned different importance to the various skill sets in determining applicants' on-the-job performance. By examining the weighted mean semi-partial correlations (Cx_i), and then calculating a percentage of total variance accounted for by each predictor, a summary of the mean usefulness of the various predictors was produced. For the student group, the four skill sets (e.g., Problem Solving, Organizational Leadership, Communication, and Managerial Skills) accounted for 73% of the variance in predicted job performance ratings. The students' ratings of Managerial Skills accounted for the most unique variance (30%) followed by Communication Skills (26%) and Organizational Leadership Skills (25%), and finally Problem Solving Skills (19%) (see Table 8). For the manager group, the four skill sets accounted for 78% of the variance in ratings of predicted job performance. Unique variance was contributed by Organizational Leadership (32%), Managerial Skills (31%), Communication Skills (23%) and Problem Solving (14%) (see Table 8). Finally, for the trained assessor group, the four skill sets accounted for 68% of the variance in predicted job performance ratings. Trained assessors focused on the dimensional skill ratings in a slightly different order (see Table 8). For this group, Managerial Skills (33%) and

Table 8. Dominance Analysis Summary: R^2 and Importance of Dimensions

	R^2	Prob	Lead	Comm	Mgr
Overall					
SME	.7924	.1026	.2644	.1111	.3143
Student	.7304	.1368	.1854	.1886	.2196
Manager	.7839	.1060	.2533	.1787	.2459
Assessor	.6752	.2070	.2158	.0965	.2231
HR Job					
SME	.7030	.0777	.3415	.0059	.2779
Student	.6172	.1085	.1972	.1400	.1715
Manager	.5757	.0639	.2348	.0790	.1980
Assessor	.5647	.1060	.2089	.0738	.1760
Admin Job					
SME	.8204	.1588	.2379	.1016	.3221
Student	.6347	.1087	.1566	.1698	.1996
Manager	.6883	.1125	.2187	.1892	.1679
Assessor	.5642	.1142	.1866	.0886	.2048
Engineer Job					
SME	.8358	.1015	.2371	.1946	.3026
Student	.8098	.1641	.1911	.2244	.2302
Manager	.8143	.1155	.2555	.2105	.2628
Assessor	.7154	.1387	.2198	.1293	.2276

Note: Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills. Sample sizes are as follows: SME (n = 8); Student (n = 59); Manager (n = 42); Assessor (n = 34).

Organizational Leadership (32%) accounted for the most unique variance, followed by Problem Solving (21%), and Communication Skills (14%).

An additional dominance analysis was computed within group for each job type. The results of this analysis show that the subject groups used different types of data within each job type to predict on-the-job performance of the applicants (see Table 8), providing additional support for Hypothesis 5. A summary of the rank order of the relative importance of the predictor variables is provided in Table 9.

Human Resource Job. For the HR job, students' ratings of the predictors of on-the-job performance accounted for 62% of the total variance and were in the following order: Organizational Leadership contributed 32% of the unique variance, Managerial Skills contributed 27%, followed by two of equal importance, Communication (27%) and Problem Solving Skills (18%) (see Table 8). Managers' ratings accounted for 58% of the total variance, and the rank order of on-the-job performance predictors was similar to the students' rankings (Organizational Leadership, 41% of variance accounted for; Managerial Skills, 34%; Communication, 14%; Problem Solving 12%) (see Table 8). Trained assessors' ratings accounted for 56% of the variance, and their strongest predictor was also Organizational Leadership (37%), followed by two equally important predictors (Managerial Skills, 31%; Problem Solving, 19%), and finally, Communication Skills (13%) (see Table 8).

Administrative Job. For the Administrative job, students' dimensional ratings accounted for 62% of the total variance. Their strongest predictor was Managerial Skills, contributing 31% of the total variance accounted for, followed by two predictors of equal importance (Communication, 27%; Organizational Leadership, 25%), and lastly, Problem

Table 9. Rank Order of the Relative Importance of Predictor Variables

Group	Prob	Lead	Comm	Mgr
SMEs (n = 8)				
Across all jobs	3rd (tie)	2nd	3rd (tie)	1st
HR job	2nd (tie)	4th	1st	2nd (tie)
Administrative job	3rd (tie)	2nd	3rd (tie)	1st
Engineering	4th	2nd (tie)	2nd (tie)	1st
Students (n = 59)				
Across all jobs	4th	2nd (tie)	2nd (tie)	1st
HR job	3rd (tie)	1st	3rd (tie)	2nd
Administrative job	4th	2nd (tie)	2nd (tie)	1st
Engineering	3rd (tie)	3rd (tie)	1st (tie)	1st (tie)
Managers (n = 42)				
Across all jobs	4th	1st	3rd	2nd
HR job	4th	1st	3rd	2nd
Administrative job	4th	1st (tie)	1st (tie)	1st (tie)
Engineering	4th	2nd	3rd	1st
Trained assessors (n = 34)				
Across all jobs	3rd	2nd	4th	1st
HR job	2nd (tie)	1st	4th	2nd (tie)
Administrative job	3rd (tie)	2nd	3rd (tie)	1st
Engineering	3rd (tie)	1st (tie)	3rd (tie)	1st (tie)

Note: Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Solving (17%) (see Table 8). Managers' ratings accounted for 69% of the total variance accounted for. They had three predictors of equal importance (Organizational Leadership, contributing 32% of the total variance accounted for; Communication, contributing 27%; and Managerial Skills, contributing 24%), followed by Problem Solving (16% of the total variance accounted for) (see Table 8). Ratings made by trained assessors accounted for 56% of the total variance. The most important predictor for trained assessors was Managerial Skills (34%), followed by Organizational Leadership (31%), and then Problem Solving (19%) and Communication (15%) (see Table 8).

Engineering Job. For the Engineering job, students' dimensional skill ratings accounted for 81% of the total variance. Students had two sets of equally important predictors (see Table 8): Managerial (28%) and Communication Skills (28%) each contributed 28% to the variance accounted for, followed by Organizational Leadership (24%) and Problem Solving (20%). Managers' ratings also accounted for 81% of the total variance, with their strongest predictor of on-the-job performance being Managerial Skills (32%), followed by Organizational Leadership (31%), then Communication (26%), and lastly, Problem Solving (14%) (see Table 8). Trained assessors' ratings accounted for 72% of the total variance. Assessors had two sets of equally important predictors (see Table 8): Managerial Skills (32%) and Organizational Leadership (31%), followed by Problem Solving (19%) and Communication Skills (18%).

A second test of Hypothesis 5 was performed with subjects' coded open-ended responses. Each subjects' open-ended responses were dummy-coded into themes, and then a score for each theme was computed. The themes "Other" under Behavioral and Non-Behavioral Rationale were not carried forward in this analysis due to their very

small sample size (see Table 2). An ANOVA was conducted to test for differences in type of information use across the subject groups. Table 10 presents the results of the ANOVA. The results of the open-ended analysis, taken in sum, provide additional support for Hypothesis 5.

More specifically, beginning with the themes subsumed under Behaviors Rationale, there were no significant differences between in the subject groups in the frequency of the General Behaviors and Skills theme ($\underline{F} = 2.205$; $\underline{p} < .115$). There were significant differences in frequency of occurrence among the groups for the Training Recommendations theme ($\underline{F} = 3.799$; $\underline{p} < .025$). A Scheffé multiple comparison test revealed that both students ($\underline{M}_{diff} = 0.822$; $\underline{p} < .05$) and managers ($\underline{M}_{diff} = 0.964$; $\underline{p} < .05$) offered specific, behaviorally-based training recommendations more frequently than assessors (see Table 11). There were also significant differences between the groups for the Integration of Job Description with Performance theme ($\underline{F} = 5.090$; $\underline{p} < .008$). The Scheffé multiple comparison test indicated that assessors ($\underline{M}_{diff} = 2.332$; $\underline{p} < .05$) offered significantly more integrative statements than students. While assessors offered more integrative rationale than managers ($\underline{M}_{diff} = 1.514$), the mean difference between the two groups, while approaching significance, failed to reach the $\underline{p} < .05$ level.

For the fours themes under Non-Behavioral Rationale, two themes failed to reach a critical F value. Both the General Statements theme ($\underline{F} = 0.725$; $\underline{p} < .486$) and the Integration of Job Description with Performance theme ($\underline{F} = 0.983$; $\underline{p} < .377$) were not significant. However, the ANOVA showed significant F values for both the Inferential Statements theme ($\underline{F} = 6.574$; $\underline{p} < .002$) and Training Recommendations theme ($\underline{F} = 5.216$; $\underline{p} < .007$) (see Table 10). A Scheffé multiple comparison test revealed that

Table 10. Summary of ANOVA for Content Analysis Themes

Source	df	SS	MS	F	
		Behavioral Rationale			
General Behaviors and Skills					
Group	2	54.95	27.48	2.205	$p < .115$
Error	119	1482.75	12.46		
Corrected total	121	1537.70			
Training Recommendations					
Group	2	18.01	9.01	3.799	$p < .025$
Error	119	282.19	2.37		
Corrected total	121	300.20			
Integration of Job Description with Performance					
Group	2	103.35	51.68	5.090	$p < .008$
Error	119	1208.03	10.15		
Corrected total	121	1311.38			
		Non-Behavioral Rationale			
General Statements					
Group	2	6.79	3.40	0.725	$p < .486$
Error	119	557.22	4.68		
Corrected total	121	564.01			
Inferential Statements					
Group	2	10.17	5.08	6.574	$p < .002$
Error	119	92.04	0.77		
Corrected total	121	102.21			
Training Recommendations					
Group	2	3.85	1.92	5.216	$p < .007$
Error	119	43.93	0.37		
Corrected total	121	47.78			
Integration of Job Description with Performance					
Group	2	2.77	1.38	0.983	$p < .377$
Error	119	167.34	1.41		
Corrected total	121	170.11			

Note: df = degrees of freedom; SS = Type III Sum of Squares; MS = Mean Square

Table 11. Scheffé Multiple Comparison for Content Analysis Themes

Comparison Groups	M_{diff}	SE	95% Confidence Interval	
			Lower Bound	Upper Bound
		Behavioral Rationale		
Training recommendations				
Student-Manager	-0.087	0.344	-0.6608	0.4864
Student-Assessor	0.822*	0.313	0.3004	1.3427
Manager-Assessor	0.964*	0.321	0.4302	1.0984
Integration of job description with performance				
Student-Manager	-0.817	0.578	-1.7804	0.1456
Student-Assessor	-2.332*	0.689	-3.4800	-1.1838
Manager-Assessor	-1.514	0.825	-2.8880	0.1407
		Non-Behavioral Rationale		
Inferential statements				
Student-Manager	0.517*	0.200	0.8500	1.0344
Student-Assessor	0.639*	0.209	0.3485	0.9871
Manager-Assessor	0.122	0.104	-0.0513	0.2943
Training recommendations				
Student-Manager	-0.334*	0.136	-0.5605	-0.01077
Student-Assessor	0.130	0.095	-0.0283	0.2887
Manager-Assessor	0.464*	0.153	0.2096	0.7190

Note: M_{diff} = mean difference; SE = standard error
*$p < .05$

students made significantly more inferential judgments than either managers (M_{diff} = 0.517; $p < .05$) or assessors (M_{diff} = 0.639; $p < .05$) (see Table 11). The Scheffé test of the Training Recommendation theme indicated that managers offered significantly more general training recommendations than students (M_{diff} = 0.334; $p < .05$) or assessors (M_{diff} = 0.464; $p < .05$).

Hypothesis 6: Differences in Starting Salary

Hypothesis 6 stated that undergraduate students and trained assessors would recommend higher starting salaries than managers. Due to the different salary ranges outlined for each job type, each job type was separately examined for this hypothesis. For each subject, only the salary of the selected applicant for each of the three jobs was used as the dependent measure; a mean salary for each job was then computed for within each subject group. The descriptive statistics for these variables are listed in Table 12. A one-way analysis of variance was performed (see Table 13). While the groups recommended significantly different salaries for the Human Resource job ($F = 3.408$; $p < .036$), there were no differences in recommended starting salary levels for the Administrative job (F = 1.157; $p < .318$) or the Engineering job ($F = 0.845$; $p < .432$). However, a Scheffé test for the Human Resource job (see Table 14) failed to reach significance for any of the group comparisons. Thus, Hypothesis 6 was not supported.

Table 12. Descriptive Statistics for Salary Variables

Group	Human Resource	Administrative	Engineering
	M (SD)	M (SD)	M (SD)
Undergraduate	$53,726 ($2165)	$47,124 ($3477)	$61,917 ($4000)
Manager	$52,584 ($2874)	$46,451 ($3504)	$61,642 ($4148)
Trained Assessor	$53,957 ($2602)	$46,026 ($3470)	$60,800 ($3805)

Note: M = mean; SD = standard deviation; all numbers rounded to the nearest whole dollar; for Undergraduates, n-size varied from 56-59 due to missing values; for Managers, n-size varied from 40-41 due to missing values; for Trained Assessors, n-size = 34.

Table 13. Summary of ANOVA for Salary Level

Source	df	SS	MS	F	
			HR Job		
Group	2	4.31×10^7	2.15×10^7	3.408	$p < .036$
Error	128	8.09×10^8	6.32×10^6		
Corrected total	130	8.52×10^8			
			Administrative Job		
Group	2	2.81×10^7	1.40×10^7	1.157	$p < .318$
Error	131	1.59×10^9	1.21×10^7		
Corrected total	133	1.62×10^9			
			Engineering Job		
Group	2	2.70×10^7	1.35×10^7	0.845	$p < .432$
Error	128	2.05×10^9	1.60×10^7		
Corrected total	130	2.07×10^9			

Note: df = degrees of freedom; SS = Type III Sum of Squares; MS = Mean Square

Table 14. Scheffé Multiple Comparison for Human Resource Salary

Comparison Groups	M_{diff}	SE	95% Confidence Interval	
			Lower Bound	Upper Bound
Student-Manager	1141.98	518.48	-142.12	2426.08
Student-Assessor	-231.12	544.70	-1580.16	1117.91
Manager- Assessor	-1373.10	586.35	-2825.30	79.10

M_{diff} = mean difference; SE = standard error

CHAPTER V

DISCUSSION

The selection literature has examined several different aspects of rating accuracy, including the effects of rater training and rater experience on rating accuracy. However, little consensus has been reached concerning the type of experience essential to making accurate selection decisions. Past assessment center research has concentrated on the accuracy of initial ratings made by raters when observing ratees' performance. How accurately assessment center performance information is used by different groups to make selection decisions has been virtually ignored by researchers. Given the rising use of the assessment center method to select employees, it is essential to understand the degree to which professional and managerial experience impacts the interpretation of assessment center data. Therefore, the main purpose of this study was to better understand how experience in management and I/O Psychology impacts selection accuracy. To this end, several hypotheses based on past selection and performance appraisal research, as well as the expert decision making literature, were examined. A secondary purpose of the current study was to examine the type of information being used across varying groups making selection decisions, in hopes of shedding light on differences in the decision making process. Lastly, the current study attempted to address several methodological limitations of past research, including the need to utilize more generalizable samples, to study more complex decision making tasks, and to examine the selection of managerial-level employees.

This chapter begins with a discussion of the study results and implications of the research findings. The second section of this chapter addresses the practical implications of the study, followed by a discussion of the study limitations. Finally, future directions for research are presented.

Discussion of Hypotheses

Assessment Center and Predicted Job Performance Ratings Among the Subject Groups

No significant differences were found between the subject groups' mean correlations with true scores. It is likely that the small n size accounted for the lack of significant findings. Each correlation was computed on a sample of 16 data points, a distribution that does not approach normality. However, the trends in the correlational data suggest that trained assessors' and managers' ratings of overall assessment center performance had a stronger relationship with true applicant performance than undergraduates' ratings. Moreover, managers' ratings of predicted job performance had a stronger relationship with true scores than either trained assessors' or undergraduates' ratings. These trends suggest that managers' experience may better equip them to assess how well an applicant could perform on the job and that managerial experience was instrumental in the accurate interpretation of managerial assessment center results. However, these conclusions are highly tentative.

Differences in Rating Accuracy Across Subject Groups

Elevation. Consistent with the study's predictions, students' ratings were significantly more elevated than either managers' ratings or trained assessors' ratings. However, contrary to the study's predications, there were no significant differences between managers' and trained assessors' elevation accuracy. Although all three groups

systematically inflated their ratings across all applicants (i.e., none of the groups had a perfect elevation accuracy of 0), the students' ratings where substantially more elevated than either managers' or assessors' ratings. These results are supportive of past selection research and multiple studies in the performance appraisal literature that show students make more lenient ratings than managers (Barr & Hitt, 1986; Bernstein et al., 1975; Dipboye et al., 1975). However, the present findings also indicate that lack of managerial experience is not the only explanation for elevation inaccuracy in ratings; trained assessors, although without significant managerial experience, did not elevate their ratings any more than managers. The assessors' specialized training and extensive practice in making ratings may explain these findings. In essence, the present findings suggest that practice in making performance ratings is an effective calibration experience, whether that experience is garnered through serving as a rater in an assessment center or as a manager participating in the company's performance appraisal system.

Differential Elevation. Differential elevation represents the propensity of a rater to generally and systematically inflate or deflate ratings of certain individuals, resulting in an inaccurate overall rank ordering of ratees. The results of this study indicate that students were significantly less accurate than trained assessors in terms of differential elevation. However, students' and managers' ratings contained similar amounts of differential elevation. Differences in managers' and trained assessors' level of differential elevation approached significance, with assessors being slightly more accurate. These findings suggest that undergraduate students provided the least accurate rank ordering of applicants while trained assessors provided the most accurate overall rank ordering. Perhaps untrained or lesser trained raters overly focus on a weakness or a

strength and this "halo" permeates all of their ratings, resulting in lower levels of differential elevation accuracy. Better trained raters may be able to use the strengths and weaknesses to inform their decisions yet not allow themselves to excessively focus on that information, resulting in more accurate dimensional differentiation.

Additionally, these results may be explained by the common experience of the SMEs and the trained assessors in their professional training. The SMEs and trained assessors that participated in this study most likely had great commonalties in their graduate school instruction and assessment center experience. More specifically, similar topical areas are covered across I/O graduate programs (e.g., individual differences, performance appraisal, etc.), and the methodology of different assessment centers is highly similar. For example, in the majority of managerial assessment centers, although specific stimuli are unique, the types of exercises used and the behavioral dimensions assessed are comparable. The similarities in background experience of the SMEs and the trained assessors could account for the similarity in their ratings.

Subjects were also asked to rank order the applicants for each job description. When these rankings were examined, the vast majority of subjects' top-ranked applicant for each job description agreed with the subject matter experts' top-ranked applicant. For example, 85% of students, 76% of managers, and 88% of trained assessors selected the top applicant for the engineering job. However, while the subject groups were able to pick out the "best" applicants for the jobs, the selection of second- and third-most qualified applicants varied greatly across subject groups. More specifically, trained assessors were more likely than the managers and undergraduates to correctly rank the second and third place applicants for the HR and administrative positions. However, for

the engineering position, undergraduates and assessors were more likely than managers to correctly select the second-place candidate, while managers were more likely than the other groups to correctly select the third-place candidate.

It is possible that the subject groups' ability to select the top candidate is a result of "transparent" applicant performance; that is, perhaps the applicants' performance was so disparate, it was immediately clear to the subjects who was the best candidate. However, as a result of the pilot study, most of the applicants with extreme performance (i.e., either at the very low end or the very high end of the performance scale) were removed from the stimulus materials, resulting in a more ambiguous task requiring complex discrimination. Perhaps the managers and undergraduates focused only on selecting the top applicant for each job position, but failed to maintain a similar level of focus when making selection decisions about other applicants.

To summarize, while there was general consensus among the groups regarding the "best" candidate for the job, trained assessors made better decisions regarding other qualified applicants. In addition, assessors were significantly more accurate than students and marginally more accurate than managers in terms of differential elevation. The presence of differential elevation inaccuracy is at the heart of personnel selection; its presence may result in hiring an unsuitable candidate. Although all subject groups appeared capable of selecting the number one candidate for the job, appropriately ranking the second- and third-ranked applicants appeared to be more difficult for the students and managers than the assessors. This finding is especially salient given that during times of low unemployment it is not uncommon to hire the second- or third-most qualified applicant for a position.

Stereotype Accuracy. Stereotype accuracy is the tendency of a rater to systematically inflate or deflate dimension ratings across individuals. As expected, students' ratings had the lowest stereotype accuracy; that is, of the three subject groups, students were less likely to correctly make relative distinctions among performance dimensions. Managers' and trained assessors' level of stereotype accuracy were significantly better than students'; however, differences between managers and assessors only approached a level of significance ($p < .10$), with managers being slightly more accurate. These findings suggest that undergraduates are not as capable as managers and assessors at distinguishing differences in dimensions across applicants. In addition, the difference between the managers and assessors suggests that assessors may not be as capable as managers at distinguishing dimensional differences in performance. These findings suggest that managerial experience may play an important role in developing an accurate image of "an effective manager". Individuals with little managerial experience may have a stereotype of an effective manager that is erroneous, while experienced managers' stereotype of an effective manager may be more realistic.

Differential Accuracy. Differential accuracy represents the best indication of how effectively a specific rater can judge patterns of performance across both ratees and dimensions (Cardy & Dobbins, 1994). In terms of differential accuracy, results indicate that while students' ratings were less accurate than trained assessors' ratings, students and managers had a similar level of differential accuracy. However, differences in managers' and trained assessors' accuracy levels only approached statistical significance ($p < .10$), with trained assessors being slightly more accurate. This implies that trained assessors were significantly better than students at making dimensional distinctions

within and across applicants. In addition, the trend in the data also suggests that trained assessors were slightly better than managers in making accurate dimensional × applicant distinctions. Findings of the present research indicate that trained assessors are better than students, and possibly better than managers, at assessing the strengths and weaknesses of individuals. Differential accuracy is essential when providing feedback to an individual concerning his or her skills and abilities. The findings of this study suggest that trained assessors appear to be more effective at providing an individual with accurate developmental feedback and performance.

Types of Information Used to Make Hiring Decisions

The results indicate that there were indeed differences in the types of information used among the groups to make their selection decisions. This was determined through both quantitative and qualitative investigation. Dominance analysis was utilized to assess the differences in the relative importance of predictors among the subject groups, while thematic content analysis was used to examine differences in rationale provided in support of subject groups' selection decisions.

Dominance Analysis. The results of the dominance analysis indicate that the subject groups did in fact use the information provided in different ways when making their selection decisions. More specifically, the skill sets accounted for different amounts of variance in predicted job performance ratings across the subject groups. For the HR job, the total variance accounted for was similar across the three groups (i.e., student: $\underline{R^2}$ = .617; manager: $\underline{R^2}$ = .578; assessor: $\underline{R^2}$ = .565). However, for the administrative and engineering jobs, the R^2 values varied among the subject groups. For the administrative job, students' ratings accounted for 63% of the variance, versus managers' ratings

accounting for 69% and assessors' ratings accounting for 56%. For the engineering job, students' ratings accounted for 81% of the variance; managers' ratings accounted for 81%; assessors' ratings accounted for 72% of the variance. The lower levels of total variance accounted for in the trained assessors' ratings may indicate assessors were concentrating on other pieces of information, like individual dimension descriptions, exercise examples, and performance extremes within an applicant's description, or perhaps they were more sensitive to interactions among dimensions.

The relative importance for each skill set also varied across the subject groups. In general terms, managers appear to have discounted managerial skills relative to other subjects groups and attended more to leadership skills. Managers also seem to have placed less emphasis on problem solving skills than other groups. However, none of the groups strongly attended to problem solving. A possible explanation for this is that all the applicants were described as meeting the minimum qualifications for the job. Subjects may have assumed that problem solving was at least a competency for all the applicants and thus, minimized the role this skill set played in their decisions.

These findings support past research (Galbraith & Alvares; 1983; Hakel, Dobmeyer, & Dunnette, 1970; Sackett & Hakel, 1979) that indicated certain dimensions are used more often than others in forming overall judgments, and the use of specific dimensions is consistent across raters. However, the results of the dominance analysis do not support Russell's (1985) findings that a single factor typically dominated ratings. Instead, the dominance analysis clearly indicates no one skill set accounted for the majority of the variance in the ratings. More significantly, the amount of emphasis placed on the four skill sets varied by subject group, indicating that the groups were

attending to different information when making their selection decisions. These findings speak to the fact that making selection decisions is a very complex task that eschews simplistic decision-making rules.

However, it seems reasonable to expect that the skill sets might be weighted differently for each managerial position. In other words, if the three jobs require differing amounts of the four skill sets, and subjects are using the job descriptions to inform their decisions, then differences in the weighting of the skill sets should emerge. Perhaps, though, the job descriptions were not distinct enough, which may account for the similarity in weightings within subject groups and the fact that no one skill set accounted for the majority of the variance in the ratings. To examine the possibility of the job descriptions being very similar, 11 doctoral students in an applied psychology program rated each job description on several dimensions.

First, the raters were asked to rate the importance of each skill set for performing effectively in each job. The scale used to rate the importance of each skill set was 5-point scale, with 1 = *Unimportant to this job*; 2 = *Of minor importance to this job*; 3 = *Important to this job*; 4 = *Very important to this job*; and 5 = *Of critical importance to this job.* Next, the raters were asked to assess the similarity of the three jobs, again using a 5-point scale (i.e., 1 = *Extremely dissimilar* to 5 = *Extremely similar*).

Separate analyses of variance were conducted on the importance ratings of the skill sets across the three jobs. Communication Skills was the only skill set with significant differences in importance across jobs ($F = 29.51$; $p < .001$), although differences in ratings of Problem Solving approached statistical significance ($F = 3.24$; $p < .053$). Both Organizational Leadership ($F = 0.30$ $p < .742$) and Managerial Skills ($F =$

0.321; $p < .728$) failed to attain significance. Additionally, the raters indicated that the jobs were similar in nature ($M = 3.11$; $SD = 0.93$; $n = 9$ due to missing responses). The results of this post hoc test indicates that the job descriptions were similar in nature, which could explain why subject groups weighted the skill sets in a similar manner across the three jobs and why no skill set dominated the ratings for any job.

Thematic Content Analysis. The results suggest that both students and managers offered more behaviorally-based, specific training recommendations for the applicants than did the trained assessors. In addition, managers offered significantly more general training recommendations than either the students or the assessors. However, trained assessors offered more behaviorally-based integrative rationale (e.g., rationale that integrated the job description with applicants' performance) than students. This pattern was the same between assessors and managers, though the difference only approached statistical significance. Finally, the student group made substantially more inferential statements than either the manager group or the assessor group.

It appears then that the managers offered rationale from a developmental viewpoint. That is, the managers were more likely to focus on future possibilities and the skills applicants needed to acquire in order to fulfill the job requirements. In contrast, the trained assessors tended to focus on the dichotomous decision to hire or not hire applicants; they were less likely to examine applicants' potential to develop into a position but instead focused on the "here and now" of the selection decision. The differences in the managers' and trained assessors' viewpoints could be attributed to a few things. The vast majority of trained assessors' practical experience was with selection-oriented assessment centers, and their role within these centers was to make

competency determinations based on concrete, behavioral observations, not on "what if's" and potential performance. On the other hand, during times of low unemployment, managers may be willing to hire an individual who meets the minimum qualifications if they feel an applicant has the ability to improve his or her skill deficiencies.

Subject matter experts' comments were examined on a post hoc basis to investigate whether their rationale patterns had similarities with either the managers or the trained assessors. While the number of comments gathered from the SMEs was not sufficient (n = 121) to allow for significance testing, the percent frequency of themes was explored to ascertain patterns in their responses. Interestingly, unlike the managers, the subject matter experts offered few behaviorally-based training recommendations (10%; n = 12). Instead, the largest percentage of their comments were behaviorally-based integrative statements that linked applicants' assessment center performance with the job description (38%; n = 46), similar to the pattern found in the trained assessor sample.

Differences in Starting Salary

Results of the study indicate that the were no statistically significant differences among the groups in level of starting salaries of the selected applicants. This was in direct contrast to the predicted relationship and to past research findings (Barr & Hitt, 1986; Hitt & Barr, 1989). Prior research found that inexperienced decision makers (i.e., undergraduates) recommended considerably higher starting salaries than experienced decision makers (i.e., managers). Due to trained assessors' lack of familiarity with this task, it was predicted that they would also recommend significantly higher starting salaries than managers; however, the trained assessors' salary suggestions were not statistically distinct from the managers' salary figures.

One potential explanation for this finding is that subjects were not given enough contextual information to provide a starting salary. In fact, several managers and trained assessors indicated that additional information was needed before a starting salary could be suggested, including work history, past salary level, and information on the company's compensation system. Another explanation for this set of findings is that the salary ranges provided in the stimulus materials were too narrow. The lower and upper ends of all three salary ranges differed by approximately $11,000. This range may have been too restricted to allow for group differences to emerge. A reexamination of Barr and Hitt's (1986) and Hitt and Barr's (1989) studies showed that their stimulus salaries covered a $25,000-$35,000 range, a figure significantly higher than the current study's salary ranges.

Practical Implications

Several implications for practitioners can be drawn from the results. First, the results of the present research speak to the strength of the assessment center as a selection methodology: the presentation of behavioral information, free from inferences and irrelevant cues, allows people regardless of type and level of experience to identify the top candidate for a position. Managers, without any specialized training in performance assessment, should be able to adequately interpret assessment center summary data and pick the best applicant. However, if there is not a clear "winner" among the candidates, managers may not be as accurate as trained assessors in rank-ordering applicants. This is especially relevant in a selection context, where the top candidate may not accept the job, and offers must be made to candidates further down the list. It seems plausible that

involving individuals with I/O psychology expertise in selection situations will help to ensure the soundness of selection decisions.

Interestingly, although managers appear to think in terms of developmental needs for job applicants, they may not be able to use assessment center data to accurately determine the strengths and weaknesses of an individual. The data trends indicate that trained assessors are slightly better than managers at using assessment center data to distinguish strengths and weaknesses of job candidates. However, the results of the content analysis indicated that assessors may not articulate these distinctions, making developmental feedback next to impossible. In addition, assessors may need to learn how to ascertain applicants' ability to learn new skills and appropriately distinguish between a skill deficiency that is a fatal flaw versus a skill deficiency that can be trained. For example, managerial skills like delegation, planning, organizing, coaching, and decisiveness, may be easier to learn on the job, while problem solving (i.e., analysis and judgment) are skills that may be more difficult to learn, if these skills can be learned at all. Perhaps assessors should be involved in determining the strengths and weaknesses of an individual, and then allow managers to step in to determine the appropriate action plans for improving deficient areas. This type of integration between managers and assessors appears very salient as managerial coaching becomes increasing popular with companies (Flilipczak, 1998).

Training programs that target specific groups is another practical indication of these research findings. Although training is never a substitute for experience, it seems that assessors and managers could benefit from learning about each others' areas of

knowledge. This may look like behavioral assessment training for managers and managerial decision making instruction for assessors.

Study Limitations

There are several limitations associated with the present study that warrant discussion. These limitations can be grouped into three areas: the nature of the sample, the nature of the task, and the nature of the criterion.

Sample Limitations

The first topic of concern is the limited sample size within each subject group. While the sample sizes were large enough to support means testing between the groups (Hayes, 1988), the samples were small enough to limit the generalizability of the study. In addition, none of samples were randomly drawn from their respective populations: the undergraduates were psychology students who volunteered for extra credit; the managers were mostly alumni from an executive MBA program; the trained assessors were current graduate students or recent graduates of an I/O Psychology Ph.D. program. Furthermore, all three samples were from the same large, southeastern public university. In addition, the undergraduate sample was primarily female (78%). However, this percentage approximates the proportion of females that typically volunteer to participate as subjects in psychology research studies at the aforementioned public university (Lawler, personal communication, November 2, 1999).

Another limitation is the self-selection process of the managerial sample. Managers were solicited for their participation via a mailing; there was no opportunity to collect demographic data on non-responders to compare them with those managers that did respond. In addition, while the managers that responded represented a wide range of

industries and companies, they also represented a wide range of managerial responsibility, as indicated by the skewed distribution of several demographic variables. More specifically, the mean number of employees supervised had a large standard deviation (M = 207.19; SD = 804.13), as did the mean number of employees hired (M = 53.36; SD = 69.69) and the mean number of performance appraisals conducted (M = 178.50; SD = 423.85).

This segues into a final limitation of the sample. An assumption was made that there was no within-cell variance in performance of the different subject groups' knowledge domains. Unfortunately, there was no feasible way to measure actual KSAs or expertise levels of subject groups or of the subject matter experts. In fact, it should be assumed that within-cell variance was present, and that subject groups consisted of a mix of novices, experts, and "practiced novices" (i.e., individuals that have more skills than the average person, but have yet to reach the level of expert) (Shanteau, 1992).

Task Limitations

The primary task limitation was the use of a paper-people task in the stimulus materials. The paper-person study paradigm has been criticized for its artificiality, especially when researching situations where personal contact with the target person is the norm (Gorman et al., 1978). It is extremely rare that an applicant for a managerial position would be hired without an interview. Even as internet-based application procedures become more prevalent, some form of personal interaction typically takes place between the job candidate and the hiring body prior to a job offer being made.

However, the advantage of the paper-people paradigm is that information cues can be tightly controlled, and researchers can manipulate the order of cues, the strength of

cues, etc. (Gorman et al., 1978). The stimulus materials for the present study were designed to minimize the presence of non-behavioral cues. For example, the purpose of this study was to examine the use of behavioral data in making selection decisions, not to examine the effects of applicant demographic variables on hiring decisions. To this end, the sex of applicants was altered to be consistent within each applicant pool, and other demographic information (e.g., age and race) was not provided. The attempt was made to provide purely behaviorally-based information cues, something that is impossible to control with the use of person-to-person interactions.

A related task limitation is the restricted amount of information provided about applicants' past work experiences, salary levels, or other assessment information (e.g., biodata, cognitive ability tests, etc.). Again, while the researcher wanted to limit the number of extraneous cues subjects could use to inform their decisions, the restrictive nature of the stimulus materials may have limited the realism of the task. However, in order to maximize realism, a complex task was developed: subjects were asked to make hiring decisions on numerous qualified applicants for three different managerial jobs. In addition, assessment center summary reports from an actual managerial assessment center were used as stimulus material. These summary reports were virtually the same paragraphs given to an extant selection committee who them used the data to inform their decisions. The resultant complex task more closely approximated a real-world selection task than past selection research. Past studies have typically examined very simple decision processes (e.g., to hire or not hire an applicant) and the use of interview transcripts (Barr & Hitt, 1984; Hitt & Barr, 1986; Gorman et al., 1978; Landy & Bates, 1983).

Another task limitation that should be discussed is the potential of a fatigue factor for the subjects completing the stimulus materials. Each subject was asked to read a description of an organization, definitions of the assessed performance dimensions and assessment exercises, and three job descriptions. Subjects were then requested to read an assessment summary report for each of 16 applicants and make a set of ratings for each applicant. Clearly, this was a very time-consuming and laborious assignment, and it is possible that attention to the task waned after a certain amount of time. This might explain the higher R^2 for the engineering job in the dominance analysis. The total variance accounted for by the skill set ratings was significantly higher for this managerial position across all three subject groups. However, fatigue did not appear to differentially impact the accuracy of the groups' ratings across job positions. An ANOVA was run to look for a job main effect, but no main effect for job type was found.

Finally, a last task limitation is the possibility of an order effect in the stimulus materials. All of the subject groups received the same set of materials, with the applicant ordering being consistent across all groups. While a better research design might have been to randomize the presentation of the applicants within each job description and to also randomize the presentation of the job descriptions, obtaining the necessary sample size for this type of design was not feasible.

Criterion Limitations

Criterion measures were the average overall ratings and dimensional ratings assigned to each candidate by a group of subject matter experts. To serve as a subject matter expert, it was necessary to meet two criteria: (1) a master's or doctorate degree in I/O Psychology and (2) current employment in a managerial position with supervisory

responsibility. An assumption was made that the experience level of these I/O psychologists/managers equated to expertise, similar to the assumption made regarding the experience levels of the managers and trained assessors. Once again, there was no feasible way to measure actual expertise levels of these subject matter experts. Thus, the SME's "truth" may not have been the best standard by which to compare the subject groups' performance. Additionally, there may have been a better way to select the experts needed for this study. Shanteau (1988) suggests that referrals by colleagues is the preferred way to identify experts for decision making studies, versus relying on knowledge tests, level of education, or job titles. However, in applied psychology research, a "subject matter expert" traditionally means an individual with specialized knowledge and/or experience performing a certain task (Borman, 1977; Lancy & Vasey, 1991). For example, in the realm of job analysis, a subject matter expert is an incumbent who performs effectively on the job or a supervisor familiar with the content of the job (Cascio, 1991). Thus, the use of a group of managers with graduate degrees in I/O psychology appears to be a sensible standard of comparison when the questions being researched concern the type of experience needed to make appropriate hiring decisions.

Avenues for Future Research

Keeping the aforementioned limitations in mind, future research should target several areas. As with any study, attempts at replicating the salient findings of the present research should occur. Other avenues for future research can be segmented into several categories. Funder (1993) has provided an excellent organizing framework for considering the different moderators of judgment accuracy, and components of his

framework have been borrowed to discuss the different research questions that should to be pursued.

Good Decision Maker

What are the properties of a good decision maker? There is some evidence that experience in rating individuals' performance improves the quality of judgments (Cardy et al., 1987; Landy & Bates, 1973; Lorenzo, 1984). However, do accurate judges have other characteristics that significantly impact the accuracy of their decisions? For example, maybe characteristics typically associated with expertise (e.g., selective attention, complex problem simplification, appropriate weighting of irrelevant information) (Shanteau, 1988) are necessary to make accurate decisions. This does not necessary imply that all hiring decision makers need to be experts, but perhaps the presence of certain expert characteristics allow experienced novices to make highly accurate decisions that approach the decisions of true experts. For example, Cardy & Kehoe (1984) found that raters high in selective attention made more accurate performance ratings. Ascertaining which characteristics have the greatest influence on a judge's accuracy appears to be a promising area of future research.

This idea segues into an additional avenue for research, that of developing and testing training that attempts to enhance characteristics associated with expertise. It is not so much what experts know, but how they put the information together and the inferences they draw that separate them from the novices. In fact, training programs that have focused on improving certain expert characteristics (e.g., discounting irrelevant information) has been successful in increasing accuracy (Gaeth & Shanteau, 1984). In addition, several frame-of-reference training endeavors have produced promising

evidence that accuracy can be increased by utilizing this method (Athey & McIntyre, 1987; Pulakos, 1984; Stamoulis & Hauenstein, 1993). While frame-of-reference training researchers do not explicitly state that they are attempting to enhance a characteristic found in expert decision makers, the very nature of frame-of-reference training serves to calibrate individuals, and calibration is a skill necessary for expert decision making to occur (Lichtenstein, Fischhoff, & Phillips, 1977; Shanteau, 1992). Although some characteristics associated with experts, like self-confidence, creativity, and handling adverse situations, may be untrainable (Shanteau, 1984), determining which cognitive structures can be modified by training appears to be a promising research arena.

Additionally, it may be that until we fully understand what information is attended to and how the information is weighted, it will be difficult, if not impossible, to effectively design training to improve decision making. For example, Funder and Sneed (1992) found that judges used some invalid behavioral cues to inform their decisions, while other valid behavioral cues were attended to, but not used to inform their decisions. Past studies have indicated that different groups attend to different information and weight information differently (Maederer, 1993; Mullins, 1982; Singer & Bruhns, 1991), implying that unique training programs may be needed for specific groups of people. Furthermore, maybe these differences in perception and use of information reflect different implicit theories decision makers hold for the job in question. There is already some evidence (Hauenstein & Alexander, 1991) that this may the case, but additional studies are needed to provide further insight into this issue.

Good Information

Maybe the blame for poor quality judgments lies not with the decision maker, but with the information available to him or her. Research is called for that examines how to maximize the quality of information used in making selection decisions. What does "better" information look like? Past studies have indicated that a larger quantity of information typically enhances decision making (Blackman & Funder, 1998; Harvey & Lozada-Larsen, 1988; Langdale & Weitz, 1973;). However, other studies have shown that judges that use only a limited number of cues to make their decisions (Galbraith & Alvares, 1983; Russell, 1985; Sackett & Hakel, 1977; Slovic & Lichtenstein, 1971). The selection arena would be well-served to reexamine these seemingly contradictory conclusions. Maybe "quantity" does not always equal "quality", implying that the usefulness of different types of information needs to be established.

Another aspect of examining the quality of information is to increase the realism of hiring decision tasks by giving more complete applicant information (e.g., interview, personality measures, work history, biodata, selection test scores). Then, by varying the amount and type of information, an evaluation can be made of whether the presence of assessment center data in association with other selection data enhances decision accuracy. This is currently an under-researched area. For example, perhaps the objective, behaviorally-based nature of assessment center data counteracts the potential bias of job-irrelevant information (e.g., race, age, gender) that can occur during the ubiquitous selection interview.

Good Assessee

Perhaps some people are easier to judge than others (Funder, 1993). It is conceivable that it is easier for a decision maker to make accurate predications about an individual demonstrating high behavioral consistency. For example, it may be easier to forecast an applicant's on-the-job performance when he or she consistently exhibits effective interpersonal skills across situations and tasks. No studies could be found that examined this concept. Additionally, it is feasible that an applicant's personality may aid or hinder accurate assessment. For example, it can be more difficult to elicit relevant behaviors from a reticent individual during an assessment center, and this may impact his or her final evaluation; however, the individual's reserved nature may have little impact on effectively performing the job. In fact, in the area of personality judgment, Colvin (1993) found that individuals with higher levels of extraversion, agreeableness, conscientiousness, and emotional stability were easier for judges to accurately rate. These findings have possible implications for the development of assessment methods to better elicit relevant behaviors from various "difficult-to-assess" personality types.

Additional Future Research

One additional research area is related to Cronbach's evaluative accuracy components. There appears to be an implicit assumption that these indices are normally distributed, and no studies could be located that examined this question. Because the calculation of Cronbach's accuracy indices involve squared terms, it would seem that a Chi-square-like distribution, rather than a normal distribution, might more closely approximate the dispersion of the terms. However, study is needed before any conclusions can be made.

Conclusion

Given the rising use of the assessment center method to select employees, it is essential to understand the degree to which professional and managerial experience impacts the accuracy of interpreting assessment center results. However, few studies have examined the utilization of assessment center data in making selection decisions. Building on research from the assessment center, performance appraisal and expertise literature, the present study extends previous selection decision making research by examining the role of professional and managerial experience in interpreting and using managerial assessment center data. The current study sheds some light on the role different types of experience plays in making accurate decisions. In addition, practical suggestions emanate for the specific role professional and managerial decision makers should take within a selection decision task. As organizations continue to use assessment center data to inform their selection decisions, research exploring how to increase the accuracy of selection judgments will remain important.

REFERENCES

REFERENCES

Anderson, J. R. (1981). Cognitive skills and their acquisition. Hillsdale, NJ: Erlbaum.

Anderson, J. R. (1982). Acquisition of cognitive skill. Psychological Review, 89, 369-406.

Athey, T. R., & McIntyre, R. M. (1987). Effect of rater training on rater accuracy: Levels-of-processing theory and social facilitation theory perspectives. Journal of Applied Psychology, 72, 567-572.

Barr, S. H., & Hitt, M. A. (1986). A comparison of selection decision models in manager versus student samples. Personnel Psychology, 39, 599-617.

Bazerman, Max. H. (1986). Judgment in managerial decision making. New York: John Wiley & Sons, Inc.

Bernardin, H. J., & Buckley, M. R. (1981). Strategies in rater training. Academy of Management Review, 6, 205-212.

Bernardin, H. J., & Pence, E. C. (1980). Rater training: Creating new response sets and decreasing accuracy. Journal of Applied Psychology, 65, 60-66.

Bernstein, V., Hakel, M. D., & Harlan, A. (1975). The college student as interviewer: A threat to generalizability? Journal of Applied Psychology, 60, 266-268.

Blackman, M. C., & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. Journal of Experimental Social Psychology, 34, 164-181.

Blanz, F., & Ghiselli, E. E. (1972). The mixed scale: A new rating system. Personnel Psychology, 25, 185-199.

Bolger, F., & Wright, G. (1992). Reliability and validity in expert judgment. In G. Wright & F. Bolger (Eds.), Expertise and decision support (pp. 47-76). New York: Plenum Press.

Borman, W. C. (1977). Format and training effects on rating accuracy and rater errors. Journal of Applied Psychology, 64, 410-421.

Boyatzis, R. E. (1998). Transforming qualitative information: Thematic analysis and code development. Thousand Oaks, CA: Sage Publications, Inc.

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).

Budescu, David V. (1993). Dominance analysis: A new approach to the problem of relative importance of predictors in multiple regression. Psychological Bulletin, 114 (3), 542-551.

Byham, W. C. (1977). Assessor selection and training. In J. L. Moses & W. C. Byham (Eds.), Applying the assessment center method (pp. 89-126). New York: Pergamon Press.

Cardy, R. L., Bernardin, H. J., Abbott, J. G., Senderak, M. P., & Taylor, K. (1987). The effects of individual performance schemata and dimension familiarization on rating accuracy. Journal of Occupational Psychology, 60, 197-205.

Cardy, R. L. & Dobbins, G.H. (1994). Performance appraisal: Alternative perspectives. Cincinnati, OH: South-Western Publishing Co.

Cardy, R. L., & Kehoe, J. F. (1984). Rater selective attention ability and appraisal effectiveness: The effect of a cognitive style on the accuracy of differentiation among ratees. Journal of Applied Psychology, 69, 589-594.

Carlson, R. E. (1967). Selection interview decisions: The effect of interviewer experience, relative quota situation, and applicant sample on interviewer decisions. Personnel Psychology, 20, 259-280.

Cascio, W. F. (1991). Applied psychology in personnel management (4th ed.). Englewood Cliffs, NJ: Prentice-Hall, Inc.

Chase, W. G., & Simon, H. A. (1973). The mind's eye in chess. In W. G. Chase (Ed.), Visual information processing (pp. 215-281). New York: Academic Press.

Chi, M. T. H., Glaser, R., & Farr, M. J. (1988). The nature of expertise. Hillsdale, NJ: Erlbaum.

Colvin, C. R. (1993). "Judgable people": Personality, behavior, and competing explanations. Journal of Personality and Social Psychology, 64, 861-873.

Cronbach, L. J. (1955). Processes affecting scores on understanding of others and assumed similarity. Psychological Bulletin, 52, 177-193.

deGroot, A. D. (1965). Thought and choice in chess. The Hague: Mouton.

Dickinson, T. L. (1987). Designs for evaluating the validity and accuracy of performance ratings. Organizational Behavior and Human Decision Processes, 40, 1-21.

Dipboye, R. L. (1990). Laboratory vs. field research in I/O psychology. International Review of Industrial and Organizational Psychology, 5, 1-34.

Dipboye, R. L., & Flanagan, M. F. (1979). Research settings in industrial and organizational psychology: Are findings in the field more generalizable than in the laboratory? American Psychologist, 60, 39-43.

Dipboye, R. L., Fromkin, H. L., & Wiback, K. (1975). Relative importance of applicant sex, attractiveness, and scholastic standing in evaluation of job applicant resumes. Journal of Applied Psychology, 60, 39-43.

Dobbins, G. H., Lane, I. M., & Steiner, D. D. (1988a). A note on the role of laboratory methodologies in applied behavioral research: Don't throw out the baby with the bath water. Journal of Organizational Behavior, 9, 281-286.

Dobbins, G. H., Lane, I. M., & Steiner, D. D. (1988b). A further examination of student babies and laboratory bath water: A response to Slade and Gordon. Journal of Organizational Behavior, 9, 377-378.

Dolan, J. G., Bordley, D. R., & Mushlin, A. I. (1986). An evaluation of clinicians' subjective prior probability estimates. Medical Decision Making, 6, 216-223.

Dugan, B. (1988). Effects of assessor training on information use. Journal of Applied Psychology, 73, 743-748.

Ebbesen, E., & Konecni, V. (1975). Decision making and information integration in the courts: The setting of bail. Journal of Personality and Social Psychology, 32, 805-821.

Einhorn, H. J. (1974). Expert judgment: Some necessary conditions and an example. Journal of Applied Psychology, 59, 562-571.

Ericsson, K., & Simon, H. A. (1980). Verbal reports on data. Psychological Review, 87, 215-251.

Filipczak, B. (1998, March). The executive coach: Helper or healer? Training, 35, 30-36.

Funder, D. C. (1987). Errors and mistakes: Evaluation the accuracy of social judgment. Psychological Bulletin, 101, 75-90.

Funder, D. C. (1993). Judgments as data for personality and developmental psychology: Error versus accuracy. In D. C. Funder (Ed.), Studying lives through time: Personality and development (pp.121-146). Washington, D. C.: American Psychological Association.

Funder, D. C., & Sneed, C. D. (1993). Behavioral manifestation of personality: An ecological approach to judgmental accuracy. Journal of Personality and Social Psychology, 64, 479-490.

Gaeth, G. J., & Shanteau, J. (1984). Reducing the influence of irrelevant information on experienced decision makers. Organizational Behavior and Human Performance, 33, 263-282.

Galbraith, J. T. (1985). Training assessment center assessors: Applying principles of human judgment (Doctoral dissertation, Bowling Green State University, 1984). Dissertation Abstracts International, 45, 3104-B.

Galbraith, J.T., & Alvares, R. M. (1983). The influence of candidate race and sex, and number of dimensions on overall assessment center ratings. Unpublished manuscript, Bowling Green State University.

Gatewood, R. D., & Feild, H. S. (1994). Human resource selection (3rd ed.). Orlando, FL: Harcourt Brace & Company.

Gaugler, B. B., Rosenthal, D. B., Thornton, G. C., III, & Bentson, C. (1987). Meta-analysis of assessment center validity. Journal of Applied Psychology, 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.

George, J. M. (1990). Personality, affect, and behavior in groups. Journal of Applied Psychology, 75, 107-116.

Gordon, M. E., Slade, L. A., & Schmitt, N. (1986). The "science of the sophomore" revisited: From conjecture to empiricism. Academy of Management Review, 11, 191-207.

Gorman, C. D., Clover, W. H., & Doherty, M. E. (1978). Can we learn anything about interviewing real people from "interviews" of paper people? Two studies of the external validity of a paradigm. Organizational Behavior and Human Performance, 22, 165-192.

Greenwood, J. M., & McNamara, W. J. (1969). Leadership styles of structure and consideration and managerial effectiveness. Personnel Psychology, 22, 141-152.

Grove, D. A. (1981). A behavioral consistency approach to decision making in employment selection. Personnel Psychology, 34, 55-64.

Guion, R. M. (1983). Editorial comments from the new editor. Journal of Applied Psychology, 68, 547-551.

Hakel, M. D., Dobmeyer, T. W., & Dunnette, M. D. (1970). Relative importance of three content dimensions in overall suitability ratings of job applicants' resumes. Journal of Applied Psychology, 54, 65-71.

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.

Harvey, R. J., & Lozada-Larsen, S. R. (1988). Influence on amount of job descriptive information on job analysis rating accuracy. Journal of Applied Psychology, 73, 457-461.

Hauenstein, N. M. A., & Alexander, R. A. (1991). Rating ability in performance judgments: The join influence of implicit theories and intelligence. Organizational Behavior and Human Decision Processes, 50, 300-323.

Hayes, W. L. (1988). Statistics (4th ed.). Orlando, FL: Holt, Rinehart and Winston, Inc.

Hedge, J. W., & Kavanagh, M. J. (1988). Improving the accuracy of performance evaluations: Comparison of three methods of performance appraiser training. Journal of Applied Psychology, 73, 68-73.

Hitt, M. A., & Barr, S. H. (1989). Managerial selection decision models: Examination of configural cue processing. Journal of Applied Psychology, 74, 53-61.

James, L. R., Demaree, R. G., & Wolf, G. (1984). Estimating within-group interrater reliability with and without response bias. Journal of Applied Psychology, 69, 85-98.

James, L. R., Demaree, R. G., & Wolf, G. (1993). r_{wg}: An assessment of within-group interrater agreement. Journal of Applied Psychology, 78, 306-309.

Kahneman, D., & Tversky, A. (1972). Subjective probability: A judgment of representativeness. Cognitive Psychology, 3, 430-454.

Kahneman, D., & Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80, 237-251.

Klimoski, R. S., & Strickland, W. (1977). Assessment centers: Valid or merely prescient. Personnel Psychology, 30, 353-363.

Kozlowski, S. W. J., & Hattrup, K. (1992). A disagreement about within-group agreement: Disentangling issues of consistency versus consensus. Journal of Applied Psychology, 77, 161-167.

Krippendorf, K. (1980). Content analysis: An introduction to its methodology. Beverly Hills, CA: Sage.

Kudisch, J. D. (1996). Factors related to participants' acceptance of developmental assessment center feedback (Doctoral dissertation, University of Tennessee, Knoxville, 1996). Dissertation Abstracts International, 58(6-B), 3349.

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., & Vasey, J. (1991). Job analysis: The composition of SME samples. Personnel Psychology, 44, 27-50.

Langdale, J. A., & Weitz, J. (1973). Estimating the influence of job information on interviewer agreement. Journal of Applied Psychology, 57, 23-27.

Lichtenstein, S., Fischhoff, B., & Phillips, L. D. (1977). Calibration of probabilities: The state of the art. In H. Jungermann & G. de Zeeuw (Eds.), Decision making and change in human affairs (pp. 306-334). Amsterdam: D. Reidel.

Lorenzo, R. V. (1984). Effects of assessorship on managers' proficiency in acquiring, evaluating, and communicating information about people. Personnel Psychology, 37, 617-634.

Maederer, C. M. (1993). What is beautiful is good...to an inexperience rater: The effect of rater experience on personnel selection decisions (Doctoral dissertation, University of South Florida, 1993). Dissertation Abstracts International, 54(4-B), 2256.

Meyer, R. E. (1983). Thinking, problem solving, cognition. New York: Freeman.

Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. Psychological Review, 63, 81-97.

Mullins, T. W. (1982). Interviewer decisions as a function of applicant race, applicant quality and interviewer prejudice. Personnel Psychology, 35, 163-174.

Murphy, K. R., & Balzer, W. K. (1989). Rater errors and raring accuracy. Journal of Applied Psychology, 74, 619-624.

Nagy, G. F. (1981). How are personnel selection decisions made? An analysis of decision strategies in a simulated personnel selection task. Unpublished doctoral dissertation, Kansas State University.

Olian, J. D., Schwab, D. P., & Haberfeld, Y. (1988). The impact of applicant gender compared to qualifications on hiring recommendations: A meta-analysis of

experimental studies. Organizational Behavior and Human Decision Process, 41, 180-195.

Phelps, R. H. (1977). Expert livestock judgment: A descriptive analysis or the development of expertise. Unpublished doctoral dissertation, Kansas State University.

Pulakos, E. D. (1984). A comparison of rater training programs: Error training and accuracy training. Journal of Applied Psychology, 69, 581-588.

Pulakos, E. D. (1986). The development of training programs to increase accuracy with different rating tasks. Organizational Behavior and Human Decision Process, 38, 76-91.

Pulakos, E. D., Schmitt, N., & Ostroff, C. (1986). A warning about the use of a standard deviation across dimensions to measure halo. Journal of Applied Psychology, 71, 29-32.

Pynes, J., & Bernardin, J. (1992). Mechanical vs. consensus-derived assessment center ratings: A comparison of job performance validities. Public Personnel Management, 21, 17-28.

Russell, C. J. (1985). Individual decision processes in an assessment center. Journal of Applied Psychology, 70, 737-746.

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.

Schiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. Psychological Review, 84, 127-190.

Shack, M. (1983). Assessment centers and information processing. Unpublished doctoral dissertation, Bowling Green State University.

Shanteau, J. (1987). Psychological characteristics of expert decision makers. In J. L. Mumpower, O., Renn, L. D. Phillips, & V. R. R. Uppuluri (Eds.), Expert judgment and expert systems (pp. 289-304). Berlin: Springer-Verlag.

Shanteau, J. (1988). Psychological characteristics and strategies of expert decision makers. Acta Psychologica, 68, 203-215.

Shanteau, J. (1992). The psychology of experts: An alternative view. In G. Wright & F. Bolger (Eds.), Expertise and decision support (pp. 11-23). New York: Plenum Press.

Shanteau, J., & Dino, G. A. (1993). Environmental stress effects on creativity and decision making. In O. Svenson & A. J. Maule (Eds.), Time pressure and stress in human judgment and decision making (pp. 203-215). New York: Plenum Press.

Shanteau, J., & Gaeth, G. J. (1981). Evaluation of the field method of soil texture classification: A psychological analysis of accuracy and consistency. (Tech. Rep. 79-1). Kansas State University, Department of Psychology.

Shanteau, J., Grier, M., Johnson, J., & Berner, E. (1991). Teaching decision making skills to student nurses. In J. Baron & R. B. Brown (Eds.), Teaching decision making to adolescents (pp. 185-206). Hillsdale, NJ: Erlbaum.

Shanteau, J., & Phelps, R.H. (1978). Livestock judges: How much information can an expert use? Organizational Behavior and Human Decision Process, 21, 209-219.

Singer, M. S., & Bruhns, C. (1991). Relative effect of applicant work experience and academic qualification on selection interview decisions: A study of between-sample generalizability. Journal of Applied Psychology, 76, 550-559.

Singer, M. S., & Sewell, C. (1989). Applicant age and selection interview decisions: Effect of information exposure on age discrimination in personnel selection. Personnel Psychology, 42, 135-154.

Slade, L. A., & Gordon, M. E. (1988). On the virtues of laboratory babies and student bath water: A reply to Dobbins, Lane, and Steiner. Journal of Organizational Behavior, 9, 373-376.

Slovic, P., & Lichtenstein, S. (1971). Comparison of Bayesian and regression approaches to the study of information processing in judgment. Organizational Behavior and Human Performance, 6, 649-744.

Spychalski, A. C., Quinones, M. A., Gaugler, B. B., & Pohley, K. (1997). A survey of assessment center practices in organizations in the United States. Personnel Psychology, 50, 71-90.

Stael von Holstein, C. S. (1971). An experiment in probabilistic weather forecasting. Journal of Applied Meteorology, 10, 635-645.

Stamoulis, D. T., & Hauenstein, N. M. A. (1993). Rater training and rater accuracy: Training for dimensional accuracy versus training for ratee differentiation. Journal of Applied Psychology, 78, 994-1003.

Task Force on Assessment Center Guidelines. (1989). Guidelines and ethical considerations for assessment center operations. Public Personnel Management, 18, 457-470.

Thomson, H. A. (1970). Comparison of predictor and criterion judgments of managerial performance using the multitrait-multimethod approach. Journal of Applied Psychology, 42, 496-502.

Thornton, G. C., & Byham, W. C. (1982). Assessment centers and managerial performance. New York: Academic Press.

Thornton, G. C., III, & Zorich, S. (1980). Training to improve observer accuracy. Journal of Applied Psychology, 65, 351-354.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. Science, 185, 1124-1131.

Tversky, A., & Kahneman, D. (1981). The framing of decisions and the psychology of choice. Science, 211, 453-463.

Wagenaar, W. A., & Keren, G. B. (1986). Does the expert know? The reliability of predictions and confidence ratings of experts. In E. Hollnagel, G. Mancini, and D. D. Woods (Eds.), Intelligent decision support in process environments (pp. 189-221). Berlin: Springer-Verlag.

Weber, R. P. (1990). Basic content analysis. Newbury Park: Sage.

Weiner, Y., & Schneiderman, M. (1974). Use of job information as a criterion in employment decisions of interviewers. Journal of Applied Psychology, 59, 699-704.

Wernimont, P. F., & Campbell, J. P. (1968). Signs, samples, and criteria. Journal of Applied Psychology, 52, 372-376.

Wexley, K., Sanders, R., & Yukl, G. (1973). Training interviewers to eliminate contrast effects in employment interviews. Journal of Applied Psychology, 57, 233-236.

Zedeck, S., & Cascio, W. F. (1982). Performance appraisal decisions as a function of rater training and purpose of the appraisal. Journal of Applied Psychology, 76, 752-758.

APPENDICES

A. EXAMPLE STIMULUS MATERIALS

INFORMATION ABOUT THE COMPANY

Southern Manufacturing Corporation (SMC) is a large manufacturing company based in Murfreesboro, Tennessee. SMC produces various types of components for other manufacturing companies across the country and internationally. Types of products include cast metal, forged metal, molded plastic, injection-molded plastic, extruded plastic and composite lay-up components. SMC employs around 15,500 workers across its nine divisions.

SMC started in the 1920s making forged metal parts for farming equipment. Since that time, it has expanded its range of products and maintained state-of-the-art-manufacturing processes to remain competitive. The average age of employees is 42 years, and the average tenure of employees is 24 years. Due to the nature of the work, the majority of employees are male.

The Extruded Plastics Division is currently undergoing a restructuring. The restructuring was driven by new market competition due to deregulation of several industries and an increasing number of foreign competitors. This reorganization and refocusing includes establishment of a new senior-level structure to direct 12 functional levels of work. In addition, a reduction of force is associated with the restructuring; it is anticipated that at completion of the reorganization, the Extruded Plastics division workforce will be reduced to 1100 employees. Currently, the Extruded Plastics Division employs approximately 1500 employees across the 12 functional departments. All Extruded Plastics employees have had to re-apply for jobs within the new organizational structure.

BEHAVIORAL EXERCISES

Five behavioral exercises and a background interview were used to assess each applicant's managerial skill and ability. The three written and two interpersonal exercises placed each individual in a rich variety of roles that were specifically developed to evoke certain managerial behaviors.

Strategic In-Basket

This exercise is a partial simulation of administrative tasks associated with upper-level managerial jobs. It requires quick analysis and action on a number of complex, high priority items. The task requires that that the applicant shows initiative and begins implementing a strategic plan. Assessing priorities, making decisions on limited information, recognizing conflicting information, and managing a very busy schedule are key elements of this exercise.

Simulation Role Plays

This exercise simulates two types of interpersonal situations that might be expected in a managerial position. Two role players, each with a defined script, serve as subordinates of the applicant. Successful resolution of each task depends on being able to correctly assess the nature of the problem from both the materials provided and information provided by the subordinate, addressing the problem in an appropriate manner, and effectively counseling the subordinate as necessary. Good interpersonal skills are required to perform well in this exercise.

Simulation Letter

The written component of the Simulation provides an opportunity to evaluate the applicant's ability to synthesize information (both written and oral) presented throughout the exercise. Successful completion of this portion of the exercise requires that the applicant correctly analyze the performance characteristics of various subordinates relative to the demands of the company, and to provide appropriate training recommendations to increase the subordinates' likelihood of success as company employees.

Analysis Exercise

This exercise required the applicant to read a summary of a consultant's visit to a production facility, suggest the nature of the problems facing the company, and propose recommendations as to how to improve the situation. The purpose of this exercise was not to evaluate the applicant's report-writing skills. A well written, polished report would have required much more time and attention. The first draft, however, allows an assessment of the applicant's ability to recognize problems and develop possible solutions.

Small Group Problem-Solving Task

This task requires the applicant to participate on a school board committee allocating a substantial sum of money to various needs of a school system. This exercise contains two major parts. First, the applicant is required to individually assess the situation and decide how the funds should be allocated among a number of proposals. Next, the "school board committee," which consists of three role players and the applicant, must arrive at a consensus decision on how funds should be distributed. This problem was designed to be relatively independent of the particular expertise of any applicant. Successful completion of this exercise requires that the applicant be able to communicate in a group setting, orally support his or her ideas, and be able to exert leadership and influence others.

ASSESSMENT CENTER DIMENSIONS

NOTE: Each of these behavioral dimensions will be either listed at the top of each assessment summary report as a Strength, Competency or Weakness, or discussed within the summary paragraph.

Oral Communication

Oral communication refers to effective expression of thoughts in individual or group situations. It includes delivery (e.g., volume, inflection, eye contact), clarity of ideas, and speaking with enthusiasm and confidence.

Written Communication

Written communication refers to the ability to clearly express ideas in writing, use good grammatical form, and use correct mechanics (e.g., proper spelling, punctuation, and sentence structure).

Analysis

Analysis refers to the ability to identify problems, secure relevant information, relate data from different sources, and identify causes of problems.

Judgment

Judgment refers to the ability to develop alternative courses of action and make decisions based on logical assumptions that reflect factual information. Judgment also includes providing rationale for decision and recommendations.

Planning & Organizing

This dimension refers to the ability to establish a course of action to accomplish a specific goal. It includes factors such as setting priorities and making appropriate allocation of time and resources.

Decisiveness

Decisiveness is the willingness to make decisions, render judgments, take action, or commit one's self. It also includes firmly stating one's opinion on an issue.

Delegation

Delegation refers to utilizing subordinates effectively. It implies direction, accountability, and control. Good delegation is clear and is directed to the best-suited individual; it includes deadlines and sets limits on authority.

Initiative

Initiative refers to the extent to which an individual is a self-starter and actively attempts to influence events to achieve goals. High initiative individuals take actions beyond those called for to achieve goals.

Confrontation

This dimension is defined as the ability and willingness to disagree or express opposing viewpoints in a tactful style. It also includes the willingness to assert and defend one's position even when challenged.

Sensitivity

Sensitivity is demonstrated with actions that indicate a consideration for the feelings and needs of others. A highly sensitive individual is not brash, rude, or threatening, asks for the opinions of others, and gives encouragement.

Leadership

Leadership refers to utilizing appropriate interpersonal styles and methods in guiding individuals or groups toward task accomplishment.

Team Building

This dimension is defined as the ability to work effectively with a group, and the willingness to act as part of a team and accept team-based decisions.

Coaching

Coaching is the ability to provide instruction, encouragement, and guidance to peers and subordinates in order to improve their competence.

Stress Tolerance

Stress tolerance refers to the ability to perform well under pressure and/or opposition.

JOB DESCRIPTION #1

Job Title: Manager, Organizational Development

Reports to: Vice President of Human Resources for Extruded Plastics Division

Position Purpose:

Responsible for communication of the Extruded Plastics Division programs, services, and capabilities for the purpose of strengthening and advancing the organization's charter, mission, and vision. Promotes organizational development through communication of the organization's vision and promoting a clear organizational identity. Supports other program functions to ensure effective communication of new programs, policies, and initiatives. Ensures timely and accurate communication of Extruded Plastics Division activities to customers, other divisions, the private sector, and the general public, promoting understanding and teamwork between these entities. Serves as the Extruded Plastics Division's liaison to corporate's communications organization.

Principal Duties and Responsibilities:

1. Ensures that Extruded Plastics Division activities are communicated to employees in an accurate and timely manner. Serves as liaison for communication between Extruded Plastics Division management and employees.

2. Ensures that Extruded Plastics Division activities are communicated to other divisions to promote interests of the Extruded Plastics Division and enhance the relationships between the Extruded Plastics Division and the rest of the company.

3. Responsible for the Extruded Plastics Division's contact with non-company entities including general public, private sector, and all levels of government. This includes customer sampling. Ensures that relations with other organizations serve to strengthen the image of the Extruded Plastics Division in the region and promotes social and economic well being of the region.

4. Responsible for ensuring that incoming and outgoing information is communicated through appropriate channels. Works to ensure that outgoing information is communicate through effective and appropriate use of written material, oral presentations, verbal interactions, and non-verbal cues so that these media reflect the Resources Group's value of quality and professionalism.

5. Promotes organization development by communicating a clear identity and mission that will stabilize and direct the efforts of Extruded Plastics Division employees.

6. Responsible for maintaining relevant technical expertise and continually establishing an up-to-date awareness of evolving human resources technology through benchmarking and participation in training or professional conferences.

7. Uses a proactive approach to work, which includes continuous improvement of products and services to ensure a progressive, nationally competitive human resource program that meets or exceeds customer expectations.

8. Establishes strong coalitions and relationships with other human resources areas within the Extruded Plastics Division, within the company, and with other organizations. Works in concert with other functional areas to enhance effectiveness of programs and ensure comprehensive implementation of new initiatives (e.g., proper communication, management of change, program evaluation, program development, etc.).

9. Provides leadership that creates an environment where innovation, cooperation, diversity, empowerment, and creativity can flourish and where employees are encouraged to develop to their full potential.

10. Ensures a diverse work environment free of discrimination and harassment.

11. Performs related duties as assigned or as the situation dictates.

Education and Experience Requirements:

Advanced degree in human resources, industrial/organizational psychology, business administration, or related field and/or equivalent experience. Demonstrated thorough knowledge of relevant Federal and company human resource policies and principles as well as equal employment opportunity laws and guidelines. Proven progressive management experience. Must be able to process information and develop effective communication media to promote Extruded Plastics Division activities.

Demonstrated managerial competencies such as leadership (individual and group), coaching/teambuilding, analysis, judgment, initiative, negotiation, flexibility, oral and written communication skills, creativity/innovation, and the ability to effectively relate to diverse groups of people across various management levels. The incumbent should also possess a strong desire to learn and grow professionally.

JOB DESCRIPTION #2

Job Title: Manager, Regulatory Guidance, Policy Development, and Compliance

Reports to: Vice President of Business Operations for Extruded Plastics Division

Position Purpose:

This position manages and coordinates a company-wide regulatory guidance, policy development, and compliance effort to ensure that appropriate environmental laws and regulations are drafted, implemented and adhered to. Serves as a liaison to federal and state regulatory agencies.

Principal Duties and Responsibilities:

1. Manages and coordinates a division-wide regulatory guidance, policy development, and compliance effort. This includes leading a well qualified, multidisciplinary group of specialists who properly interpret the regulations and disseminate to all appropriate units, ensure that laws and regulations are appropriately implemented throughout the division, and ensure compliance with division policies.

2. Serves as the division's principal spokesperson on federal and state environmental regulations. Serves as the division's primary liaison to federal and state regulatory agencies.

3. Communicates and coordinates with key environmental managers in the division to quickly identify and solve problems. If they are not solved, brings them to the attention of appropriate senior managers.

4. Responsible for maintaining relevant technical expertise and continually establishing an up-to-date awareness of evolving environmental legislation and technologies through participation in training, professional conferences, and thorough benchmarking with leading public and private organizations.

5. Evaluates the effectiveness of and modifies current programs and activities as necessary to accomplish division objectives.

6. Ensures that activities are properly coordinated, internally and externally, and that they are sufficiently integrated with related division, state, and federal programs to efficiently and cost-effectively accomplish organizational objects.

7. Conducts cost-benefit analyses on proposed and on going projects and programs.

8. Serves as the liaison to the Finance and Contracts Departments in budget and contract preparation, review, and analysis.

9. Provides leadership that creates an environment where innovation, cooperation, diversity, empowerment, and creativity can flourish and where employees are encouraged to develop to their full potential.

10. Ensures a diverse work environment free of discrimination and harassment.

11. Ensures a safe and healthy work environment where work practices are safe and free from hazardous situations.

12. Performs related duties as assigned or as the situation dictates.

Education and Experience Requirements:

An advance degree, M.S. highly desired, in environmental science, environmental planning, related field, or equivalent experience. Experience in environmental auditing is also desirable. Demonstrated successful record managing environmental programs. Must possess knowledge of specific and broad environmental laws, regulations, and policies, and the regulatory process. Knowledge of pollution prevention technologies and waste management is highly desirable.

Demonstrated managerial competencies such as leadership (individual and group), analysis, judgment, initiative, negotiation, coaching/teambuilding, oral and written communication skills, confrontation, and creativity, innovation. Must possess a high degree of interpersonal sensitivity and flexibility in order to interact with and serve a diverse group of internal and external customers. Must also demonstrate diplomacy/tact when dealing with critical/sensitive issues.

JOB DESCRIPTION #3

Job Title: Manager, Systems Engineering and Special Analysis

Reports to: Vice President of Engineering Services for Extruded Plastics Division

Position Purpose:

Provide technical, administrative, and functional direction to develop mechanical, electrical, and integrated systems engineering and analyses products and services to the Extruded Plastics Division and their customers.

Principal Duties and Responsibilities:

1. Ensures that policies and procedures are developed and administered in a manner that supports and enhances organizational objectives, the division's strategic direction and meets customer needs.

2. Accountable for planning and organizing all Systems and Special Analysis-related activities in the division. Develops budgets and controls costs of these activities and ensures that all programs are administered consistently.

3. Develops, implements, and evaluates the effectiveness of and modifies products and activities in the Extruded Plastics Division as necessary to accomplish organizational objectives.

4. Ensures effective and efficient administration and continuous improvement of activities as well as planning for the development of new products to meet the long-range needs and strategic directions of the company.

5. Responsible for implementing the Total Quality initiative throughout the Extruded Plastics Division including identifying and resolving problems, reviewing and improving processes, establishing and supporting teams, identifying customer needs and requirements, and taking appropriate steps to ensure that quality is a prime consideration in all activities.

6. Directs the design and development of electrical, mechanical, and/or integrated systems to support the success of the Extruded Plastics Division and meet the needs of its customers.

7. Directs the development and implementation of state-of-the-are analyses to support the needs of the Extruded Plastics Division and its customers.

8. Provides leadership that creates an environment where innovation, cooperation, diversity, empowerment, and creativity can flourish and where employees are encouraged to develop to their full potential.

9. Integrates activities across engineering services functions to remove barriers to efficiency improvements.

10. Ensures a diverse work environment free of discrimination and harassment.

11. Ensures a safe and healthy work environment where work practices are safe and free from hazardous situations.

12. Performs related duties as assigned or as the situation dictates.

Education and Experience Requirements:

An advanced degree in mechanical, electrical, or systems engineering or a related field, or equivalent experience. Five to seven years' experience in the development and implementation of engineering designs associated with plastics manufacturing also desirable.

Demonstrated managerial competencies such as creativity, sensitivity, coaching/teambuilding, leadership (group and individual), etc.

All applicants assessed at this stage of the restructuring process met or exceeded the minimum educational and technical qualifications for this position.

Applicant #1

- Strengths: Oral Communication, Initiative, Confrontation, Sensitivity, Leadership
- Competencies: Analysis, Judgment, Planning & Organizing, Decisiveness, Coaching & Team Building, Stress Tolerance
- Weaknesses: Written Communication

The applicant performed well in the assessment center. Except for the written communication dimension, her performance was consistently evaluated as above average. Her performance on many of the assessment center exercises indicates that she should be an effective HR problem solver. Her writing was poorly organized and contained a few minor grammatical errors, but was otherwise acceptable.

Two items of concern were raised during the assessment center. First, the applicant overfocused on the human resource issues in the Case Analysis and consequently ignored several more important issues. Her suggestions were incomplete and she did not fully solve the problem situation.

Second, the applicant had real difficulty delegating responsibility for issues on which she held technical competence. That is, the applicant did not delegate HR issues to others, even though it was clearly inappropriate for her to assume responsibility for the issues. On items for which she did not possess expertise, she was clearly willing to delegate. In these cases, she gave clear instructions regarding how the issues were to be handled, set appropriate time limitations, and specified the degree of feedback she required. If she is to be a successful HR manager, the applicant must apply these excellent delegation skills within her area of expertise as well.

Applicant #2

- Strengths: None Observed
- Competencies: Oral Communication, Written Communication, Analysis, Planning & Organizing, Decisiveness, Delegation, Initiative, Confrontation, Stress Tolerance
- Weaknesses: None Observed

The applicant performed well on most of the assessment center dimensions. While few exceptional strengths were noted, and no glaring deficiencies were observed, she performed well above average on many dimensions. She was particularly effective at analyzing complex situations.

One major concern emerged with the applicant during the assessment center. Specifically, her leadership, sensitivity, and judgment varied considerably across the exercises. In the Simulation and Small Group Discussion she demonstrated acceptable to very good leadership, sensitivity, and judgment. However, in the Small Group Discussion, she was somewhat condescending to certain role players. She also displayed weaker leadership and judgment by accepting several bad ideas offered by the "power" member of the committee. The applicant may place too much emphasis on early impressions of individuals or the perceived power role of an individual.

Applicant #3

- Strengths: None Observed
- Competencies: Oral Communication, Written Communication, Analysis, Judgment, Planning & Organizing, Delegation, Initiative, Confrontation, Sensitivity, Leadership, Stress Tolerance
- Weaknesses: Decisiveness, Coaching & Team Building

The applicant performed well on most of the assessment center dimensions; few strengths were identified, and no major weaknesses were noted. Hence, the applicant can be described as a very consistent performer.

Two problems surfaced with the applicant during the assessment center exercises. First, she was indecisive and did not provide coaching to subordinates. She appeared uncomfortable making decisions without complete data. This low level of decisiveness may cause difficulty when decisions have to be made based upon partial information.

While the applicant exhibited satisfactory analysis and judgment, she had a tendency to apply general human resource models to problems that were not produced by human resource issues. This tendency resulted in poor analysis and judgment in some of the exercises. It should be emphasized, however, that the applicant exhibited solid analysis and judgment when she did not rely upon her preconceived notions concerning problems.

B. SURVEY MEASURES

A. Based upon the above report, how would you characterize this applicant's overall performance in the assessment center? Circle **one** response on the scale below.

1	2	3	4	5	6
Weak, Poor	Moderate, Marginal	Competent, Satisfactory	Good, Strong	Very Strong, Outstanding	Exceptional, Exemplary

B. How would you characterize this applicant's level of performance for each of the following skills? Circle **one** response using the scale below.

1	2	3	4	5	6
Weak, Poor	Moderate, Marginal	Competent, Satisfactory	Good, Strong	Very Strong, Outstanding	Exceptional, Exemplary

1. Problem Solving Skills.....	1	2	3	4	5	6
2. Organizational Leadership Skills....	1	2	3	4	5	6
3. Communication Skills.....	1	2	3	4	5	6
4. Managerial Skills.....	1	2	3	4	5	6

C. The salary range for this job is $50,910 - $61,650 per year. If you were to hire this applicant, what starting salary would you recommend?

$_____ per year

D. If you were to hire this applicant, how well do you think the applicant would perform on the job? Circle **one** response on the scale below:

1	2	3	4	5	6
Weak, Poor	Moderate, Marginal	Competent, Satisfactory	Good, Strong	Very Strong, Outstanding	Exceptional, Exemplary

E. Without regard to any information you have read about other applicants for this job, is this person qualified for the job for which she applied?

❑ Yes, this applicant is qualified. ❑ No, this applicant is not qualified.

F. In the space below, please provide your reason(s) for why you would or would not hire this applicant:

You will now be asked to rank order the applicants for this position.

If you could hire **only one** of the applicants, which applicant would it be? In other words, which applicant is your first choice to fill the position? Write the applicant number in the blank below.

1st choice: _____

If your first choice declines the job, which applicant is your second choice to fill the job?

2nd choice: _____

And third choice?

3rd choice: _____

C. PILOT STUDY QUESTIONNAIRE

Please respond to the following statements about the hiring decision exercise you just completed. Use the following scale:

1 = Strongly Disagree
2 = Disagree
3 = Neither Disagree nor Agree
4 = Agree
5 = Strongly Agree

1.	The material was easy to read.....	1	2	3	4	5
2.	The information about each applicant was realistic	1	2	3	4	5
3.	The directions for completing the exercise were clear.....	1	2	3	4	5
4.	I understood the task at hand	1	2	3	4	5
5.	The exercise took too long to complete.....	1	2	3	4	5
6.	There was too much information for me to deal with....	1	2	3	4	5
7.	The exercise was an appropriate length	1	2	3	4	5
8.	I was able to complete the exercise in 60-90 minutes ...	1	2	3	4	5
9.	The directions for ranking applicants made sense to me	1	2	3	4	5
10.	The questionnaire about each applicant was easy to follow.....	1	2	3	4	5

Please record how long it took you to complete the exercise. If you did not finish within 90 minutes, mark the last page you read, and put 90+ in the blank: _____minutes

Which part of the directions was not clear?

What other sort of information did you need to help you make your decisions?

What would you change about the exercise or the materials?

D. MATHEMATICAL DEFINITIONS OF CRONBACH'S EVALUATIVE ACCURACY COMPONENTS

Cronbach (1955) isolated four orthogonal variance components indicative of inaccuracy in ratings. Use and derivation of these components requires that ratings be obtained on several individuals (groups, companies) and on two or more dimensions. Further, it is assumed that all ratings share the same scale. Finally, determination of accuracy requires that true performance scores (true scores) be known on all dimensions for all individuals.

Given the above, we define:

r_{ij}	rating of the i^{th} individual on the j^{th} dimension
$\bar{r}_{i\bullet}$	mean rating of the i^{th} individual across all dimensions
$\bar{r}_{\bullet j}$	mean rating on the j^{th} dimension across all dimensions
$\bar{r}_{\bullet\bullet}$	grand mean rating across all dimensions and individuals
t_{ij}, $\bar{t}_{i\bullet}$, $\bar{t}_{\bullet j}$, $\bar{t}_{\bullet\bullet}$	as the true scores associated with the above ratings
n	number of individuals (ratees)
k	number of dimensions

Then, inaccuracy due to *Elevation* is defined as:

$$E^2 = \left(\bar{r}_{\bullet\bullet} - \bar{t}_{\bullet\bullet}\right)^2,$$

inaccuracy due to *Differential Elevation* as:

$$DE^2 = \frac{1}{n}\sum_{i=1}^{n}\left[\left(\bar{r}_{i\bullet} - \bar{r}_{\bullet\bullet}\right) - \left(\bar{t}_{i\bullet} - \bar{t}_{\bullet\bullet}\right)\right]^2,$$

inaccuracy due to *Stereotyping (Stereotype Accuracy)* as:

$$SA^2 = \frac{1}{k}\sum_{j=1}^{k}\left[\left(\bar{r}_{\bullet j} - \bar{r}_{\bullet\bullet}\right) - \left(\bar{t}_{\bullet j} - \bar{t}_{\bullet\bullet}\right)\right]^2,$$

and inaccuracy due to *Differential Accuracy* as:

$$DA^2 = \frac{1}{kn}\sum_{i=1}^{n}\sum_{j=1}^{k}\left[\left(\bar{r}_{\bullet\bullet} + r_{ij} - \bar{r}_{i\bullet} - \bar{r}_{\bullet j}\right) - \left(\bar{t}_{\bullet\bullet} + t_{ij} - \bar{t}_{i\bullet} - \bar{t}_{\bullet j}\right)\right]^2$$

E. DOMINANCE ANALYSIS TABLES BY SUBJECT GROUP

Dominance Analysis for Undergraduate Student Sample

Variable(s)	R^2	Contribution of Prob	Lead	Comm	Mgr
Overall					
----	0	.4150	.5193	.5281	.5751
Prob	.4150	----	.1905	.1950	.2310
Lead	.5193	.0862	----	.1279	.1474
Comm	.5281	.0819	.1190	----	.1376
Mgr	.5751	.0709	.0916	.0907	----
Prob, Lead	.6055	----	----	.0761	.0944
Prob, Comm	.6100	----	.0715	----	.0890
Prob, Mgr	.6460	----	.0539	.0530	----
Lead, Comm	.6472	.0344	----	----	.0644
Lead, Mgr	.6667	.0331	----	.0449	----
Com, Mgr	.6657	.0333	.0458	----	----
Prob, Lead, Comm	.6815	----	----	----	.0489
Prob, Lead, Mgr	.6998	----	----	.0306	----
Prob, Comm, Mgr	.6990	----	.0314	----	----
Lead, Comm, Mgr	.7116	.0189	----	----	----
Mean Squared Semi-Partial		.0967	.1404	.1433	.1735
Importance (Cx_i)		.1368	.1854	.1886	.2196
Relative Importance		19%	25%	26%	30%
HR Job					
----	0	.2729	.4423	.3682	.4233
Prob	.2729	----	.2600	.1886	.2158
Lead	.4423	.0906	----	.0829	.1087
Comm	.3682	.0932	01569	----	.1432
Mgr	.4233	.0654	.1277	.0881	----
Prob, Lead	.5329	----	----	.0476	.0607
Prob, Comm	.4615	----	.1190	----	.0900
Prob, Mgr	.4887	----	.1049	.0628	----
Lead, Comm	.5252	.0553	----	----	.0601
Lead, Mgr	.5510	.0426	----	.0342	----

Note: n = 59; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Dominance Analysis for Undergraduate Student Sample (continued)

		Contribution of			
Variable(s)	R^2	Prob	Lead	Comm	Mgr
Com, Mgr	.5514	.0400	.0739	----	----
Prob, Lead, Comm	.5805	----	----	----	.0367
Prob, Lead, Mgr	.5936	----	----	.0235	----
Prob, Comm, Mgr	.5514	----	.0657	----	----
Lead, Comm, Mgr	.5853	.0319	----	----	----
Mean Squared Semi-Partial		.0865	.1688	.1120	.1423
Importance (Cx_i)		.1085	.1972	.1400	.1715
Relative Importance		18%	32%	23%	27%
Administrative Job					
----	0	.3114	.3934	.4235	.4496
Prob	.3114	----	.1766	.1686	.2211
Lead	.3934	.0946	----	.1448	.1591
Comm	.4235	.0565	.1148	----	.1506
Mgr	.4496	.0829	.1029	.1244	----
Prob, Lead	.4481	----	----	.0770	.1067
Prob, Comm	.4800	----	.0850	----	.1179
Prob, Mgr	.5325	----	.0622	.0654	----
Lead, Comm	.5383	.0268	----	----	.0820
Lead, Mgr	.5526	.0422	----	.0677	----
Com, Mgr	.5740	.0238	.0463	----	----
Prob, Lead, Comm	.5650	----	----	----	.0697
Prob, Lead, Mgr	.5948	----	----	.0400	----
Prob, Comm, Mgr	.5979	----	.0369	----	----
Lead, Comm, Mgr	.6203	.0144	----	----	----
Mean Squared Semi-Partial		.0816	.1273	.1389	.1696
Importance (Cx_i)		.1087	.1566	.1698	.1996
Relative Importance		17%	25%	27%	31%

Note: n = 59; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Dominance Analysis for Undergraduate Student Sample (continued)

Variable(s)	R^2	Contribution of Prob	Lead	Comm	Mgr
Engineering Job					
----	0	.5383	.6053	.6623	.6779
Prob	.5283	----	.1533	.2136	.2194
Lead	.6053	.0762	----	.1453	.1449
Comm	.6623	.0796	.0883	----	.0889
Mgr	.6779	.0698	.0723	.0733	----
Prob, Lead	.6816	----	----	.0986	.0978
Prob, Comm	.7419	----	.0383	----	.0498
Prob, Mgr	.7477	----	.0316	.0439	----
Lead, Comm	.7506	.0296	----	----	.0392
Lead, Mgr	.7502	.0292	----	.0397	----
Com, Mgr	.7512	.0405	.0387	----	----
Prob, Lead, Comm	.7802	----	----	----	.0396
Prob, Lead, Mgr	.7794	----	----	.0304	----
Prob, Comm, Mgr	.7917	----	.0818	----	----
Lead, Comm, Mgr	.7898	.0200	----	----	----
Mean Squared Semi-Partial		.1092	.1307	.1634	.1684
Importance (Cx_i)		.1641	.1911	.2244	.2302
Relative Importance		20%	24%	28%	28%

Note: n = 59; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Dominance Analysis for Manager Sample

Variable(s)	R^2	Contribution of Prob	Lead	Comm	Mgr
Overall					
----	0	.3217	.6214	.4661	.6082
Prob	.3217	----	.3440	.2400	.3397
Lead	.6214	.0443	----	.0907	.0966
Comm	.4661	.0955	.2460	----	.2386
Mgr	.6082	.0532	.1099	.0965	----
Prob, Lead	.6657	----	----	.0690	.0758
Prob, Comm	.5617	----	.1731	----	.1701
Prob, Mgr	.6614	----	.0801	.0704	----
Lead, Comm	.7122	.0226	----	----	.0583
Lead, Mgr	.7181	.0234	----	.0524	----
Com, Mgr	.7047	.0271	.0658	----	----
Prob, Lead, Comm	.7348	----	----	----	.0491
Prob, Lead, Mgr	.7415	----	----	.0424	----
Prob, Comm, Mgr	.7318	----	.0521	----	----
Lead, Comm, Mgr	.7705	.0134	----	----	----
Mean Squared Semi-Partial		.0752	.2116	.1409	.2046
Importance (Cx_i)		.1060	.2533	.1787	.2459
Relative Importance		14%	32%	23%	31%
HR Job					
----	0	.1304	.4053	.1370	.3658
Prob	.1304	----	.3163	.1047	.2909
Lead	.4053	.0414	----	.0710	.0951
Comm	.1370	.0980	.3393	----	.2914
Mgr	.3658	.0554	.1345	.0626	----
Prob, Lead	.4467	----	----	.0613	.0843
Prob, Comm	.2350	----	.2729	----	.2384
Prob, Mgr	.4212	----	.1098	.0522	----
Lead, Comm	.4762	.0317	----	----	.0750
Lead, Mgr	.5003	.0307	----	.0509	----

Note: n = 42; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Dominance Analysis for Manager Sample (continued)

Variable(s)	R^2	Contribution of Prob	Lead	Comm	Mgr
Com, Mgr	.4284	.0450	.0739	.0509	----
Prob, Lead, Comm	.5080	----	----	----	.0677
Prob, Lead, Mgr	.5310	----	----	.0446	----
Prob, Comm, Mgr	.4734	----	.1022	----	----
Lead, Comm, Mgr	.5512	.0244	----	----	----
Mean Squared Semi-Partial		.0571	.2254	.0730	.1886
Importance (Cx_i)		.0639	.2348	.0790	.1980
Relative Importance		12%	41%	14%	34%
Administrative Job					
----	0	.2593	.4635	.3729	.3901
Prob	.2593	----	.2792	.2339	.2367
Lead	.4635	.0749	----	.1433	.0810
Comm	.3729	.1202	.2339	----	.1833
Mgr	.3901	.1059	.1544	.1662	----
Prob, Lead	.5385	----	----	.1133	.0619
Prob, Comm	.4931	----	.1586	----	.1261
Prob, Mgr	.4960	----	.1044	.1232	----
Lead, Comm	.6068	.0449	----	----	.0454
Lead, Mgr	.5445	.0558	----	.1077	----
Com, Mgr	.5562	.0630	.0960	----	----
Prob, Lead, Comm	.6517	----	----	----	.0365
Prob, Lead, Mgr	.6003	----	----	.0879	----
Prob, Comm, Mgr	.6192	----	.0690	----	----
Lead, Comm, Mgr	.6522	.0360	----	----	----
Mean Squared Semi-Partial		.0950	.1949	.1685	.1451
Importance (Cx_i)		.1125	.2187	.1892	.1679
Relative Importance		16%	32%	27%	24%

Note: n = 42; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Dominance Analysis for Manager Sample (continued)

Variable(s)	R^2	Contribution of Prob	Lead	Comm	Mgr
Engineering Job					
----	0	.4170	.7225	.6346	.7267
Prob	.4170	----	.3287	.2754	.3319
Lead	.7225	.0233	----	.0647	.0930
Comm	.6346	.0578	.1525	----	.1648
Mgr	.7267	.0223	.0888	.0728	----
Prob, Lead	.7458	----	----	.0522	.0746
Prob, Comm	.6924	----	.1055	----	.1157
Prob, Mgr	.7490	----	.0713	.0592	----
Lead, Comm	.7872	.0108	----	----	.0546
Lead, Mgr	.8155	.0048	----	.0263	----
Com, Mgr	.7994	.0087	.0424	----	----
Prob, Lead, Comm	.7980	----	----	----	.0464
Prob, Lead, Mgr	.8203	----	----	.0241	----
Prob, Comm, Mgr	.8081	----	.0362	----	----
Lead, Comm, Mgr	.8418	.0026	----	----	----
Mean Squared Semi-Partial		.0684	.1935	.1512	.2010
Importance (Cx_i)		.1155	.2555	.2105	.2328
Relative Importance		14%	31%	26%	32%

Note: n = 42; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Dominance Analysis for Trained Assessor Sample

Variable(s)	R^2	Contribution of Prob	Lead	Comm	Mgr
Overall					
----	0	.3627	.5139	.2832	.5175
Prob	.3627	----	.2236	.1262	.2236
Lead	.5139	.0724	----	.0321	.1211
Comm	.2832	.2057	.2328	----	.2800
Mgr	.5175	.0688	.1175	.0457	----
Prob, Lead	.5863	----	----	.0259	.0780
Prob, Comm	.4889	----	.1233	----	.1331
Prob, Mgr	.5864	----	.0780	.0357	----
Lead, Comm	.5460	.0662	----	----	.0998
Lead, Mgr	.6350	.0294	----	.0109	----
Com, Mgr	.5632	.0588	.0826	----	----
Prob, Lead, Comm	.6122	----	----	----	.0629
Prob, Lead, Mgr	.6644	----	----	.0108	----
Prob, Comm, Mgr	.6220	----	.0531	----	----
Lead, Comm, Mgr	.6459	.0293	----	----	----
Mean Squared Semi-Partial		.1117	.1819	.0713	.1895
Importance (Cx_i)		.1398	.2158	.0965	.2231
Relative Importance		21%	32%	14%	33%
HR Job					
----	0	.2323	.4324	.2011	.4070
Prob	.2323	----	.2725	.1173	.2380
Lead	.4324	.0724	----	.0374	.0760
Comm	.2011	.1486	.2688	----	.2343
Mgr	.4070	.0634	.1015	.0284	----
Prob, Lead	.5048	----	----	.0270	.0481
Prob, Comm	.3497	----	.1822	----	.1436
Prob, Mgr	.4703	----	.0826	.0230	----
Lead, Comm	.4698	.0620	----	----	.0527

Note: n = 34; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Dominance Analysis for Trained Assessor Sample (continued)

		Contribution of			
Variable(s)	R^2	Prob	Lead	Comm	Mgr
Lead, Mgr	.5085	.0445	----	.0141	----
Com, Mgr	.4354	.0579	.0872	----	----
Prob, Lead, Comm	.5318	----	----	----	.0329
Prob, Lead, Mgr	.5529	----	----	.0118	----
Prob, Comm, Mgr	.4933	----	.0715	----	----
Lead, Comm, Mgr	.5226	.0422	----	----	----
Mean Squared Semi-Partial		.0904	.1873	.0575	.1541
Importance (Cx_i)		.1060	.2089	.0738	.1760
Relative Importance		19%	37%	13%	31%
Administrative Job					
----	0	.2837	.3701	.2198	.3764
Prob	.2837	----	.1589	.0997	.1844
Lead	.3701	.0726	----	.0628	.1844
Comm	.2198	.1367	.2130	----	.2300
Mgr	.3764	.0917	.1785	.0734	----
Prob, Lead	.4426	----	----	.0444	.1331
Prob, Comm	.3835	----	.1036	----	.1306
Prob, Mgr	.4681	----	.1076	.0460	----
Lead, Comm	.4329	.0542	----	----	.1439
Lead, Mgr	.5549	.0208	----	.0219	----
Com, Mgr	.4498	.0643	.1270	----	----
Prob, Lead, Comm	.4871	----	----	----	.1071
Prob, Lead, Mgr	.5757	----	----	.0185	----
Prob, Comm, Mgr	.5141	----	.0801	----	----
Lead, Comm, Mgr	.5768	.0174	----	----	----
Mean Squared Semi-Partial		.0961	.1673	.0733	.1863
Importance (Cx_i)		.1142	.1866	.0886	.2048
Relative Importance		19%	31%	15%	34%

Note: n = 34; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

Dominance Analysis for Trained Assessor Sample (continued)

Variable(s)	R^2	Contribution of Prob	Lead	Comm	Mgr
Engineering Job					
----	0	.4123	.5974	.4173	.6006
Prob	.5283	----	.2395	.1681	.2305
Lead	.6053	.0543	----	.0172	.0941
Comm	.6623	.1631	.1973	----	.2245
Mgr	.6779	.0422	.0909	.0412	----
Prob, Lead	.6816	----	----	.0163	.0580
Prob, Comm	.7419	----	.0877	----	.0987
Prob, Mgr	.7477	----	.0670	.0633	----
Lead, Comm	.7506	.0534	----	----	.0812
Lead, Mgr	.7502	.0183	----	.0042	----
Com, Mgr	.7512	.0372	.0539	----	----
Prob, Lead, Comm	.7802	----	----	----	.0473
Prob, Lead, Mgr	.7794	----	----	.0055	----
Prob, Comm, Mgr	.7917	----	.0362	----	----
Lead, Comm, Mgr	.7898	.0195	----	----	----
Mean Squared Semi-Partial		.1000	.1712	.0883	.1794
Importance (Cx_i)		.1387	.2198	.1293	.2276
Relative Importance		19%	31%	18%	32%

Note: n = 34; Prob = Problem Solving Skills; Org Lead = Organizational Leadership Skills; Comm = Communication Skills; Mgr = Managerial Skills

VITA

Erin Kate Patton Atchley was born in Washington, D. C. on October 2, 1968. She attended elementary school in Stillwater, Oklahoma, and junior high and high school in Knoxville, Tennessee. She graduated from Farragut High School in Knoxville in 1986. Kate entered The University of Tennessee, Knoxville in August 1986 as a Tennessee Scholar and Bonham Scholar. She received her Bachelor of Science in College Scholars with a concentration in Health and Physiological Psychology from The University of Tennessee, Knoxville in May, 1990. Kate returned to The University of Tennessee from August, 1990 to August, 1992 to obtain a master's degree in College Student Personnel. During the 1991-1992 academic year, she received the Hilton A. Smith Fellowship. In August 1992, she entered the Industrial and Organizational Psychology Program at The University of Tennessee, Knoxville. She officially received her Doctoral degree in December, 1999.

While completing her doctorate, Kate was awarded the Hilton A. Smith Fellowship for 1992-1993 and 1993-1994. In addition, she was the recipient of a College of Business Administration Scholarly Research Grant in 1999. She has been involved in a variety of applied projects, including designing and evaluating workforce training programs, conducting managerial assessment centers, and conducting leadership development and executive coaching programs. Kate has consulted for various companies both nationally and in Bangkok, Thailand, including Shinawatra Company, Inc., Lockheed-Martin Energy Systems, BellSouth's L. M. Berry and Company, Ruby Tuesdays, Inc., and the Tennessee Valley Authority.

During her tenure at The University of Tennessee, Knoxville, Kate presented several papers at national and regional conferences and published a book chapter. She is affiliated with the Society for Industrial and Organizational Psychology, the American Psychological Association, and the Southern Management Association.

Kate is presently employed as the Director of Assessment and Development for Graduate MBA Programs at The University of Tennessee. In her free time, she enjoys gardening, reading, and spending time with her husband and cat.