INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
313/761-4700  800/521-0600
A test of a model of employment interview information gathering

Spychalski, Annette Christine, M.A.

Rice University, 1994
RICE UNIVERSITY

A TEST OF A MODEL OF EMPLOYMENT INTERVIEW INFORMATION GATHERING

by

ANNETTE SPYCHALSKI

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE MASTER OF ARTS

APPROVED, THESIS COMMITTEE

B. B. Gaugler, Assistant Professor, Chair Industrial and Organizational Psychology

R. L. Dipboye, Professor Industrial and Organizational Psychology

R. N. Taylor, Professor Jones Graduate School of Administration

Houston, Texas

May, 1994
Abstract

A Test of a Model of Employment Interview Information Gathering

by

Annette C. Spychalski

Interviewers' questioning behavior and predictive validity of applicant ratings varies considerably in unstructured interviews. A model hypothesizing a relationship between these variables is tested in this study. The model proposes that the relationship between interviewer questioning behavior and evaluation validity is mediated by the diagnosticity of applicant information that is collected during the interview.

The process and content of three interviewers' questioning of 149 candidates for an entry-level correctional officer position was examined. Although the complete information gathering model was not supported, a robust relationship between questioning behavior and information diagnosticity emerged. Furthermore, the validity of individual interviewers' applicant evaluations varied considerably.

These results reinforce the existence of differences in interviewers' questioning behavior and differences in the quality of applicant information they gather. Because differences in questioning behavior correspond to differences in the predictive validity of applicant ratings, both these variables should be monitored at the individual interviewer level.
Acknowledgments

I would like to thank Doctors Barbara Gaugler, Robert Dipboye, and Ronald Taylor for their effort and guidance. Their input was critical to the completion of this project. For assistance in data collection, I am forever grateful to Elizabeth Knapp, Laura Galarza, and Jim Day. I would also like to thank my family for its unwavering support in my studies and all other respects.
# Table of Contents

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Models of the Employment Interview</td>
<td>1</td>
</tr>
<tr>
<td>Interviewer Questioning Style in Unstructured Interviews</td>
<td>7</td>
</tr>
<tr>
<td>Diagnosticity in the Employment Interview</td>
<td>14</td>
</tr>
<tr>
<td>Level of Analysis in Employment Interview Studies</td>
<td>21</td>
</tr>
<tr>
<td>The Present Investigation</td>
<td>24</td>
</tr>
<tr>
<td>Method</td>
<td>25</td>
</tr>
<tr>
<td>Subjects</td>
<td>25</td>
</tr>
<tr>
<td>Application Process</td>
<td>27</td>
</tr>
<tr>
<td>Criteria</td>
<td>28</td>
</tr>
<tr>
<td>Materials and Coding Procedures</td>
<td>29</td>
</tr>
<tr>
<td>Rating Accuracy</td>
<td>30</td>
</tr>
<tr>
<td>Results</td>
<td>32</td>
</tr>
<tr>
<td>Relationships Among Study Variables</td>
<td>32</td>
</tr>
<tr>
<td>Aggregate Interview Validity</td>
<td>33</td>
</tr>
<tr>
<td>Individual Interviewer Validity</td>
<td>40</td>
</tr>
<tr>
<td>Individual Use of Interview Questions and Elicitation of</td>
<td>45</td>
</tr>
<tr>
<td>Diagnostic Information</td>
<td>45</td>
</tr>
<tr>
<td>Data Reduction</td>
<td>56</td>
</tr>
<tr>
<td>Hierarchical Regression Tests of the Information Gathering</td>
<td>59</td>
</tr>
<tr>
<td>Model</td>
<td>59</td>
</tr>
<tr>
<td>Model Testing With Correlational Tests</td>
<td>68</td>
</tr>
<tr>
<td>Aggregate Interviewer Relationships Between Diagnosticity</td>
<td>73</td>
</tr>
<tr>
<td>and Questioning Behavior</td>
<td>73</td>
</tr>
<tr>
<td>Discussion</td>
<td>76</td>
</tr>
</tbody>
</table>
Table of Contents, Continued

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Footnotes</td>
<td>90</td>
</tr>
<tr>
<td>References</td>
<td>92</td>
</tr>
<tr>
<td>Appendices</td>
<td></td>
</tr>
<tr>
<td>Coding Conventions and Coding Forms</td>
<td></td>
</tr>
<tr>
<td>Interview Questions and Applicant Responses</td>
<td>99</td>
</tr>
<tr>
<td>Coding Conventions and Coding Forms</td>
<td></td>
</tr>
<tr>
<td>True Score Estimates</td>
<td>110</td>
</tr>
<tr>
<td>Intercorrelations Among Study Variables</td>
<td>121</td>
</tr>
<tr>
<td>Hierarchical Regression of Performance Appraisals</td>
<td></td>
</tr>
<tr>
<td>on Interview Score</td>
<td>136</td>
</tr>
</tbody>
</table>
List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1:</td>
<td>Aggregate Unpartialled Validity Coefficients</td>
<td>35</td>
</tr>
<tr>
<td>Table 2:</td>
<td>Aggregate Semi-partial Validity Coefficients</td>
<td>38</td>
</tr>
<tr>
<td>Table 3:</td>
<td>Individual Interviewer Unpartialled Validity Coefficients</td>
<td>42</td>
</tr>
<tr>
<td>Table 4:</td>
<td>Individual Interviewer Corrected Semi-partial Validity Coefficients</td>
<td>43</td>
</tr>
<tr>
<td>Table 5:</td>
<td>Mean Proportion and Standard Deviations for Questioning Variables</td>
<td>46</td>
</tr>
<tr>
<td>Table 6:</td>
<td>Mean Proportion and Standard Deviations for Diagnosticity Variables</td>
<td>48</td>
</tr>
<tr>
<td>Table 7:</td>
<td>Test of Differences in Interviewer Questioning Behavior</td>
<td>49</td>
</tr>
<tr>
<td>Table 8:</td>
<td>Interviewer Differences in the Use of Question Type</td>
<td>50</td>
</tr>
<tr>
<td>Table 9:</td>
<td>Test of Differences in Interviewer Information Diagnosticity</td>
<td>53</td>
</tr>
<tr>
<td>Table 10:</td>
<td>Interviewer Differences in Diagnosticity of Information Elicited From Applicants</td>
<td>54</td>
</tr>
<tr>
<td>Table 11:</td>
<td>Rotated Factor Matrix of Question Variables (varimax rotation)</td>
<td>58</td>
</tr>
<tr>
<td>Table 12:</td>
<td>Rotated Factor Matrix of Interview Judgment-specific Diagnosticity Variables</td>
<td>60</td>
</tr>
<tr>
<td>Table 13:</td>
<td>Hierarchical Regression of Accuracy on Diagnosticity and Questioning Style (aggregate interviewer level)</td>
<td>62</td>
</tr>
<tr>
<td>Table 14:</td>
<td>Hierarchical Regression of Accuracy on Diagnosticity and Questioning Style (Interviewer 1)</td>
<td>64</td>
</tr>
<tr>
<td>Table 15:</td>
<td>Hierarchical Regression of Accuracy on Diagnosticity and Questioning Style (Interviewer 2)</td>
<td>65</td>
</tr>
</tbody>
</table>
List of Tables, Continued

Table 16: Hierarchical Regression of Accuracy on Diagnosticity and Questioning Style (Interviewer 3) 66
Table 17: Regression of Accuracy on Questioning Factors 69
Table 18: Regression of Accuracy on Diagnosticity Factors 70
Table 19: Set Correlations Among Model Components 72
Table 20: Intercorrelations Among Question and Diagnosticity Factors 74
List of Figures

Figure 1: Model of Employment Interview Information Gathering  3
Introduction

The traditional employment interview has rarely produced criterion-related validity coefficients greater than .20 (Dreher, Ash, & Hancock, 1988; Hunter & Hunter, 1984). Despite this poor record, the employment interview is the most widely used method of employee selection (Ulrich & Trumbo, 1965). In the face of more valid selection devices such as mental ability tests (uncorrected mean $r = .53$), peer ratings (uncorrected mean $r = .49$), biodata (uncorrected mean $r = .37$), and reference checks (uncorrected mean $r = .26$) (Hunter & Hunter, 1984), widespread devotion to the employment interview as a method of employee selection seems puzzling. Perhaps adherence to this method is due to its intuitive appeal. People generally like to believe that they are good judges of character and that they can estimate the suitability of a potential employee more accurately than can a paper-and-pencil measure. They may feel that a "human element" which cannot be measured through a written exam provides valuable information regarding the fit between the potential employee and the job. Whatever the reason for the popularity of the selection interview, it is clear that those responsible for selecting employees do not intend to abandon this technique.

Models of the Employment Interview

Since people persist in using employment interviews for selection purposes, researchers have studied them in order to better understand how they operate. Several researchers have proposed models depicting the components involved in the interview process (Arvey & Campion, 1982; Dipboye & Gaugler, 1993; Schmitt, 1976). These models illustrate the interplay among applicant, interviewer, and situational variables that can influence the outcome of the interview.
This study examines in depth one aspect of the interview process: the relationship between interviewers' methods of applicant questioning and the usefulness of the information provided by the applicant. The usefulness of information provided by the applicant, in part, determines the quality of the interviewer evaluation of applicant suitability for the job. Thus, three basic components of the interview process will be examined: interview questions, applicant information, and interviewer evaluation validity. The goal of this study is to examine the relationships among these components of employment interviews.

The proposed relationships among interview questions, applicant information, and interviewer evaluation validity are illustrated in Figure 1. This model applies primarily to traditional (unstructured) employment interviews, during which the interviewer has considerable latitude regarding what questions are asked and in what manner. In this type of interview, the interviewer also sets the tone of the interview, decides what the topic of conversation will be, and when that topic will change. These factors reflect the interviewer's opinion of the best way to assess the job-applicant fit.

The interviewer's belief regarding the best way to evaluate applicants can be discovered by examining the discourse in the interview. That is, the questions that the interviewer asks or does not ask is a function of his or her approach to interviewee evaluation. For example, an interviewer may think that intimidating applicants is the best way to assess their suitability for a job. This interviewer would probably refrain from engaging in chitchat with interviewees and simply hurl questions regarding job qualifications at them. Conversely, an interviewer may think that creating an open, relaxed atmosphere is the best way to assess applicant abilities.
Figure 1. Model of Employment Interview Information Gathering

Individual interviewer differences

Situational factors

Question type

Interviewee information

Interviewer evaluation validity

Individual interviewee differences

Situational factors
This interviewer might inquire about applicants' non-work interests in an attempt to put them at ease before exploring their suitability for a job.

Another dimension of interviewing strategy is question style. Interviewers can choose from several types of questions to inquire about virtually any topic. For example, the interviewer may ask primarily broad, unstructured questions that can be answered in a number of ways. Alternately, he or she may ask many questions that can be answered sufficiently with a simple "yes" or "no". The type of questions that are asked in the interview is the first basic component of the model.

The second basic component of the model addresses the product of the interview, or the information that the interviewer accumulates about the applicant during the interview. Some information can be gathered from the nonverbal behavior of the applicant, such as amount of eye contact, body orientation, and posture. However, the primary means of gathering information about candidates and their suitability for a job is through their answers to the questions that the interviewer asks. Because interviewers are in control of which questions are asked, and thus answered, they exert some influence over the information that they gain from the interview. For example, an interviewer may ask a recent college graduate who is applying for a managerial position about his or her extra-curricular activities. If this inquiry reveals that the applicant directed a university production of a play, planning and organizing skills that his or her work history did not expose may be discovered. However, if the interviewer does not provide an opportunity to discuss this event, the applicant's planning and organizing skills are likely to go unrecognized.

Following the interview, the interviewer is faced with the task of evaluating the candidate's suitability for the job. This is the final basic
component of the model. The interviewer combines information gained from the interview with any other job-relevant information that may be available (e.g., letters of recommendation, resume, test scores). Then the interviewer assesses the applicant's suitability for the job based on his or her prediction of how well the interviewee would perform on the job.

However, because interviewers are human, they are anything but unbiased gatherers and processors of information. Therefore, the chain of events depicted in the model (see Figure 1) may be influenced by other factors that are theoretically irrelevant to rendering an employment evaluation. One such factor may be interviewer training. Trained interviewers may be cautioned of the biases that can compromise the validity of their evaluations (e.g., confirmatory and disconfirmatory biases). Due to this training, they ask questions that should produce unbiased information. However, because they have formed a strong impression of the applicant prior to the interview (based on test scores, resume, letters of recommendation), they selectively attend to this information. Thus, it is entirely possible to ask questions that provide high quality information, and are therefore conducive to producing employment evaluations of high validity, yet actually produce evaluations of low validity.

Situational factors, such as the accountability that interviewers feel for their applicant evaluations, may moderate the proposed relationship as well. In many organizations, interviewers are not aware of how well the individuals they select perform, much less are held responsible for the quality of the applicants they selected (Arvey & Campion, 1982). If interviewers are held accountable for their employment evaluations and
receive feedback on their quality of their hires, they may change their questioning style.

Yet another factor that may affect the quality of interviewer decisions is individual differences in interviewers, such as interviewing experience. Experience in rating others has been found to correlate negatively with bias in ratings (Borman & Hallam, 1991). Furthermore, individual differences in stereotypes (e.g., gender, race) can affect performance ratings. For example, in a study by Dobbins, Cardy, and Truxillo (1988), performance ratings of women were less accurate when subjects held traditional female stereotypes than when they held nontraditional stereotypes of women.

These biases may also operate in those appraising the performance of employees. These biases may compromise the accuracy of the performance appraisal rating, and consequently attenuate the validity of the interview rating.

Other factors may compromise the validity of the performance appraisal rating as well. For example, the appraiser's motivation to rate employees accurately can play a substantial role in determining the quality of the performance appraisal rating. Oftentimes, raters are not rewarded for completing performance appraisals well (Heneman, Schwab, Fossum, & Dyer, 1989). Furthermore, performance appraisals are frequently viewed as a useless exercise because the organization does nothing with the performance appraisal information: good performers are not rewarded, and poor performers are not reprimanded (Schultz & Schultz, 1990). Moreover, supervisors frequently have difficulty in assigning poor ratings to employees. Consequently, the performance appraisal process is often lightly dismissed, and ratings are haphazardly assigned only when sufficient
pressure is applied to the rater (Schultz & Schultz, 1990). Given these organizational realities, it is understandable that performance appraisal ratings frequently do not accurately represent employee performance.

Situational factors may also moderate the relationship between interview and performance appraisal ratings. For example, employees simply may not have the opportunity to perform certain tasks on the job, and may be given a low rating on those particular skills because they have not been exhibited. Alternatively, the appraiser may not observe employees performing all relevant tasks, and rate them poorly with respect to certain tasks only because their execution had not been witnessed.

Despite these shortcomings, this study purports to examine the information gathering portion of the employment interview process and its relation to the validity of interviewer judgments. The theoretical rationale for expecting events in the model to affect interview evaluation validity will be presented subsequently.

**Interviewer Questioning Style in Unstructured Interviews**

In traditional (unstructured) employment interviews, the questions asked are determined by the interviewer's caprices. Interviewers may utilize many types of questions in assessing job-applicant fit, or rely on only a few types. Questions interviewers ask may be differentiated with respect to numerous dimensions, such as open-endedness, standardization, and relation to the job. They can also be categorized as directed, situational, or secondary. The frequency of use of each question type may differ among interviews, which may help account for differences in evaluation validity.

For example, interviewers may affect the validity of their evaluations through the environment they create with their questioning
style. For example, interviewers may improve the validity of their evaluations if they create an environment where the interviewee does most of the speaking. Often the interviewer does much of the speaking during employment interviews (Arvey & Campion, 1982; Dipboye, 1992). In making a valid judgment, however, the interviewer must gather a large amount of information about the interviewee. Indeed, Motowidlo et al. (1992) propose that questions framed and delivered in ways more likely to elicit large amounts of interviewee information contribute to interview validity. More information about the interviewee can be discovered when he or she is allowed to speak rather than when the interviewer speaks. Interviewers can create this situation by limiting their own comments and by encouraging the interviewee to speak (Dipboye, 1992). Under these conditions, an interviewer can gather maximal information about the applicant.

Similar to test scores based on a large number of questions versus those based on fewer questions, ratings based on a large amount of applicant information tend to be more reliable than those based on less applicant information (Allen & Yen, 1979). Thus, when interviewers limit their speaking time in the interview and encourage applicant disclosure, they may produce more valid evaluations than when they do not. This is because evaluations can only be valid to the extent that they are reliable (Feldman, 1981).

To this end, interviewers may ask open-ended questions. These questions are very broad and allow the applicant to decide what type and amount of information to provide when answering them. For example, the interviewer might ask the applicant to explain why he or she is a suitable candidate for the job. Answers to open-ended questions potentially provide
rich information to the interviewer by indicating what the applicant perceives is important in answering the question, as well as the content of the answer itself (Dipboye, 1992). This additional information may contribute to accurate evaluations of applicants because it provides more insight regarding the interviewee than questions which limit the applicant’s answers to a particular topic (e.g., Describe your past jobs that entailed duties similar to those you would execute in this job).

Another tool that interviewers may use as an effective means of assessing the applicant’s suitability for a job is the secondary question. This type of question is useful when interviewees fail to answer a previous question completely. Additional clarification or elaboration of an answer may be desired. For example, the interviewer may ask the interviewee to elaborate on job duties that were mentioned in response to a previous question. Jablin and Miller (1990) recommend the use of secondary questions because they concentrate on answers which the interviewee has already provided and allow the applicant to elaborate on the information that he or she feels is most important. Yet, the interviewer controls the discourse. Furthermore, Motowidlo et al. (1992) found that interviewers who skillfully asked probing questions elicited more relevant information about interviewees than those who were not as skilled in use of this type of question. Therefore, appropriate use of secondary questions may enhance the validity of interviewer evaluations.

The job-relatedness of interview questions may also influence the validity of interviewer evaluations. It seems obvious that questions asked during the interview should assess the candidate’s qualifications for the vacant position. To ensure that this is the case, a job analysis must be performed on the target job. Ideally, this information is then used to
construct questions which tap the knowledge, skills, abilities, and other characteristics that are required to perform the target job.

Intuitively, questions regarding job characteristics should produce information that is more useful for assessing the job-applicant fit than questions about topics unrelated to the job. Indeed, Weisner and Cronshaw (1988) found that use of questions in structured interviews based on formal job analysis resulted in more valid judgments than those constructed after a less "systematic assessment" of job requirements. In their meta-analysis, McDaniel, Whetzel, Schmidt, and Maurer (1991) also found higher validity of job related interviews compared to psychologically-based interviews. Furthermore, the use of information relevant to the job has been shown to decrease the effects of irrelevant information on employment decisions (Wiener & Scheiderman, 1974).

What is less apparent is that inquiring about topics unrelated to the job may actually decrease the probability of the interviewer producing a valid employment evaluation. This can happen when an evaluator is given information irrelevant to the decision at hand, such as interviewee marital status. This information, when accompanied by other information pertinent to the forthcoming evaluation, may actually dilute the influence of the relevant information on the evaluation (e.g., Locksley, Hepburn, & Ortiz, 1982; Nisbett, Zukier, & Lemley, 1981). This can happen when information that is irrelevant to making a judgment appears useful because it had been helpful in making evaluations in other contexts. For example, during an interview for an unskilled position, an interviewer may learn a candidate's high school grade point average. This information is presumably irrelevant to assessing qualifications for the job. Nonetheless, it is then combined with the other relevant information available to the
evaluator (e.g., job skills), and subsequently dilutes the effects of the clearly relevant information on the evaluation. Therefore, questions that are relevant to assessing job qualifications should be more helpful in making valid employment evaluations than those irrelevant to estimating job qualifications.

Another question that interviewers sometimes use is the situational question. When interviewers use this type of question, they may produce more valid evaluations than when they do not. Situational questions pose job situations to interviewees, who then describe how they would respond. For example, the interviewer may ask, "Suppose your boss asked you to supervise your co-workers while she was on vacation. During this period, your best friend and office-mate is an hour late 50% of the time. What would you do if the boss asked for a report of the activities that occurred in her absence?"

It is assumed that interviewees will behave on the job in the way they proposed during the interview. Latham & Saari’s research on situational questions (1984) supports this assumption. Actions which applicants said they would take in response to situational questions correlated well with supervisor behavioral observations ($r = .39$) and peer behavioral observations ($r = .42$).

Research on the validity of interviews utilizing situational questions is encouraging as well. Numerous studies (e.g., Latham & Saari, 1984; Latham, Saari, Pursell, & Campion, 1980; Weekley & Gier, 1987) reveal that use of situational questions in interviews produce employment evaluations of higher criterion-related validity than those made in traditional employment interviews.
Another element of an interviewing approach refers to the degree of standardization that an interviewer uses when evaluating applicants. Some interviewers adjust their evaluation techniques for each applicant, while others rely on similar questions for all interviews. When interviewers ask questions that have been composed prior to the interview of all applicants in the same order, a structured interview approach is used. Jablin and Miller (1990) found that interviewers using structured interview approaches made more accurate and less biased applicant evaluations than those utilizing an unstructured questioning strategy. Numerous other researchers and reviewers (e.g., Arvey & Campion, 1982, Dipboye & Gaugler, 1992; McDaniel et al., 1991; Weisner & Cronshaw, 1988) have consistently found higher criterion-related validity coefficients with the use of structured interviews as opposed to less structured interviews. This approach may influence evaluation validity for a number of reasons.

Under ideal conditions, job analysis information is used to construct structured questions which tap the knowledge, skills, abilities, and other characteristics that are required to perform the target job (Wright, Lichtenfels, & Pursell, 1989). All applicants are compared on this information rather than on domains about which interviewers spontaneously decide to inquire. Answers are then scored according to a predetermined scale. Because all applicants are asked the same questions, more accurate and fair comparisons can be performed, as opposed to those made with differing information bases (Arvey & Campion, 1982). In their meta-analysis, Weisner & Cronshaw (1988) found that structured interviews conducted by single interviewers produced mean validity coefficients (corrected for range restriction and criterion unreliability).
three times the size of those found with unstructured interviews ($r = .63$ v. $r = .20$).

Structure can help interviewers gather similar information from interviewees and assist them in discriminating among applicants who are well and poorly suited for the job. However, interview structure is only as good as the content of the questions asked and the method of information gathering. For example, interviewers may use a "structured" interview in which they consistently ask all applicants about issues that are unrelated to the job. Alternatively, they may focus on identical job-related topics for each applicant, yet phrase questions differently for each, thus influencing the answer of each in a different manner. Therefore, in practice, "structure" may not mean the same thing in all interviewing situations, and may not be equally beneficial in all interviews.

Nonetheless, standardized, job-related, open-ended, situational, and secondary questions are hypothesized to contribute to valid judgments made in the employment interview. Other types of questions may compromise the validity of selection evaluations. Perhaps when interviewers make more valid applicant evaluations, they use these questions less frequently than when they make less valid evaluations.

One type of question that may not assist interviewers in predicting employee performance is the directed question. This type of question provides clues about which answer the interviewer thinks is most appropriate. For example, the interviewer might ask, "You wouldn’t fire an employee before discussing performance problems and allowing time to correct them, would you?" Directed questions may be leading, and imply that a specific answer is desired. They may be loaded, suggesting that one form of response is more socially desirable than another (Jablin & Miller,
1990). In either case, directed questions shape the applicants’ responses. They act as demand characteristics to the interviewee. When interviewers rely on these types of questions, they may bias their own evaluations in a way that decreases their validity.

In light of these findings, use of different types of questions may result in employment interview evaluations that are differentially valid. Specifically, the greater the number of job-related, open-ended, situational, and secondary questions that are asked, the more valid the employment evaluations should be; the fewer the number of questions of these types that are asked, the less valid the employment evaluations should be. Furthermore, when interviews are highly standardized with respect to the order, number, and content of questions, the interviewer should produce employment evaluations of higher validity; when the interview is standardized to a lesser degree, the interview evaluations should be less valid.¹ This study proposes to examine these possibilities.

Thus, one goal of this study is to ascertain the relationship between interview question style and evaluation validity. A second goal is to understand this relationship by examining an intermediate variable: the diagnosticity of information provided by the interviewee.

Diagnostics in the Employment Interview

As the model in Figure 1 depicts, interviewers ask questions of candidates to gather information that will assist them in evaluating the job-applicant fit. Interview question style helps to determine what information is gained from the execution of the interview. Because question style can influence the quality and amount of information gained from the interviewee, it is a critical element of the interview process. If the interviewer receives accurate information regarding how well the applicant
can perform the job duties, his or her judgments are potentially sound. If, however, the interview does not reveal such information, the interviewer's evaluation is likely to be less valid. Indeed, Motowidlo et al. (1992) propose that a critical feature of an interview is the type of information that it elicits. Therefore, it is important to examine the relationship between question style and the quality of information provided by the applicant.

There is a body of research addressing this issue, in both the social cognition and interview literature. It concerns the diagnosticity of interviews, or the extent to which the information that is gathered allows interviewers to accurately categorize interviewees. Highly diagnostic information appears to be the ideal product of an employment interview.

Diacnosticity has been studied and defined in many ways. Some researchers approach this variable from the perspective of the questions that subjects ask. For example, Trope and Bassok (1983) defined diagnosticity by the extent to which questions addressed features that were judged to discriminate between two categories of individuals (introverts and extroverts). They measured subjects' preference for questions high in diagnosticity over those lower in diagnosticity. Similarly, Trope, Bassok, and Alon (1984) asked subjects to generate questions that they thought would help distinguish extroverts from introverts. Raters then estimated the diagnosticity of each question (i.e., to what extent can each question discriminate among individuals in each category).

Skov and Sherman (1986) also measured the diagnosticity of questions that subjects generated. They manipulated information diagnosticity by changing the probability of the presence or absence of a given characteristic in a sample. In their study, subjects were told to ask
questions of a member of a group of aliens with different characteristics (e.g., had webbed feet, wore pink suspenders). The proportion of each type of alien with each characteristic was varied. Subjects' task was to ask questions answerable with "yes" or "no" in order to determine which type of alien a target was.

Devine, Hirt, and Gherke (1990) addressed question diagnosticity in another way. They considered question diagnosticity a function of question focus. Questions inquiring about the less restrictive of the traits tested in a hypothesis were considered more diagnostic than those addressing the more restrictive of the traits. A nonrestrictive trait is that which is associated with a wider range of behaviors than its more restrictive opposite. Subjects generated questions to find if a target possessed a given trait. Question diagnosticity was assessed in terms of whether the question was phrased in terms of the more or less restrictive trait.

Swann and Guliano (1987) examined questions that subjects generated and questions that subjects chose from a list with respect to how informative they were in distinguishing targets in terms of two traits. The choices were also examined with respect to the trait they addressed (i.e., the hypothesized trait versus the alternative trait). Similarly, Kruglanski and Mayseless (1988) examined subjects' question preferences. Subjects in their study chose questions with predetermined diagnostic values from a list of questions intended to distinguish between individuals in two separate occupations.

Still other researchers have examined diagnosticity from the perspective of the information about a target person. For example, in their review, Skowronski and Carlston (1987) discuss behavioral information in terms of its category diagnosticity, or the ability of the information to
distinguish a target person with respect to alternate categories (e.g., honest v. dishonest). Conversely, Fischhoff and Bar-Hillel (1984) consider diagnosticity synonymous with representativeness. They categorized information in terms of this variable by the degree to which it was differentially suggestive of membership in different occupations (e.g., business executive, university professor).

In previous studies of diagnosticity in interviews (e.g., Trope & Bassok, 1983; Skov & Sherman, 1986; Trope, Bassok & Alon, 1984), diagnosticity has been operationalized as a characteristic of questions. That is, a diagnostic question is one that potentially elicits diagnostic information from the target. It is important to note that the quality of diagnosticity, which refers to a type of information about the target, has typically been measured in terms of the questions that are asked.

In the employment interview literature, studies on diagnostic questioning have produced mixed results (Day, 1990; Dipboye & Jackson, 1990). Dipboye and Jackson (1990) rated this variable in terms of the extent to which interviewer questions inquired about gaps that had been written into the target's resume. Day (1990) addressed diagnosticity in several ways: in terms of the extent to which subjects addressed key criteria and details in the target's application, with a rating of subject confidence in their questions' ability to evaluate target qualifications, and with a direct subject rating of how well their questions could distinguish among well and poorly qualified applicants. Authors of these studies have reported problems in measuring diagnosticity with these methods (personal communication with Day, July, 1992; personal communication with Jackson, September, 1992).
Hilton and Fein (1989) have further refined the operationalization of information diagnosticity. They have differentiated between types of diagnostic information: typically diagnostic information and judgment-specific diagnostic information. Typically diagnostic information is helpful for predicting individual membership across a large number of social judgment categories. A student's grade point average (GPA) may be considered typically diagnostic information: it helps others to predict membership in academic clubs, likelihood of future employment or admission to graduate school, etc. What the student ate for lunch, however, is only helpful in predicting whether or not he or she is a vegetarian, or perhaps a gourmand. This latter type of information, judgment-specific information, is useful in categorizing individuals in only a small number of categories. Its value is situationally-driven. Information can be high in typical diagnosticity and low in judgment-specific diagnosticity, or vice versa. It may also be high or low in both types of diagnosticity.

For example, GPA information is high in typical diagnosticity and also high in judgment-specific diagnosticity if one is assessing likelihood of membership in an academic honorary club. However, in assessing physical fitness, GPA is low in judgment-specific diagnosticity, but it is still high in typical diagnosticity.

Although a person's lunch is low in typical diagnosticity, when one wishes to assess likelihood of subscribing to a gourmet cooking magazine, this information is high in judgment-specific diagnosticity. However, if one wishes to predict this person's mathematical skill, the lunch contents become low in both types of diagnosticity.
Typically diagnostic information is more likely to be useful in any categorization task (e.g., honest v. dishonest person) than judgment-specifically diagnostic information. However, judgment-specifically diagnostic information is generally more useful than typically diagnostic information when one is in a situation appropriate for using the judgment-specifically diagnostic information. When judging information with respect to diagnosticity, then, Hilton and Fein argue that both types of diagnosticity must be addressed in order to assess the effect that information can have on judgments.

In Hilton and Fein's (1987) study, when subjects were asked to rate stereotyped targets on a given characteristic (and were given information clearly irrelevant to performing this task), they rated the targets according to the stereotype information only. When subjects were provided with information that was typically diagnostic, but not diagnostic for that situation, effects of the stereotype were obliterated: targets were rated equally on the characteristic. This finding illustrates the influence of typically diagnostic information and also the necessity for a dual assessment of information diagnosticity. In other studies addressing diagnosticity (e.g., Fischoff & Bar-Hillel, 1984; Skov & Sherman, 1986; Trope & Bassok, 1983; Trope, Bassok, & Alon, 1984), this variable has been examined unidimensionally, from only the judgment-specific diagnosticity perspective.

In an employment interview context, judgment-specifically diagnostic information is that which helps interviewers accurately categorize applicants in terms of job qualifications. According to Hilton and Fein, applicant information may be high in typical diagnosticity, judgment-specific diagnosticity, both, or neither.
For example, if an interviewee indicates that he or she has never used illegal drugs, the interviewer would gain information high in typical diagnosticity. This information can be used to categorize the applicant as a responsible, law-abiding citizen, as well as someone who is not likely to frequently miss days of work due to a "bad high." Furthermore, in predicting whether or not this person is a drug abuser, this information becomes high in judgment-specific diagnosticity as well. However, when trying to determine if the individual is physically capable of performing strenuous job duties, this information loses its judgment-specific diagnosticity.

Information indicating that an applicant would only accept a position in his or her current geographical location is high in judgment-specific diagnosticity if one is predicting how easy it will be to place this person if he or she is hired, or perhaps whether or not he or she would accept a transfer. However, it is low in typical diagnosticity because the situations in which it is helpful in categorizing the individual are extremely limited.

Even less useful may be the information that the individual had been laid off from a previous position. It is low in both typical and judgment-specific diagnosticity because it reveals nothing about the applicant except that he or she is unemployed, ostensibly through no fault of his or her own. It is not particularly useful in predicting category membership in any job-related domain.

In the present investigation, diagnosticity is defined and measured directly as a characteristic of applicant information rather than as a quality of interviewer questions. Although it is hypothesized that certain types of questions are more likely to elicit highly diagnostic information than others (e.g., standardized, open-ended), the potential of interview questions to
elicit information useful for categorizing interviewees is not the variable under scrutiny. Diagnosticity will be measured as a direct quality of applicant information. To the best of our knowledge, there are no studies in the employment interviewing literature that approach diagnosticity in this way.

In summary, diagnostic information helps the interviewer determine the applicant's degree of suitability for the job, which in turn helps the interviewer produce an employment evaluation of high validity. This will be the second proposition tested in this study.

Level of Analysis in Employment Interview Studies

Many studies regarding the criterion-related validity of structured and unstructured interviews have emerged from research which has been conducted at the aggregate interviewer level (e.g., Orpen, 1985; Weisner & Cronshaw, 1988; Wright, Lichtenfels, & Pursell, 1989). That is, the results of several interviewers' decisions are pooled in creating a validation sample. The validity coefficient computed on this sample is then assumed to be representative of the validity achieved when using a particular type of interview (e.g., structured vs. unstructured). That is, the validity of a particular style of interview is tested.

This practice makes a number of erroneous assumptions regarding interviewers. First, it assumes that all interviewers possess the same type of decisional biases and that they have roughly equal ability to make accurate evaluations (Dreher et al., 1988). Research indicates that individuals differ on many characteristics that may influence the accuracy of their judgments, such as intelligence, verbal listening skill, investigative interest, interest in observing others, and detail orientation (Borman, 1979; Schuh, 1978; Taft, 1950; Wedeck, 1947). Furthermore, Zedeck, Tziner, &
Middlestadt (1983) suggest that interviewers differ with respect to response biases which may cause systematic differences in interviewer ratings. For example, interviewers may assign different values to identical true levels of performance due to bias toward or against certain scale values. Zedeck et al. (1983) argue that these and other biases (e.g., halo, leniency, central tendency) cannot be identified, much less rectified, unless analyses are conducted at the individual interviewer level.

Second, the practice of combining interviewer judgments presupposes that all interviewers use the same type of information to make their evaluations. However, researchers have found individual differences in information use (Dougherty, Ebert, & Callender, 1986; Zedeck et al., 1983). For example, Zedeck et al. (1983) found that interviewers differ in their use and integration of information gathered in employment interviews. Specifically, interviewers relied on different dimensions to make their ratings and were influenced most heavily by different dimensions. Policy-capturing studies of decision makers confirm that individuals are influenced by information differently in making judgments (Dougherty et al., 1986).

A third reason to abandon the technique of collapsing evaluations across interviewers is individual differences in the conduct of the interview. Conduct of the interview can significantly affect interviewer rating accuracy. For example, Schuh (1978) found that differences in the use of techniques such as note taking and choice of work environment influenced listening accuracy. In his study, subjects who took notes during the interview recalled the interview content more accurately than those who did not take notes. Similarly, Macan and Dipboye (1986) found that note taking can lead to more accurate recall of information in the
interview. Schuh (1978) also found that subjects who were interrupted during the interview were less accurate in recall of interview content than those who were not interrupted. The use of aggregate validation samples assumes that either there are no significant differences in interview conduct or that any such differences do not affect interviewers' evaluations of applicants. This may not be the case.

Fourth, interview data should not be aggregated across interviewers due to possible variations in the degree to which interviewers structure their interviews. Researchers have repeatedly found that structured (hence, standardized) interviews yield criterion-related validity coefficients greater than those obtained through the use of unstructured interviews (McDaniel et al., 1991). In unstructured interviews, the interviewer has the freedom to impose a degree of standardization on the interview process. Because standardizing the interview is optional, interviewers are likely to use standardization to different degrees. Since this factor can significantly affect the validity of interviewers' judgments, interviewer evaluations derived from interviews with different degrees of standardization should not be collapsed in a validation sample.

Finally, as previously discussed, interview data should be analyzed at the individual interviewer level because the type of questions that interviewers ask may vary in unstructured interviews. Some interviewers may primarily ask questions that can be answered in a few words. Conversely, others may rely upon questions that encourage substantial elaboration on the part of the applicant. In unstructured interviews, interviewers often develop an individual method or policy for conducting interviews (Dougherty et al., 1986; Zedeck et al., 1983). This policy may include reliance on some question types more than others, or avoidance of
particular question types. Differences in use of question type may affect the quality and quantity of applicant information on which the employment decision is based. This, in turn, may affect the validity of interviewers' evaluations.

Because individuals differ in the way they make evaluations and in their ability to make accurate ratings, one might infer that the effectiveness of individuals' judgments differ as well. Researchers have found support for this notion: interviewers do not produce equally valid employment evaluations (Dipboye, Gaugler, Hayes, & Parker, 1990; Jablin & Miller, 1990; Zedeck et al., 1983).

Thus, the practice of combining data across interviewers appears misleading in that it masks differences in predictive validity, use of information, evaluation biases, and execution of interviews among individual interviewers. Therefore, the results of meta-analytic investigations of employment interviews in which unstructured interviews are grouped together may be confounded by these differences.

In trying to produce valid employment evaluations, the notion of a valid interview format may be erroneous; the critical element may be the individual performing the interview. Thus, perhaps we should be seeking the best individual strategies for conducting interviews in addition to the best format (e.g., structured vs. unstructured) for interviewing. This study attempts to do exactly that by examining individual interviewer questioning style and its relationship with evaluation validity.

The Present Investigation

The purpose of this study is to test the basic premise of the employment interview information gathering model presented in Figure 1: that interview question style affects the validity of interviewer decisions
through the quality of the applicant information it provides. Two hypotheses were examined (at the aggregate interviewer level):  
1. The more interviewers use questions that are open-ended, job-related, situational, secondary, and standardized, the more diagnostic the information gathered from the interview will be; the less interviewers use questions that are open-ended, job-related, situational, secondary, and standardized, the less diagnostic the information gathered from the interview will be.  
2. The more diagnostic the information that the interviewer gathers from the interview, the more accurate and valid judgments of applicants will be; the less diagnostic the information that the interviewer gathers from the interview, the less accurate and valid judgments of applicants will be.\textsuperscript{2}  

As an exploratory analysis, these hypotheses will be tested at the individual interviewer level. Insufficient number of interviewers available for the present study precludes a more rigorous examination among model components at this level of analysis.

The relationships among the three basic components in the proposed interviewer information gathering model will be examined from several perspectives. Factors thought to moderate the relationship among the three basic components of the model will not be tested.

Method

Subjects

Three interviewers at a large correctional institution participated in the study. All were white men. Their average age was 49 years, ranging from 34 - 55 years. The educational background of the interviewers varied: one interviewer had a bachelor’s degree in criminology, another had a bachelors’ degree in history, and the third had no post-secondary
education. On average, the interviewers had 7.22 years of interviewing experience prior to employment at this organization (range: 0 - 21.6 years). All three interviewers received on-the-job training at the target organization, which consisted of observing a trained interviewer for at least one day and receiving supervision and coaching from a supervisor in the personnel department for approximately one week. During training, interviewers were given feedback regarding their verbal and nonverbal communication, appropriateness of probes, and legality of questions. The interviewers who participated in this investigation had been on the job for an average of 2.8 years at the beginning of the study (range: 2.25 - 3.8 years).

The interviewers had each recommended for hire approximately 50 applicants, thus warranting participation in this study. Although there were actually four interviewers in this organization available for participation in the study, insufficient statistical power precluded studying the fourth interviewer, who had recommended fewer than 40 individuals for hire. On average, 50 of the applicants that each interviewer had evaluated favorably were included in the study (range: 49 - 50 applicants). There were additional applicants available for study for two of the interviewers in the study. However, the intention in this study was to compare interviewers’ behavior. To allow for meaningful statistical comparison of questioning behavior and elicitation of diagnostic information, roughly equal sample sizes were maintained for all interviewers.

One hundred forty nine hirees participated in the study. Seventy-one percent were male. Seventy-one percent were white, 16 percent were black, 11 percent were hispanic, and two percent were of an unidentified
race. Their average height was 69 inches; their average weight was 177 pounds.

**Application Process**

Candidates applied for an entry-level correctional officer position. Applicants were either former employees or were referred to the target organization by the Texas Employment Commission or by a college education program designed to train individuals for the target position. Applicants first completed an application blank and provided supporting documents such as court depositions and proof of high school graduation. Upon satisfactory completion of the application, applicants took the Basic Employment Security Test (BEST). They were required to score above the 25th percentile on the test to be further considered for the position.

Those scoring acceptably on the BEST and showing no outstanding criminal warrants or felonies on their record were invited to interview for the position. Their height and weight was compared to recommended proportions specified by the Federal Bureau of Prisons (cited in Texas Department of Criminal Justice, 1986). Those applicants within the acceptable height and weight range were fingerprinted and administered a cognitive ability test. Five sections comprised this test: (a) observation and memory, (b) situational reasoning, (c) reading comprehension, (d) vocabulary, and (e) arithmetic. Applicants scoring above the 25th percentile on each section of this test were then interviewed by a single interviewer. The content and structure of the interviews were determined by individual interviewer discretion.

Interviewers rated interviewees on ten dimensions identified through job analysis as important for successful performance in the correctional officer position: (a) attitude/motivation, (b) perceptiveness/alertness, (c)
personal integrity, (d) learning ability, (e) physical ability, (f) emotional stability/stress tolerance, (g) interpersonal skills, (h) judgment/common sense, (i) responsibility, and (j) communication skills. Applicants' evaluations were based on their behavior in the interview, scores on the cognitive ability test, reference letters, height and weight information, and other ancillary data. The sources of information that interviewers were to consider in rating applicants on each dimension were specified in written rater guidelines. Applicants were rated either low, medium, or high on each of the interview dimensions. Ratings were then weighted according to their relative importance in contributing to effective job performance. These weights were determined by upper-level supervisors (subject matter experts) in the target organization. The weighted scores were combined to form an overall interview score. This overall score was utilized in all validity analyses because it is how the target organization used the interview information to make selection decisions.

Criteria

The validity of interviewer judgments was calculated against two types of criteria. The first was performance appraisal ratings which had been created by the employees' immediate and/or higher level supervisor. Ratings were based on employees' performance on the following dimensions: (a) quality, (b) productivity, (c) knowledge, (d) cooperation, (e) responsibility, and (f) dependability. Ratings ranged from 1 (unsatisfactory) to 5 (excellent) on each of the dimensions. Employees were also given an overall rating of their performance on the job. On average, employees were rated after 120 days of employment the first time; after 180 days of employment the second time.
The applicants' overall scores on the interview were validated against the overall score on the employees' first two performance appraisal ratings. Two appraisals were chosen for validation purposes because the supervisors articulated that the first performance appraisal was conducted too soon after employees completed training school to be meaningful. Furthermore, they indicated that scores on this appraisal varied minimally.

A nonjudgmental measure of performance was also examined. Employee turnover was measured up to two years from the employee's date of hire. This was the second criterion utilized in this study.

Materials and Coding Procedures

One hundred forty nine audio taped interviews were used in this study. All met the following criteria: they were audible and contained some applicant identification information. Pre-interview information was available for the applicants on the tapes, the applicants were rated with the same interviewing form, and the interviewee had worked in the organization long enough to receive at least one performance appraisal rating. These tapes were coded for the presence of the six interviewer questioning variables examined in this study: (a) open-ended, (b) secondary, (c) situational, (d) job-related, (e) standardized, (f) directed. Open-ended questions are broad questions that typically encourage the applicant to speak at length. Secondary questions are probing questions that the interviewer asks to gain more information on a topic of interest. Situational questions are those posing job situations to candidates and asking how the candidate would respond. Job-related questions are those inquiring about topics related to job performance along dimensions identified through job analysis. Standardized questions are those which the
interviewer asks of most applicants. Directed questions are those which are posed with cues regarding the desired or expected response.

In addition, applicant responses were coded in terms of their typical and judgment-specific diagnosticity. Diagnosticity was coded in terms of the number of social judgment situations in which the applicant information could be used to categorize the interviewee, and the extent to which the interviewee information was helpful in categorizing the interviewee across all job domains. The coding conventions and coding forms for these variables are presented in Appendix A.

Question and diagnosticity variables were measured by four coders. Each coder gathered data from 48, 43, 30, and 28 interviews. All coders gathered data from interviews conducted by each of the three interviewers in the study. Interview assignment was not restricted in any other way.

An inter-rater agreement check was conducted three times for each coder (each check was performed with a different interviewer's data). For the question variables, a measure of inter-rater agreement was calculated with the method described by Tinsley and Weiss (1975). The average inter-rater agreement coefficient was .90; coefficients ranged from .71 (directed questions) to .99 (integrity). For the diagnosticity variables, inter-rater agreement measures were calculated with the method described by James, Demaree, and Wolf (1984); all coefficients were .99.

Rating Accuracy

Validity of the interview ratings was estimated for each applicant with a measure of rating accuracy. This measure was chosen because, 
"...reliability and validity are necessary, but insufficient conditions for accuracy" (Sulsky & Balzer, 1988, p. 498). Although rating accuracy is an imperfect representation of validity, it does provide some indirect evidence
regarding the quality of the interviewers' applicant ratings. The accuracy measure represented how similar each interviewee's overall rating was compared to a corresponding true score estimate.

True score estimates for each interview were made by the primary investigator in the study. The investigator had studied the selection system at the target organization, had taken graduate-level coursework in selection, and has had some experience in rating human performance. One other expert in industrial and organizational psychology assisted in creating true score estimates for each interviewee. This expert has had extensive experience with the selection procedures at the target organization and has taught graduate level selection courses. Both coders gathered data from interviews conducted by each of the three interviewers in the study. Interview assignment was determined by assigning the interviews that the primary investigator had coded with respect to question and diagnosticity variables to the second coder. The first coder created true score estimates for the remainder of the interviewees.

The true score estimates were based on review of the interview rating guidelines that were provided to the interviewers in the target organization. The rating form that the interviewers in the study had used to rate applicants were employed by the coders. These materials are presented in Appendix B. The coders were blind to the rating that had been assigned to the applicants by the interviewers. Pre-interview information that the interviewers had been given (i.e., test scores, application, etc.) were provided to the raters with the exception of any visual applicant information (e.g., interviewee dress, eye contact). According to the interview scoring guidelines, this information may be
relevant in determining ratings for the dimensions of attitude/motivation, judgment/common sense, and physical ability.

Accuracy of single interviewee scores was calculated according to Cronbach's (1955) overall index of accuracy. This measure refers to the absolute difference between the applicants' true score and the score that had been assigned by the interviewer.

\[ D^2 = \sum_{kn} \sum_{nk} (x_{nk} - t_{nk})^2 \]

where \( n = \text{ratee}, k = \text{dimension}, x = \text{subject rating}, \) and \( t = \text{true score}. \) This calculation was divided by the number of interview dimensions to yield an average accuracy rating for each interview.

An inter-rater agreement check was conducted three times for each coder (each check was performed with a different interviewer's data). Inter-rater agreement was calculated with the method described by Tinsley and Weiss (1975); the mean coefficient was .85 (range: .78 to .89).

Results

**Relationships Among Study Variables**

The intercorrelations among all variables measured in this study are presented in Appendix C. Examination of this appendix reveals that the interview dimensions of judgment/common sense, responsibility, communication skill, and learning ability correlated the highest with the applicants’ overall interview score (mean \( r = .43 \)); the dimensions of perceptiveness/alertness, physical ability, and emotional stability/stress tolerance correlated the lowest with the applicants' overall interview score (mean \( r = .18 \)). The true score ratings for the dimensions of attitude/motivation, perceptiveness/alertness, communication skill, and learning ability correlated the most strongly with the applicants’ overall
true score estimate (mean $r = .58$), while the dimensions of physical ability and interpersonal skill correlated the least strongly with the applicants' overall true score estimate (mean $r = .09$). This suggests that the interviewers and the true score coders relied on somewhat different information to create the applicant evaluations. Unfortunately, neither the applicant ratings made by the interviewers nor those made by the accuracy coders significantly correlated with job performance ($r = -.03$; $r = .05$, respectively) or turnover behavior ($r = -.03$; $r = -.05$, respectively).

The question and diagnosticity variables correlated significantly with one another (absolute value mean $r = .26$). However, none of these variables correlated significantly with interviewer accuracy (absolute value mean $r = .05$).

**Aggregate Interview Validity**

The validity of the interviewers' ratings at the aggregate interviewer level was calculated with the traditional validation method, and also with that of Dreher et al. (1988). The traditional validation method (computing the bivariate correlation between predictor and criterion scores) is often employed with pooled predictor and/or criterion scores. That is, scores for interviewees created by multiple interviewers are pooled, then correlated with performance scores generated by multiple supervisors. This procedure assumes that the rater or interviewer factors (e.g., individual differences among interviewers) do not influence applicant ratings. Independence among observations made by each interviewer or rater is also assumed.

The Dreher et al. method of validation reflects the fact that multiple interviewers' judgments comprise the pool of predictor scores. When using this technique, the mean and standard deviation of both the predictor
and the criterion scores, as well as the correlation between the two sets of scores, are calculated for each interviewer. These statistics are then weighted according to the sample size for each interviewer. These figures are combined such that individual interviewer differences in predictive validity are reflected in the aggregate total. This method of calculating validity coefficients was also employed in this study.\(^3\)

After validity coefficients were calculated, they were corrected for the effects of criterion unreliability. In their meta-analysis of employment interviews, McDaniel and colleagues (1991) adopted Pearlman's (1980) estimate of average reliability of supervisor ratings of employees (.60). The validity coefficients in this study (using supervisor ratings for the criterion) were corrected for unreliability at that level. Then, validity coefficients were corrected for the effects of range restriction in the predictor. Applicants included in this study were hired; they performed at a minimally acceptable level or higher in the interview. Therefore, validity coefficients were adjusted to offset the attenuating affects that may have resulted. The ratio of the unrestricted sample's standard deviation of overall interview scores to that of the restricted sample was 1.72.

Validity coefficients were computed on both standardized and unstandardized interview and performance scores. Because the differences in the resulting coefficients were too small to warrant separate treatment (e.g., the aggregate validity coefficients for applicants' first performance appraisal score based on both unstandardized and standardized ratings are .01 and .01, respectively), only the results calculated with the unstandardized scores were reported. The validity coefficients calculated at the aggregate interviewer level are presented in Table 1.
Table 1

Aggregate Unpartialed Validity Coefficients

N = 145-149

<table>
<thead>
<tr>
<th>Criterion</th>
<th>Appraisal 1</th>
<th>Appraisal 2</th>
<th>Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Predictor</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Aggregate Interviewer</td>
<td>.01</td>
<td>-.07</td>
<td>-.03</td>
</tr>
<tr>
<td>Corrected Aggregate</td>
<td>.02a(.01)ab</td>
<td>-.12a(-.09)ab</td>
<td>-.07a(-</td>
</tr>
<tr>
<td>.04)ab Interviewer</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dreher, Ash, &amp; Hancock (Aggregate)</td>
<td>.002</td>
<td>-.08</td>
<td>-</td>
</tr>
<tr>
<td>Corrected Dreher, Ash, &amp;</td>
<td>.01a(.003)ab</td>
<td>-.14a(-.10)ab</td>
<td>-</td>
</tr>
<tr>
<td>Hancock Tests</td>
<td>-.04</td>
<td>-.08</td>
<td>.01</td>
</tr>
<tr>
<td>Corrected tests</td>
<td>-.09a(-.05)ab</td>
<td>-.17a(-.10)ab*</td>
<td>.02a(.01)ab</td>
</tr>
</tbody>
</table>

Note. aCorrelations have been corrected for the effects of criterion unreliability.

bCorrelations have not been corrected for the effects of range restriction in the predictor.

* p < .05
According to the traditional approach, the uncorrected correlation coefficients between the aggregate interviewer ratings and the first and second performance appraisal scores were .01 and -.07, respectively. The uncorrected correlation between interviewer rating and turnover was -.03. When corrected for the effects of range restriction and criterion unreliability, these correlations became .02, -.12, and -.07, respectively. None of these relationships were statistically significant.

Validity of the interview ratings for predicting performance scores using the Dreher et al. technique was .002 and -.08. When corrected for the effects of range restriction and criterion unreliability, the magnitude of these coefficients increased to .01 and -.14, respectively. None of these relationships were statistically significant. Thus, at the aggregate level, the interviewers' ratings did not predict job performance, regardless of the technique utilized to perform this calculation.4

In creating the interviewees' ratings, the interviewers were explicitly instructed to consider test score information. For example, the scoring guidelines provided to interviewers indicate whether a high, medium, or low rating should be given to an applicant with a particular test score. Since interviewers were told to rely on this information, the relationship between test scores and criteria was examined.

The various tests differed conceptually, so a measure of internal consistency was calculated. Because the resulting coefficient alpha was not high (.56), a principal components analysis with varimax rotation was applied to the individual test scores. Only one factor resulted, so the test scores were treated in aggregate form in all of the analyses.

The poor predictive validity of applicants' test scores partially explained the failure of the aggregate interviewer ratings to adequately
predict performance criteria: the bivariate correlations between test scores and performance ratings were -.04 and -.08. Neither did test score information predict turnover behavior; the correlation between test scores and turnover was .01. When corrected for the effects of range restriction and criterion unreliability, the coefficients increased to -.09, -.17, and .02, respectively. The correlation between test scores and the second performance rating (-.17) was statistically significant (p < .05): applicants scoring higher on the pre-interview tests also received lower performance appraisal ratings at six months on the job than those who scored lower on the tests. Based on these results, one may question the wisdom of relying on this test score information to create interview ratings. Nonetheless, the information gathering model (to be tested subsequently) concerns the predictive validity of the interviewers' ratings independent of the test score information. Therefore, test scores were partialled from the interview ratings. The results of these analyses are presented in Table 2.

For exploratory purposes, the validity of the individual pre-interview tests was examined. None of these scores significantly predicted the criteria, with the exception of the vocabulary test. As stated above, this test correlated at -.17 (p < .05) with the second performance appraisal score. The uncorrected semi-partial correlations between the overall interview ratings and the two performance appraisal scores were .03 and -.04 (n. s.), respectively.

The uncorrected semi-partial correlation between overall interview rating and turnover was -.03. When corrected for range restriction and criterion unreliability, these coefficients increased to .07, -.09, and -.09, respectively. None of these relationships was statistically significant. However, the aggregate interviewer ratings predicted job performance
### Table 2

**Aggregate Semi-partial Validity Correlations**

N = 145-149

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Appraisal 1</th>
<th>Appraisal 2</th>
<th>Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate Interviewer</td>
<td>.03</td>
<td>-.04</td>
<td>-.03</td>
</tr>
<tr>
<td>Corrected Aggregate Interviewer</td>
<td>.07&lt;sup&gt;a&lt;/sup&gt;(.04)&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>-.09&lt;sup&gt;a&lt;/sup&gt;(-.05)&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>-.09&lt;sup&gt;a&lt;/sup&gt;(-</td>
</tr>
<tr>
<td>Dreher, Ash, &amp; Hancock (aggregate)</td>
<td>.01</td>
<td>-.05</td>
<td>-</td>
</tr>
<tr>
<td>Corrected Dreher, Ash, &amp; Hancock (aggregate)</td>
<td>.02&lt;sup&gt;a&lt;/sup&gt;(.01)&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>-.10&lt;sup&gt;a&lt;/sup&gt;(-.06)&lt;sup&gt;ab&lt;/sup&gt;</td>
<td>-</td>
</tr>
</tbody>
</table>

**Note.**<sup>a</sup>Correlations have been corrected for the effects of criterion unreliability.

<sup>b</sup>Correlations have not been corrected for the effects of range restriction in the predictor.
scores on the first performance appraisal better than scores on the second performance appraisal and better than turnover behavior. In fact, the relationship between interview ratings and the second performance score and that between interview ratings and turnover behavior was negative.

When the Dreher et al. method for validation was applied to the semi-partial correlations between interviewer scores and performance ratings, the uncorrected coefficients were .01 and -.05. When corrected for statistical artifacts, these coefficients became .02 and -.10, none of which was statistically significant.

There were slight differences in the magnitude of the validity coefficients calculated with the traditional and the Dreher et al. methods. For the first performance appraisal, the traditional technique yielded a larger validity coefficient than that produced by the Dreher et al. method, while the reverse was true for the validity coefficient for the second performance appraisal. However, the signs of the coefficients were identical, regardless of the technique used. Furthermore, neither technique produced statistically significant results, even when the test scores were partialled from the predictor.

The validity of the applicants' overall interview score was calculated in these analyses because this is the manner in which interview information was used in the target organization. To better estimate the validity of interview information as it was actually used by the interviewers, the validity of interviewers' ratings in factor form was calculated. The factor structure of interviewer ratings was taken from a factor analysis of the interview dimensions that was calculated on a sample larger than the one employed in this study, where both hires and non-hires were included (Kohn, 1993). Two interview factors were identified by retaining
interview dimensions loading on factors at .42 or higher. The first was comprised by the communication skill and learning ability factors; the second was comprised by the interpersonal skill, responsibility, personal integrity, and emotional stability dimensions. Then, hierarchical regression analyses were performed. The pre-interview test scores were entered at the first step, then the interview factors. The test scores and interview factors did not account for a significant proportion of variance in the first or second performance appraisal scores, or in turnover behavior ($R^2 = .04, R^2 = .04, R^2 = .23$, respectively).

**Individual Interviewer Validity**

Researchers have discovered that interviewers differ in many ways, including the validity of their applicant evaluations. Therefore, regression analyses were performed to determine if the proportion of variance in the criteria explained by the interview scores differed across interviewers. This determination can be made by regressing the criteria on the interaction between the overall interview score and interviewer. Two coded vectors were created to represent the three interviewers; interaction terms were created by multiplying these vectors with the interview score. Performance scores and turnover behavior were regressed on these interaction terms, the interviewer vectors, and the interview score. The results of these analyses are presented in Appendix D.

The interaction was nonsignificant in analyses with the first ($F = 2.23; F = 1.7$) and the second ($F = 1.22; F = .10$) performance scores, as well as turnover behavior ($F = 1.9; F = .03$). These results indicate that the proportion of variance in the criteria accounted for by the interview ratings does not differ according to interviewer. Nonetheless, the analyses performed on the aggregate interview ratings were repeated at the
individual interviewer level for exploratory purposes. It should be noted that the statistical power of the analyses conducted at the individual interviewer level is low due to the small number of applicants available for each interviewer (ns ranged from 49-50 applicants). The results of these analyses are presented in Tables 3 and 4.

The uncorrected bivariate correlations between the ratings of Interviewer 1 and the first and second performance appraisal ratings and turnover behavior were .14, .05, and -.19, respectively. When the effects of the pre-interview tests were removed from the predictor, the resulting validity coefficients changed to .04, .11, and -.07, respectively. When these semi-partial correlations were corrected for the effects of range restriction and criterion unreliability, the validity coefficients of Interviewer 1's evaluations increased to .09, .24, and -.15. None of these relationships was statistically significant.

It is interesting to note that the first two validity coefficients were positive, while only the validity coefficient for the first performance appraisal rating was positive at the aggregate interviewer level. Another difference between Interviewer 1's ratings and those at the aggregate interviewer level was that Interviewer 1's ratings predicted the second performance appraisal scores better than scores on the first performance appraisal. At the aggregate level, performance on the first performance appraisal was predicted better than performance on the second performance appraisal.

For Interviewer 2, the uncorrected bivariate correlation between interview ratings, performance appraisal scores, and turnover behavior were .09, -.10, and .03, respectively. Removing the effects of the pre-interview tests from the predictor had little effect: the coefficients were
Table 3

Individual Interviewer Unpartialled Validity Coefficients

<table>
<thead>
<tr>
<th>Predictor</th>
<th>n</th>
<th>Appraisal 1</th>
<th>Appraisal 2</th>
<th>Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer 1</td>
<td>50</td>
<td>.14</td>
<td>.05</td>
<td>-.19</td>
</tr>
<tr>
<td>Interviewer 2</td>
<td>49-50</td>
<td>.09</td>
<td>-.10</td>
<td>.03</td>
</tr>
<tr>
<td>Interviewer 3</td>
<td>46-49</td>
<td>-.15</td>
<td>-.15</td>
<td>.06</td>
</tr>
</tbody>
</table>
**Table 4**

**Individual Interviewer Corrected Semi-partial Validity Coefficients**

<table>
<thead>
<tr>
<th>Criteria</th>
<th>n</th>
<th>Appraisal 1</th>
<th>Appraisal 2</th>
<th>Turnover</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interviewer 1</td>
<td>50</td>
<td>.04</td>
<td>.11</td>
<td>-.07</td>
</tr>
<tr>
<td>Corrected 1</td>
<td>50</td>
<td>.09(^a)(.05)^ab</td>
<td>.24(^a)(.14)^ab</td>
<td>-.15(^a)(-</td>
</tr>
<tr>
<td>Interviewer 2</td>
<td>49-50</td>
<td>.08</td>
<td>-.10</td>
<td>.04</td>
</tr>
<tr>
<td>Corrected 2</td>
<td>49-50</td>
<td>.17(^a)(.10)^ab</td>
<td>-.22(^a)(-.13)^ab</td>
<td>.09(^a)(.05)^a</td>
</tr>
<tr>
<td>Interviewer 3</td>
<td>46-49</td>
<td>-.05</td>
<td>-.12</td>
<td>-.03</td>
</tr>
<tr>
<td>Corrected 3</td>
<td>46-49</td>
<td>-.10(^a)(-.06)^ab</td>
<td>-.25(^a)(-.15)^ab</td>
<td>-.07(^a)(-.04)^ab</td>
</tr>
</tbody>
</table>

**Note.** Test scores have been partialled from the predictor.

\(^a\)Correlations have been corrected for the effects of criterion unreliability.

\(^b\)Correlations have not been corrected for the effects of range restriction in the predictor.
.08, -.10, and .04. However, when Interviewer 2's validity coefficients were corrected for the effects of range restriction and criterion unreliability, their magnitude increased to .17, -.22, and .09, respectively. Unfortunately, none of these correlations was statistically significant.

Interviewer 3's ratings correlated with performance appraisal ratings and turnover behavior at -.15, -.15, and .06, respectively. When the effects of the test scores were removed from the predictor, the magnitude of these validity coefficients decreased to -.05, -.12, and -.03, respectively. When Interviewer 3's validity coefficients were corrected for the effects of range restriction and criterion unreliability, they increased to -.10, -.25, and -.07, respectively. Again, none of these correlations is statistically significant.

For both Interviewers 2 and 3, the predictions of job performance became worse the longer the incumbent remained on the job. Furthermore, both these interviewers' ratings correlated negatively with the scores on the second performance appraisal, as is the case at the aggregate interviewer level.

Previous analyses revealed that the proportion of variance in the criteria accounted for by the predictor did not significantly differ according to interviewer. However, examination of the results of analyses conducted at the individual interviewer level reveals that the validity coefficients for individual interviewers vary considerably in both magnitude and sign. This provides support for the idea that interview processes should be examined at the individual interviewer level. In an effort to understand these differences, the questioning behavior of the interviewers and the diagnosticity of applicant information for each interviewer was compared.
Individual Use of Interview Questions and Elicitation of Diagnostic Information

Both interview questioning process and question content were examined for each interviewer. Questioning process refers to the manner in which the interviewers posed their questions. In this study, four process characteristics were examined for each question: directed, open-ended, situational, and standardized. All other question variables in this study assessed whether or not the questions inquire about the following job-related dimensions: attitude/motivation, learning ability, physical ability, emotional stability/stress tolerance, interpersonal skill, judgment/common sense, responsibility, and personal integrity. Each interview was coded for the presence of these characteristics. The total number of occurrences of each question characteristic per interview was calculated. Then, for each question variable, the total number of occurrences was divided by the number of questions asked in the interview. The result of this calculation represents the proportion of questions asked during each interview that contain each characteristic. This figure was summed across interviews. The means and standard deviations of the average proportion of questions containing each question characteristic across interviews is presented in Table 5.

Similarly, coders indicated the extent to which each applicant response contained diagnostic information. The total score for each diagnosticity variable was calculated for each interview. Then, the eight values representing interview judgment-specific diagnosticity were summed, and both resulting diagnosticity measures were divided by the number of responses in the interview. These calculations represent the average diagnosticity of the interviewee responses in each interview. These
<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Process variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Directed</td>
<td>.14</td>
<td>.09</td>
</tr>
<tr>
<td>Open-ended</td>
<td>.33</td>
<td>.11</td>
</tr>
<tr>
<td>Secondary</td>
<td>.59</td>
<td>.13</td>
</tr>
<tr>
<td>Situational</td>
<td>.17</td>
<td>.12</td>
</tr>
<tr>
<td>Standardized</td>
<td>.49</td>
<td>.12</td>
</tr>
<tr>
<td><strong>Content variables:</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Attitude/motivation</td>
<td>.22</td>
<td>.09</td>
</tr>
<tr>
<td>Learning ability</td>
<td>.03</td>
<td>.04</td>
</tr>
<tr>
<td>Physical ability</td>
<td>.09</td>
<td>.08</td>
</tr>
<tr>
<td>Emotional stability/stress tolerance</td>
<td>.31</td>
<td>.12</td>
</tr>
<tr>
<td>Interpersonal skill</td>
<td>.24</td>
<td>.12</td>
</tr>
<tr>
<td>Judgment/common sense</td>
<td>.37</td>
<td>.13</td>
</tr>
<tr>
<td>Responsibility</td>
<td>.30</td>
<td>.13</td>
</tr>
<tr>
<td>Personal Integrity</td>
<td>.18</td>
<td>.09</td>
</tr>
</tbody>
</table>

**Note.** Scale = 0 to 1.
averages were summed across interviews. The means and standard deviations of the average diagnosticity of responses across interviews for each interviewer are presented in Table 6.

To determine whether there were differences in the questioning behavior among interviewers, ANOVAs were performed on the mean proportion of each type of question characteristic present in each interviewer's interviews. The results of these analyses are presented in Table 7.

Examination of this table reveals that there were significant differences among interviewers in their use of several question characteristics. To further investigate these differences, t-tests were performed on the use of these question characteristics among all pairs of interviewers. The results of this analysis are presented in Table 8.

Examination of this table reveals that Interviewers 1 and 2 were very similar in their questioning style. However, they significantly differed in their use of situational questions (t = 3.50, p < .01). Interviewer 1 asked a significantly lower proportion of situational questions (M = .14, SD = .09) than did Interviewer 2 (M = .10, SD = .05). Also, interviewers 1 and 2 differed in their use of questions inquiring about learning ability (t = 1.82, p < .05). Interviewer 2 asked a significantly greater proportion of these questions (M = .05, SD = .05) than did Interviewer 1 (M = .03, SD = .03). Otherwise, the differences in questioning behavior between these interviewers were not significantly different.

Interviewers 2 and 3 differed in both questioning process and question content. The tests for use of directed (t = 2.83, p < .01) and situational (t = 8.03, p < .01) questions were both significant. Interviewer 2 asked a lower proportion of situational questions (M = .10, SD = .05)
Table 6
Mean Proportion and Standard Deviations for Diagnosticity Variables
(N=149)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Scale</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical Diagnosticity</td>
<td>1 to 5</td>
<td>2.19</td>
<td>.30</td>
</tr>
<tr>
<td>Interview-Judgment Specific Diagnosticity</td>
<td>8 to 40</td>
<td>9.99</td>
<td>.60</td>
</tr>
</tbody>
</table>
Table 7

Test of Differences in Interviewer Questioning Behavior

<table>
<thead>
<tr>
<th>Variable</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directed</td>
<td>28.10</td>
<td>.00</td>
</tr>
<tr>
<td>Open-ended</td>
<td>49.80</td>
<td>.00</td>
</tr>
<tr>
<td>Secondary</td>
<td>37.16</td>
<td>.00</td>
</tr>
<tr>
<td>Situational</td>
<td>40.37</td>
<td>.00</td>
</tr>
<tr>
<td>Attitude/motivation</td>
<td>1.98</td>
<td>.14</td>
</tr>
<tr>
<td>Learning ability</td>
<td>16.73</td>
<td>.00</td>
</tr>
<tr>
<td>Physical ability</td>
<td>114.06</td>
<td>.00</td>
</tr>
<tr>
<td>Emotional stability/stress tolerance</td>
<td>39.37</td>
<td>.00</td>
</tr>
<tr>
<td>Interpersonal skill</td>
<td>25.41</td>
<td>.00</td>
</tr>
<tr>
<td>Judgment/common sense</td>
<td>38.96</td>
<td>.00</td>
</tr>
<tr>
<td>Responsibility</td>
<td>2.39</td>
<td>.09</td>
</tr>
<tr>
<td>Personal Integrity</td>
<td>.59</td>
<td>.55</td>
</tr>
<tr>
<td>Standardized</td>
<td>4.05</td>
<td>.02</td>
</tr>
</tbody>
</table>
Table 8

Interviewer Differences in the Use of Question Type

<table>
<thead>
<tr>
<th>Variable</th>
<th>x1</th>
<th>n1</th>
<th>SD1</th>
<th>x2</th>
<th>n2</th>
<th>SD2</th>
<th>x3</th>
<th>n3</th>
<th>SD3</th>
<th>t12</th>
<th>t23</th>
<th>t13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Directed</td>
<td>.16</td>
<td>50</td>
<td>.08</td>
<td>.19</td>
<td>50</td>
<td>.09</td>
<td>.08</td>
<td>49</td>
<td>.06</td>
<td>1.55</td>
<td></td>
<td>2.83*1.80**</td>
</tr>
<tr>
<td>Open-ended</td>
<td>.29</td>
<td>50</td>
<td>.07</td>
<td>.28</td>
<td>50</td>
<td>.08</td>
<td>.43</td>
<td>49</td>
<td>.09</td>
<td>1.18</td>
<td>1.49</td>
<td>1.73</td>
</tr>
<tr>
<td>Secondary</td>
<td>.59</td>
<td>50</td>
<td>.11</td>
<td>.50</td>
<td>50</td>
<td>.10</td>
<td>.68</td>
<td>49</td>
<td>.10</td>
<td>1.10</td>
<td>1.05</td>
<td>1.08</td>
</tr>
<tr>
<td>Situational</td>
<td>.14</td>
<td>50</td>
<td>.09</td>
<td>.10</td>
<td>50</td>
<td>.05</td>
<td>.27</td>
<td>49</td>
<td>.13</td>
<td>3.50**</td>
<td>8.07**</td>
<td>2.39**</td>
</tr>
<tr>
<td>Learning ability</td>
<td>.03</td>
<td>50</td>
<td>.03</td>
<td>.05</td>
<td>50</td>
<td>.04</td>
<td>.01</td>
<td>49</td>
<td>.02</td>
<td>1.82*</td>
<td>3.54**</td>
<td>1.95*</td>
</tr>
<tr>
<td>Physical ability</td>
<td>.11</td>
<td>50</td>
<td>.05</td>
<td>.15</td>
<td>50</td>
<td>.06</td>
<td>.01</td>
<td>49</td>
<td>.03</td>
<td>1.22</td>
<td>4.49**</td>
<td>2.98**</td>
</tr>
<tr>
<td>Emotional stability/stress tolerance</td>
<td>.29</td>
<td>50</td>
<td>.09</td>
<td>.24</td>
<td>50</td>
<td>.08</td>
<td>.41</td>
<td>49</td>
<td>.12</td>
<td>1.28</td>
<td>2.08*</td>
<td>1.72</td>
</tr>
<tr>
<td>Interpersonal skill</td>
<td>.19</td>
<td>50</td>
<td>.09</td>
<td>.20</td>
<td>50</td>
<td>.08</td>
<td>.32</td>
<td>49</td>
<td>.13</td>
<td>1.14</td>
<td>2.52**</td>
<td>2.27**</td>
</tr>
<tr>
<td>Judgment/common sense</td>
<td>.36</td>
<td>50</td>
<td>.09</td>
<td>.28</td>
<td>50</td>
<td>.08</td>
<td>.47</td>
<td>49</td>
<td>.14</td>
<td>1.50</td>
<td>2.85**</td>
<td>2.13*</td>
</tr>
<tr>
<td>Standardized</td>
<td>.45</td>
<td>50</td>
<td>.11</td>
<td>.51</td>
<td>50</td>
<td>.10</td>
<td>.51</td>
<td>49</td>
<td>.13</td>
<td>1.22</td>
<td>1.68</td>
<td>1.37</td>
</tr>
</tbody>
</table>

Note: Numbers represent average proportion of questions per interview containing each characteristic.

Scale = 0 to 1

*p < .05.  **p < .01
than did Interviewer 3 ($M = .27$, $SD = .13$). Yet he asked a greater proportion of directed questions ($M = .19$, $SD = .09$) than did Interviewer 3 ($M = .08$, $SD = .06$). Moreover, Interviewers 2 and 3 differed in their use of questions regarding learning ability ($t = 3.54$, $p < .01$), physical ability ($t = 4.49$, $p < .01$), emotional stability/stress tolerance ($t = 2.08$, $p < .01$), interpersonal skill ($t = 2.52$, $p < .01$), and judgment/common sense ($t = 2.85$, $p < .05$). Interviewer 2 asked a greater proportion of questions regarding learning ability ($M = .05$, $SD = .04$) and physical ability ($M = .15$, $SD = .06$) than did Interviewer 3 ($M = .01$, $SD = .02$; $M = .01$, $SD = .03$, respectively). However, Interviewer 2 also asked a smaller proportion of questions regarding emotional stability/stress tolerance ($M = .24$, $SD = .08$) and judgment/common sense ($M = .28$, $SD = .08$) than did Interviewer 3 ($M = .41$, $SD = .12$; $M = .47$, $SD = .14$, respectively).

Interviewers 1 and 3 differed on almost the same question variables as did Interviewers 2 and 3. Also, tests of the significance of differences in the use of directed ($t = 1.80$, $p < .05$) and situational ($t = 2.39$, $p < .01$) questions were both significant. Interviewer 1 asked a significantly greater proportion of directed questions ($M = .16$, $SD = .08$) than did Interviewer 3 ($M = .08$, $SD = .06$). But, Interviewer 1 asked a significantly smaller proportion of situational questions ($M = .14$, $SD = .09$) than did Interviewer 3 ($M = .27$, $SD = .13$). Tests of differences in use of questions regarding learning ability ($t = 1.95$, $p < .05$), physical ability ($t = 2.98$, $p < .01$), interpersonal skill ($t = 2.27$, $p < .01$), and judgment/common sense ($t = 2.13$, $p < .05$) were also statistically significant. Interviewer 1 asked a significantly greater proportion of questions regarding learning ability ($M = .03$, $SD = .03$) and physical ability ($M = .11$, $SD = .05$) than did Interviewer 3 ($M = .01$, $SD = .02$; $M = .01$, $SD = .03$, respectively).
Interviewer 1 asked a significantly smaller proportion of questions regarding emotional stability/stress tolerance ($M = .29$, $SD = .09$), interpersonal skill ($M = .19$, $SD = .09$), and judgment/common sense ($M = .36$, $SD = .09$) than did Interviewer 3 ($M = .41$, $SD = .12$; $M = .32$, $SD = .13$; $M = 47$, $SD = .14$, respectively).

These results indicate that more of the differences among interviewers' questioning behavior lie in question content as opposed to questioning process. Furthermore, there were many differences in questioning behavior between Interviewers 2 and 3 and Interviewers 1 and 3, but relatively few differences between the questioning behavior of Interviewers 1 and 2. This pattern is somewhat surprising since the ratings made by Interviewers 2 and 3 correlated negatively with applicants' second performance appraisal ratings, whereas interview ratings made by Interviewer 1 correlated positively with this criterion. However, Interviewer 3 is the only interviewer whose ratings correlated negatively with the first performance appraisal rating. This pattern of results seems to suggest that differences in questioning style primarily affect information relevant to performance immediately following hire.

ANOVAs were also performed on the mean proportion of each type of information diagnosticity elicited in each interviewer's interviews. These results are presented in Table 9. They indicate that interviewers differ significantly in both the typical diagnosticity ($F = 19.26$, $p < .01$) and the interview-judgment specific diagnosticity ($F = 18.92$, $p < .01$) of the information that is elicited in their interviews.

T-tests were performed to investigate these differences, the results of which are presented in Table 10. Analyses revealed that neither the typical diagnosticity ($t = 1.65$, n.s.) nor the interview-judgment specific
Table 9

Test of Differences in Interviewer Information Diagnosticity

<table>
<thead>
<tr>
<th>Variable</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical diagnosticity</td>
<td>19.26</td>
<td>.00</td>
</tr>
<tr>
<td>Interview judgment-specific diagnosticity</td>
<td>18.92</td>
<td>.00</td>
</tr>
</tbody>
</table>
Table 10

**Interviewer Differences in Diagnosticity of Information Elicited From Applicants**

<table>
<thead>
<tr>
<th>Variable</th>
<th>x1</th>
<th>n1</th>
<th>SD1</th>
<th>x2</th>
<th>n2</th>
<th>SD2</th>
<th>x3</th>
<th>n3</th>
<th>SD3</th>
<th>t12</th>
<th>t23</th>
<th>t13</th>
</tr>
</thead>
<tbody>
<tr>
<td>Typical Diagnosticitya</td>
<td>2.05</td>
<td>50</td>
<td>.24</td>
<td>2.16</td>
<td>50</td>
<td>.19</td>
<td>2.38</td>
<td>49</td>
<td>.36</td>
<td>1.65</td>
<td>3.79**</td>
<td>2.26**</td>
</tr>
<tr>
<td>Interview-Judgmentb</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specific Diagnosticity</td>
<td>9.86</td>
<td>50</td>
<td>.48</td>
<td>9.75</td>
<td>50</td>
<td>.39</td>
<td>10.37</td>
<td>49</td>
<td>.71</td>
<td>1.5</td>
<td>2.22**</td>
<td>3.36**</td>
</tr>
</tbody>
</table>

*Note.* Numbers represent average diagnosticity per interviewee response.

aScale range = 1-5.
bScale range = 8-40

* p < .05. ** p < .01.
diagnosticity (t = 1.5, n.s.) of applicant information elicited by
Interviewers 1 and 2 differed significantly. Interviewers 2 and 3,
however, differed in both the typical diagnosticity (t = 3.79, p < .01) and
interview-judgment specific diagnosticity (t = 2.22, p < .01) of
information gathered from their applicants. Interviewer 2 elicited less
typically diagnostic information from his applicants (M = 2.16, SD = .19)
than did Interviewer 3 (M = 2.38, SD = .36). The information that
Interviewer 1 gathered from applicants was also lower in interview-
judgment specific diagnosticity (M = 9.75, SD = .39) than that gathered by
Interviewer 3 (M = 10.37, SD = .71).

Similarly, the applicant information that Interviewers 1 and 3
gathered significantly differed in both types of diagnosticity (typical
diagnosticity: t = 2.26, p < .01; interview judgment-specific diagnosticity:
t = 3.36, p < .01). The information that Interviewer 1 elicited from
applicants was lower in both typical diagnosticity (M = 2.05, SD = .24) and
interview-judgment specific diagnosticity (M = 9.86, SD = .48) than that
gathered by Interviewer 3 (M = 2.38, SD = .36; M = 10.37, SD = .71,
respectively).

This pattern of results is similar to that resulting from examining
individual interviewer questioning behavior. There were significant
differences in the diagnosticity of information elicited by Interviewers 2
and 3 and Interviewers 1 and 3, but not between that of Interviewers 1 and
2. These results suggest that there may be a relationship between
questioning behavior and interviewee information diagnosticity.

Examination of the results of all analyses conducted at the individual
interviewer level suggest that Interviewer 3 is the most "deviant"
interviewer. His questioning style differed most from that of the other two
interviewers. Furthermore, the information he elicited from interviewees was higher in diagnosticity than the information gathered by the other two interviewers. The differences among the quality of this interviewer's applicant ratings and those of the other two interviewers are most visible when examining the correlation between applicant ratings and the first performance ratings that employees receive. While Interviewer 3's questioning style seems to elicit information higher in diagnosticity than the questioning behavior of the other interviewers, it did not lead to statistically valid interview ratings.

Interviewer 2 behaved similarly to Interviewer 1 in applicant questioning. The diagnosticity of his applicants' responses was similar to that of Interviewer 1's applicants. Yet, the interview scores produced by Interviewer 2 negatively related to the applicants' second performance appraisal ratings, while interview ratings produced by Interviewer 1 always positively correlated with the performance scores. This suggests that the differences in questioning style that do exist between Interviewers 1 and 2 (the proportion of situational questions and the proportion of questions regarding applicant learning ability) are critical influences on the differences in validity between these interviewers' applicant evaluations. Additional tests of the information gathering model were performed to investigate this proposition.

Data Reduction

The correlations among the question variables (absolute value $r = .29, p < .01$; range = .01 to .82) and those among the diagnosticity variables (absolute value $r = .26, p < .01$; range = .00 to .74) raised concern regarding multicollinearity effects in tests of the information gathering model. To prevent this problem, factor analyses were
performed on the sets of question and diagnosticity variables. The information gathering model was tested using the resulting factor scores.

The first step in the factor analysis of the question variables was to separate them with respect to process and content. That is, variables reflecting question subject were separated from those representing question form. A principal components analysis (using varimax rotation) was applied to each group of variables. Variables loading at .61 or higher on a factor were considered representative of that factor. The loading of each question variable on the emergent factors are presented in Table 11.

The factor analysis produced two process factors. Variables loading highly on this first factor were open-ended and situational question characteristics. Since situational questions are almost always open-ended in nature, this factor was termed the "openness" factor. The secondary variable loaded highly but negatively on the second factor, and the standardized variable loaded highly and positively on this factor. Because interviewers have less discretion in probing applicants as interviews become more standardized, this factor was termed the "standardized" factor.

In the analysis of the question content variables, one variable (physical ability) was complex (i.e., it loaded highly on more than one variable) and therefore was removed from the analysis as suggested by Tabachnik and Fidell (1989). A second variable ("other") was removed from the analysis because its inclusion rendered the resulting factor structure uninterpretable.

Three content factors emerged from the analysis of the remaining variables. The first was comprised by the emotional stability/stress tolerance, interpersonal skill, and judgment/common sense variables. Since
Table 11

Rotated Factor Matrix of Question Variables (varimax rotation)

**Process variables**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Openness</th>
<th>Standardization</th>
</tr>
</thead>
<tbody>
<tr>
<td>Open-ended</td>
<td>.8781</td>
<td>.0110</td>
</tr>
<tr>
<td>Secondary</td>
<td>.4213</td>
<td>-.8012</td>
</tr>
<tr>
<td>Situational</td>
<td>.9051</td>
<td>-.0415</td>
</tr>
<tr>
<td>Standardized</td>
<td>.2580</td>
<td>.8873</td>
</tr>
</tbody>
</table>

**Content variables**

<table>
<thead>
<tr>
<th>Factor</th>
<th>Job situation</th>
<th>Conformity</th>
<th>Performance potential</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude/motivation</td>
<td>.2712</td>
<td>.3694</td>
<td>.7111</td>
</tr>
<tr>
<td>Learning ability</td>
<td>-.3731</td>
<td>.2277</td>
<td>.6103</td>
</tr>
<tr>
<td>Emotional stability/stress tolerance</td>
<td>.9154</td>
<td>-.2121</td>
<td>-.0780</td>
</tr>
<tr>
<td>Interpersonal skill</td>
<td>.8367</td>
<td>-.3385</td>
<td>.1609</td>
</tr>
<tr>
<td>Judgment/common sense</td>
<td>.8440</td>
<td>.3258</td>
<td>-.1239</td>
</tr>
<tr>
<td>Responsibility</td>
<td>.0095</td>
<td>.9100</td>
<td>-.0461</td>
</tr>
<tr>
<td>Personal integrity</td>
<td>.2181</td>
<td>.8184</td>
<td>-.3034</td>
</tr>
</tbody>
</table>
interviewers were instructed to use answers to situational questions in rating applicants on these interview dimensions, this factor was termed the "job situation" factor. The second factor was represented by the responsibility and personal integrity variables. These variables appear on the conformity scale of the California Personality Inventory (CPI; Gough, 1987), so this factor was termed the "conformity" factor. The final content factor was comprised by the attitude/motivation and learning ability question characteristics. Since these characteristics represent basic, but possibly undeveloped attributes, this was termed the "performance potential" factor.

Diagnosticity variables were subjected to a similar analysis. They were separated according to Hilton and Fein’s (1987) division regarding typical and judgment-specific diagnosticity. Only the latter category was comprised by multiple variables. A principal components analysis with varimax rotation was applied to these variables. Those loading at .45 or higher on a factor were considered representative of that factor. The loading of each diagnosticity variable on the emergent factors is presented in Table 12. This analysis revealed three interview-judgment specific diagnosticity factors. Because the information characteristics that loaded highly on the first two factors were very similar to those loading highly on the first two question content factors, they were also named "job situation" and "conformity". Their learning ability and physical ability variables loaded on the final factor. Since this factor was represented by variables measuring basic abilities, it was termed "ability".

Hierarchical Regression Tests of the Information Gathering Model

The interview information gathering model was tested with hierarchical regression in a manner suggested by James and Brett (1984).
Table 12

Rotated Factor Matrix of Interview Judgment-specific Diagnosticity

<table>
<thead>
<tr>
<th>Variables</th>
<th>Job situation</th>
<th>Conformity</th>
<th>Ability</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attitude/motivation</td>
<td>.2309</td>
<td>.4589</td>
<td>-.2564</td>
</tr>
<tr>
<td>Learning ability</td>
<td>.1026</td>
<td>.1590</td>
<td>.8591</td>
</tr>
<tr>
<td>Physical ability</td>
<td>-.3244</td>
<td>-.1818</td>
<td>.7829</td>
</tr>
<tr>
<td>Emotional stability/stress tolerance</td>
<td>.8743</td>
<td>.0440</td>
<td>-.1375</td>
</tr>
<tr>
<td>Interpersonal skill</td>
<td>.8935</td>
<td>.0256</td>
<td>-.0588</td>
</tr>
<tr>
<td>Judgment/common sense</td>
<td>.7230</td>
<td>.2966</td>
<td>-.0119</td>
</tr>
<tr>
<td>Responsibility</td>
<td>-.0534</td>
<td>.8788</td>
<td>.1584</td>
</tr>
<tr>
<td>Personal integrity</td>
<td>.1723</td>
<td>.8089</td>
<td>-.0084</td>
</tr>
</tbody>
</table>
Because the model involves the validity of interviewers’ ratings independent of the test score information, test scores were covaried from the examinations of the model. The information gathering model predicts that question factors will explain variance in interviewer decision accuracy after test scores have been entered into the regression equation. It also predicts that when diagnosticity factors are added, they will account for additional variance in interviewer accuracy. However, when diagnosticity factors precede question factors in the regression equation, the diagnosticity factors, but not the question factors, should explain a significant amount of variance in interviewer accuracy. This is because the information gathering model posits that variance in interviewer accuracy is explained by question factors primarily through information diagnosticity. The results of this analysis are presented in Table 13. In this table, MIN Change in $R^2$ refers to the proportion of variance in evaluation accuracy that is accounted for by the variable set when it is entered last into the regression equation. MAX Change in $R^2$ refers to the proportion of variance in evaluation accuracy that is accounted for by the variable set when it is entered first into the regression equation.

Accuracy of interviewer rating was regressed on test scores at the first step ($R^2 = .01$, n.s.), question factors on the second step ($R^2_{change} = .03$, n.s.), and then diagnosticity factors on the third step ($R^2_{change} = .03$, n.s.). Together, these variables accounted for only 6% of the variance in decision accuracy (n.s.). Next, accuracy was regressed on test scores at the first step ($R^2 = .01$, n.s.), diagnosticity factors on the second step ($R^2_{change} = .02$, n.s.), and then question factors on the third step ($R^2_{change} = .03$, n.s.). Neither pre-interview tests, question factors, nor the diagnosticity factors significantly predicted interviewer accuracy at the
Table 13
Hierarchical Regression of Accuracy on Diagnosticity and Questioning Style

Aggregate interviewer level (N=149)
Whole model: F (11, 138) = .95, R^2 = .06, n.s.

<table>
<thead>
<tr>
<th>Step/source</th>
<th>R</th>
<th>R^2</th>
<th>MIN Change R^2</th>
<th>MAX Change R^2</th>
<th>b</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test Scores:</td>
<td>.08</td>
<td>.01</td>
<td>.01</td>
<td>.01</td>
<td>.004</td>
<td>.85</td>
</tr>
<tr>
<td>2. Diagnosticity</td>
<td>.03</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical diagnosticity</td>
<td>.99</td>
<td>2.33</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Situation</td>
<td>-.14</td>
<td>.38</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity</td>
<td>.02</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability</td>
<td>.03</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Questioning Style:</td>
<td>.04</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td>-.45</td>
<td>2.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardization</td>
<td>-.26</td>
<td>3.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Situation</td>
<td>.32</td>
<td>1.5</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity</td>
<td>-.07</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance potential</td>
<td>-.05</td>
<td>.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Change MIN R^2 = Increment in variance accounted for when predictor set is entered last into the regression equation.

Change MAX R^2 = Increment in variance accounted for when predictor set is entered first into the regression equation.

R^2 = Variance accounted for by the predictor set upon entry into the regression equation.

b = regression weight after all variables have been entered in the equation.
aggregate interviewer level.

There is no straightforward method of testing for differences in individual interviewers' adherence to the information gathering model. Because interviewers had demonstrated such variability in predictive validity, the model was tested at the individual interviewer level, despite the low statistical power of such a test. The results are presented in Tables 14-16.

For Interviewer 1, evaluation accuracy was regressed on test scores at the first step ($R^2 = .04$, n.s.), the question factors on the second step ($R^2_{\text{change}} = .01$, n.s.), and diagnosticity factors on the third step ($R^2_{\text{change}} = .09$, n.s.). Collectively, they accounted for 14% of the variance in evaluation accuracy (n.s.). Then, accuracy was regressed on test scores at the first step ($R^2 = .04$, n.s.), the diagnosticity factors on the second step ($R^2_{\text{change}} = .09$, n.s.), and question factors on the third step ($R^2_{\text{change}} = .02$, n.s.). None of these variables significantly predicted evaluation accuracy.

For Interviewer 2, evaluation accuracy was regressed on test scores at the first step ($R^2 = .00$, n.s.), the question factors on the second step ($R^2_{\text{change}} = .29$, $p < .05$), and diagnosticity factors on the third step ($R^2_{\text{change}} = .03$, n.s.). Combined, these factors accounted for 32% of the variance in evaluation accuracy. The question factors predicted a significant amount of variance in evaluation accuracy. Then, accuracy was regressed on test scores at the first step ($R^2 = .00$, n.s.), the diagnosticity factors on the second step ($R^2_{\text{change}} = .09$), and question factors on the third step ($R^2_{\text{change}} = .23$, $p < .05$). Contrary to the model predictions, question factors accounted for a significant amount of variance in
Table 14

Hierarchical Regression of Accuracy on Diagnosticity and Questioning Style

Interviewer 1 (n=50)

Whole model: F (11, 39) = .38, R² = .14, n.s.

<table>
<thead>
<tr>
<th>Step/source</th>
<th>MIN R²</th>
<th>MIN Change R²</th>
<th>MAX R²</th>
<th>MAX Change R²</th>
<th>b</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test Scores:</td>
<td>.19</td>
<td>.04</td>
<td>.04</td>
<td>.04</td>
<td>.01</td>
<td>.93</td>
</tr>
<tr>
<td>2. Diagnosticity:</td>
<td>.09</td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical diagnosticity</td>
<td></td>
<td></td>
<td>1.67</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Situation</td>
<td></td>
<td></td>
<td>.22</td>
<td>.32</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity</td>
<td></td>
<td></td>
<td>.49</td>
<td>.63</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability</td>
<td></td>
<td></td>
<td>-.41</td>
<td>.81</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Questioning Style:</td>
<td>.02</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td></td>
<td></td>
<td>.22</td>
<td>.12</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardization</td>
<td></td>
<td></td>
<td>-.04</td>
<td>.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Situation</td>
<td></td>
<td></td>
<td>-.30</td>
<td>.21</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity</td>
<td></td>
<td></td>
<td>-.23</td>
<td>.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Potential</td>
<td></td>
<td></td>
<td>-.08</td>
<td>.04</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Change MIN R² = Increment in variance accounted for when predictor set is entered last into the regression equation.

Change MAX R² = Increment in variance accounted for when predictor set is entered first into the regression equation.

R² = Variance accounted for by the predictor set upon entry into the regression equation.

b = regression weight after all variables have been entered in the equation.
Table 15

Hierarchical Regression of Accuracy on Diagnosticity and Questioning Style

Interviewer 2 (n=50)

Whole model: F (11, 39) = 1.84, R² = .32, (n.s.)

<table>
<thead>
<tr>
<th>Step/source</th>
<th>MIN R</th>
<th>MIN R²</th>
<th>MIN Change R²</th>
<th>MAX R²</th>
<th>MAX Change R²</th>
<th>b</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test Scores:</td>
<td>.03</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2. Diagnosticity:</td>
<td></td>
<td>.03</td>
<td></td>
<td>.09</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical diagnosticity</td>
<td></td>
<td></td>
<td></td>
<td>.04</td>
<td>.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Situation</td>
<td></td>
<td></td>
<td></td>
<td>-.21</td>
<td>.14</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity</td>
<td></td>
<td></td>
<td></td>
<td>-.17</td>
<td>.29</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability</td>
<td></td>
<td></td>
<td></td>
<td>.39</td>
<td>1.24</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Questioning Style:</td>
<td>.23*</td>
<td></td>
<td>.29**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td></td>
<td>-.96</td>
<td>3.72</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardization</td>
<td></td>
<td>-.22</td>
<td>.56</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Situation</td>
<td></td>
<td>.63</td>
<td>1.09</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity</td>
<td></td>
<td>.76</td>
<td>6.71**</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Potential</td>
<td></td>
<td>.17</td>
<td>.47</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note. Change MIN R² = Increment in variance accounted for when predictor set is entered last into the regression equation.

Change MAX R² = Increment in variance accounted for when predictor set is entered first into the regression equation.

R² = Variance accounted for by the predictor set upon entry into the regression equation.

b = regression weight when all variables are entered into the equation.

**p < .01. *p < .05
### Table 16

**Hierarchical Regression of Accuracy on Diagnosticity and Questioning Style**

Interviewer 3 (n=49)

Whole model: F (11, 38) = .74, $R^2 = .16$, n.s.

<table>
<thead>
<tr>
<th>Step/source</th>
<th>MIN R</th>
<th>MIN R²</th>
<th>MAX R²</th>
<th>MAX R²</th>
<th>b</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <strong>Test Scores:</strong></td>
<td>.03</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
<td>.00</td>
</tr>
<tr>
<td>2. <strong>Diagnosticity:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Typical diagnosticity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Situation</td>
<td>.05</td>
<td>.03</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ability</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. <strong>Questioning Style:</strong></td>
<td>.13</td>
<td>.11</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Openness</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Standardization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Job Situation</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Conformity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Performance Potential</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.17</td>
<td>.13</td>
</tr>
</tbody>
</table>

**Note.** Change MIN $R^2 = $ Increment in variance accounted for when predictor set is entered last into the regression equation.

Change MAX $R^2 = $ Increment in variance accounted for when predictor set is entered first into the regression equation.

$R^2 = $ Variance accounted for by the predictor set upon entry into the regression equation.

$b = $ regression weight when all variables have been entered into the equation.
evaluation accuracy even after information diagnosticity had been entered into the equation.

For Interviewer 2, the regression weight for the question conformity factor ($F = 9.1, p < .01$) was significant. The more this interviewer used questions loading on the conformity factor (questions regarding responsibility and personal integrity), the more accurate his applicant evaluations.

For Interviewer 3, evaluation accuracy was regressed on test scores at the first step ($R^2 = .00$, n.s.), the question factors on the second step ($R^2_{\text{change}} = .11$, n.s.), and diagnosticity factors on the third step ($R^2_{\text{change}} = .05$, n.s.). Together, the independent variables accounted for 16% of the variance in evaluation accuracy (n.s.). Then, accuracy was regressed on test scores at the first step ($R^2 = .00$, n.s.), the diagnosticity factors on the second step ($R^2_{\text{change}} = .03$, n.s.), and question factors on the third step ($R^2_{\text{change}} = .13$, n.s.). None of these variables significantly predicted evaluation accuracy.

In an effort to understand these results, the validity of the true score estimates was then investigated. For neither the first nor second performance appraisal ratings were the uncorrected bivariate ($r = .08, r = .02$) or uncorrected semi-partial validity coefficients ($r = .14, r = .05$) significant. When corrected for range restriction and criterion unreliability, the semi-partial correlation between the true score estimate and the first ($r = .32, p < .01$), but not the second ($r = .10$, n.s.) performance appraisal score was statistically significant. However, as previously discussed, the supervisors at the target organization had commented that the incumbents' first performance appraisal score was not meaningful. Since the true score estimates were not adequately predicting
a useful measure of employee performance, one may question the appropriateness of measuring interviewers’ deviation from the true score estimates in this study.

**Model Testing With Correlational Tests**

To further investigate the lack of significant results at the aggregate interviewer level and to tease out differences in information gathering among individual interviewers, relationships among model components were tested with set correlations. At the aggregate level, both the relationship between interview diagnosticity factors and accuracy and the relationship between question factors and evaluation accuracy ($R^2 = .02$, n.s.; $R^2 = .04$, n.s., respectively) were minimal. However, the relationship between interview diagnosticity factors and interview question factors was extremely high ($R^2 = .81$, $p < .01$).

Tests were then performed to determine if relationships among model components differed at the individual interviewer level. There is no straightforward method of testing for differences in the relationship between the question and diagnosticity factors among interviewers. However, regression analyses were performed to find if relationships between question factors and accuracy or relationships between diagnosticity factors and accuracy differed at the individual interviewer level. The results of these analyses are presented in Tables 17 and 18.

The coded interviewer vectors, question factors, and the interaction between question factors and interviewer vectors accounted for 12% of the variance in evaluation accuracy (n.s.). The regression weight for the interaction between one of the interviewer vectors and the second question content factor (conformity) was significant ($b = .97$, $F = 6.1$, $p < .01$), thus indicating that the relationship between this factor and accuracy was not the
Table 17

Regression of Accuracy on Questioning Factors

Aggregate interviewer level (N = 149)

Whole model: F (18, 131) = 1.05, $R^2 = .12$, n.s.

<table>
<thead>
<tr>
<th>Variable</th>
<th>b</th>
<th>F</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Intercept</td>
<td>-4.34</td>
<td>110.01</td>
<td>.00</td>
</tr>
<tr>
<td>2. Openness</td>
<td>-.83</td>
<td>2.83</td>
<td>.09</td>
</tr>
<tr>
<td>3. Standardization</td>
<td>-.41</td>
<td>2.59</td>
<td>.11</td>
</tr>
<tr>
<td>4. Job Situation</td>
<td>.65</td>
<td>2.25</td>
<td>.14</td>
</tr>
<tr>
<td>5. Conformity</td>
<td>-.32</td>
<td>1.76</td>
<td>.19</td>
</tr>
<tr>
<td>6. Performance Potential</td>
<td>-.13</td>
<td>.25</td>
<td>.62</td>
</tr>
<tr>
<td>7. Interviewer vector 1</td>
<td>.09</td>
<td>.04</td>
<td>.85</td>
</tr>
<tr>
<td>8. Interviewer vector 2</td>
<td>-.09</td>
<td>.02</td>
<td>.89</td>
</tr>
<tr>
<td>9. 2x7</td>
<td>.79</td>
<td>1.20</td>
<td>.27</td>
</tr>
<tr>
<td>10. 3x7</td>
<td>.43</td>
<td>1.40</td>
<td>.24</td>
</tr>
<tr>
<td>11. 4x7</td>
<td>.35</td>
<td>.31</td>
<td>.58</td>
</tr>
<tr>
<td>12. 5x7</td>
<td>.12</td>
<td>.09</td>
<td>.76</td>
</tr>
<tr>
<td>13. 6x7</td>
<td>-.16</td>
<td>.17</td>
<td>.69</td>
</tr>
<tr>
<td>14. 2x8</td>
<td>-.29</td>
<td>.12</td>
<td>.73</td>
</tr>
<tr>
<td>15. 3x8</td>
<td>.07</td>
<td>.03</td>
<td>.87</td>
</tr>
<tr>
<td>16. 4x8</td>
<td>-.22</td>
<td>.10</td>
<td>.75</td>
</tr>
<tr>
<td>17. 5x8</td>
<td>.97</td>
<td>6.1</td>
<td>.01</td>
</tr>
<tr>
<td>18. 6x8</td>
<td>.50</td>
<td>.91</td>
<td>.17</td>
</tr>
</tbody>
</table>

Note. b = regression weight after all variables have been entered into the equation
Table 18

Regression of Accuracy on Diagnosticity Factors

Aggregate interviewer level (N = 149)

Whole model: F (15, 134) = .71, R² = .07, n.s.

<table>
<thead>
<tr>
<th>Variable</th>
<th>b</th>
<th>F</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Intercept</td>
<td>-5.42</td>
<td>6.96</td>
<td>.01</td>
</tr>
<tr>
<td>2. Typical Diagnosticity</td>
<td>.39</td>
<td>.18</td>
<td>.67</td>
</tr>
<tr>
<td>3. Job Situation</td>
<td>-.22</td>
<td>.65</td>
<td>.42</td>
</tr>
<tr>
<td>4. Conformity</td>
<td>.11</td>
<td>.24</td>
<td>.62</td>
</tr>
<tr>
<td>5. Ability</td>
<td>-.13</td>
<td>.19</td>
<td>.67</td>
</tr>
<tr>
<td>6. Interviewer vector 1</td>
<td>-2.65</td>
<td>.64</td>
<td>.42</td>
</tr>
<tr>
<td>7. Interviewer vector 2</td>
<td>1.86</td>
<td>.24</td>
<td>.63</td>
</tr>
<tr>
<td>8. 2x6</td>
<td>1.46</td>
<td>.89</td>
<td>.35</td>
</tr>
<tr>
<td>9. 3x6</td>
<td>.04</td>
<td>.01</td>
<td>.93</td>
</tr>
<tr>
<td>10. 4x6</td>
<td>-.09</td>
<td>.04</td>
<td>.84</td>
</tr>
<tr>
<td>11. 5x6</td>
<td>-.08</td>
<td>.04</td>
<td>.84</td>
</tr>
<tr>
<td>12. 2x7</td>
<td>-.81</td>
<td>.22</td>
<td>.64</td>
</tr>
<tr>
<td>13. 3x7</td>
<td>-.01</td>
<td>.00</td>
<td>.98</td>
</tr>
<tr>
<td>14. 4x7</td>
<td>.04</td>
<td>.01</td>
<td>.93</td>
</tr>
<tr>
<td>15. 5x7</td>
<td>.58</td>
<td>1.81</td>
<td>.18</td>
</tr>
</tbody>
</table>

*Note. b = regression weight after all variables have been entered into the equation*
same for all interviewers.

To investigate these differences, set correlations between question factors and evaluation accuracy were performed at the individual interviewer level. The results of these analyses are presented in Table 19. For Interviewer 1, this relationship was $R^2 = .04$ (n.s.); for Interviewer 2, $R^2 = .29$ (p < .01); for Interviewer 3, $R^2 = .10$ (n.s.). For Interviewers 1 and 3, questioning style had no significant effect on decision accuracy. However, use of questions loading on the conformity factor significantly predicted the accuracy of Interviewer 2's applicant ratings.

The t-tests previously performed indicated that Interviewers 1 and 3 differed from Interviewer 2 in the use of questions regarding responsibility (which loads highly on the conformity factor). Interviewer 2 asked a smaller proportion of these questions ($M = .29$) than did Interviewer 1 ($M = .33$), but almost the same proportion as did Interviewer 3 ($M = .28$). However, there was less variability in the use of this question by Interviewer 2 ($SD = .10$) than by Interviewer 3 ($SD = .17$). Perhaps these differences exert important influences on evaluation accuracy.

Similar regression analyses were performed to test for differences in the relationship between information diagnosticity factors and decision accuracy at the individual interviewer level. The coded interviewer vectors, diagnosticity factors, and the interaction between diagnosticity factors and interviewer vectors accounted for 7% of the variance in decision accuracy (n.s.). The regression weights for the interaction between interviewer and the diagnosticity factors were not significant, indicating that the relationship between these factors and decision accuracy did not differ according to interviewer.
Table 19
Set Correlations Among Model Components

<table>
<thead>
<tr>
<th>Level of analysis</th>
<th>n</th>
<th>Question and Diagnostics factors</th>
<th>Diagnostics and Accuracy</th>
<th>Question factors and Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aggregate</td>
<td>149</td>
<td>.81**</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>Interviewer 1</td>
<td>50</td>
<td>.79**</td>
<td>.08</td>
<td>.04</td>
</tr>
<tr>
<td>Interviewer 2</td>
<td>50</td>
<td>.82**</td>
<td>.09</td>
<td>.29**</td>
</tr>
<tr>
<td>Interviewer 3</td>
<td>49</td>
<td>.77**</td>
<td>.04</td>
<td>.10</td>
</tr>
</tbody>
</table>

** = p < .01
For exploratory purposes, set correlations between diagnostically factors and accuracy were performed at the individual interviewer level. For Interviewer 1, this relationship was $R^2 = .08$ (n.s.); for Interviewer 2, $R^2 = .09$ (n.s.); for Interviewer 3, $R^2 = .04$ (n.s.). Although the magnitude of the relationship between diagnostically factors and accuracy was low for each of the interviewers, Interviewer 2 had the highest relationship between diagnostically factors and evaluation accuracy. He also predicted performance scores on the first performance appraisal better than did the other two interviewers. Although these relationships are not statistically significant, this interviewer's information gathering process seems to approximate the model proposed in this study. This may account for the superior predictive validity of his applicant evaluations over that of the other two interviewers'.

**Aggregate Interviewer Relationships Between Diagnostically and Questioning Behavior**

To further investigate the strong correlation between question and diagnostically factors, intercorrelations among them were computed. They are presented in Table 20.

Examination of this table reveals that the openness and job situation factors correlate most strongly with the typical diagnostically of interviewee information ($r = .38, p < .01$; $r = .38, p < .01$, respectively). The more interviewers used questions loading on these factors, the higher the typical diagnostically of the information they gathered from their applicants. Not surprisingly, the openness factor also correlated at .62 ($p < .01$) with information regarding job situations.

It is not surprising that the correlations between the question factor regarding conformity and the information factor regarding conformity ($r$
<table>
<thead>
<tr>
<th>Factor</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1. Openness</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Standardization</td>
<td>0.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Job situation</td>
<td>.81**</td>
<td>.04</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Conformity</td>
<td>-.14</td>
<td>-.15</td>
<td>0.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Performance potential</td>
<td>-.13</td>
<td>.15</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnosticy:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Typical diagnosticity</td>
<td>.38**</td>
<td>.11</td>
<td>.36**</td>
<td>-.01</td>
<td>.05</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Job situation</td>
<td>.62**</td>
<td>.02</td>
<td>.59**</td>
<td>-.13</td>
<td>-.05</td>
<td>.48**</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Conformity</td>
<td>-.02</td>
<td>-.06</td>
<td>.05</td>
<td>.61**</td>
<td>.07</td>
<td>.41**</td>
<td>0.0</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>9. Ability</td>
<td>-.42**</td>
<td>-.07</td>
<td>-.45**</td>
<td>-.11</td>
<td>.12</td>
<td>-.12</td>
<td>0.0</td>
<td>0.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* = p < .05, ** = p < .01
=.62, p < .01), as well as the correlation between the question factor regarding job situations and the information factor regarding job situations (r = .59, p < .01) were considerable. These results support those found in tests of differences in interviewer questioning behavior and information diagnosticity: differences in questioning behavior are related to differences in the diagnosticity of interviewee information.

Contrary to other interview research recommending the use of questions loading on the openness and the job situation factors are the moderately strong, negative correlations between the openness factor and the ability information factor (r = -.42, p < .01) and the job situation question factor and the ability information factor (r = -.44, p < .01). The more questions loading on the openness and job situation factors that are used, the less diagnostic the information gained about applicant ability. We suspect that these results are largely due to the negative relationship between the openness and job potential question factors (r = -.13). The more questions loading on the openness factor that are asked, the fewer questions loading on the performance potential factor are asked. The fewer questions loading on the performance potential factor that are asked, the less diagnostic the information regarding ability gathered. Logic dictates that the more questions loading on the openness factor are used, the less diagnostic the information gathered regarding ability. Since the openness and the job situation question factors are strongly and positively related (r = .81, p < .01), similar rationale can be used to explain the negative relationship between the job situation factor and the ability information factor.

The complete interview information gathering model did not receive much support in this study. However, the consistently high correlation
between question and diagnosticity factors suggests that questioning behavior influences information diagnosticity. This correlation also provides some of the only field evaluation of questions that have been typically recommended to interviewers.

Discussion

This study made several meaningful contributions to the employment interview literature. First, it examined a model of the relationship between interviewer questioning behavior and interviewer rating accuracy, as well as a proposed mediator of this relationship—diagnosticity of interviewee information. This model was tested in an effort to reveal an additional theoretical link in the employment interview process. Second, diagnosticity in employment interviews was approached differently from other studies addressing this variable. Diagnosticity was measured directly, in terms of applicant information, rather than indirectly, through the potential of questions to elicit diagnostic information from applicants. Finally, in examining relationships among interviewer questioning behavior and evaluation accuracy, this study examined some of the prescriptions put forth in the interviewing literature regarding the best way to conduct employment interviews. Results from this study shed some light on the usefulness of these prescriptions.

In this study, the information gathering model received little support. According to the model, interviewer questioning behavior should affect interviewer rating accuracy through the diagnosticity of the applicant information that the questions elicit. Although correlational tests among the question and diagnosticity factors were statistically significant, neither of these model components were significantly related to interviewer rating accuracy. Therefore, diagnosticity does not appear to mediate the
relationship between interviewer questioning behavior and interviewer rating accuracy.

Social cognition researchers have directly measured information diagnosticity and have shown that individuals prefer to use highly diagnostic information over less diagnostic information in making judgments about others. Hilton and Fein (1987) found that subjects relied upon information high in typical diagnosticity (as opposed to stereotype information) when judging targets. Skowronski and Carlson (1989) found that more diagnostic cues influence impression formation more than does less diagnostic information. However, information diagnosticity did not significantly affect the accuracy of interviewers' applicant evaluations in this study.

The failure to obtain results similar to those discovered in the Hilton and Fein (1987) and Skowronski and Carlson (1989) research may be due to differences in the operationalization and manipulation of the diagnosticity variable. For example, in Hilton and Fein's study, diagnosticity was measured and manipulated according to the number of social judgment domains in which information was useful in categorizing a target. Then, subjects were given the opportunity to attend to specific pieces of information of precise levels of diagnosticity. In the current investigation, information diagnosticity was only measured: interviewers were free to ignore or attend to a large amount of applicant information of uncontrolled levels of diagnosticity. It is possible that the lack of exposure to information of particular levels of diagnosticity (i.e., information very high or very low in diagnosticity) in this investigation can partially account for the results.
Another factor that may contribute to the differences in the results of this study compared to those found in other studies addressing diagnosticity is differences in the experimental task. For example, in the Hilton and Fein (1987) and Skowronski and Carlston (1989) studies, subjects were presented with information about a target and were asked to make one or two judgments about each target. In the current study, the interviewers were introduced to a great amount of applicant information in a short period of time. Then, they were required to make ten judgments about each target. The complexity of this task may have obliterated any effect that information diagnosticity may have had on the quality of interviewers’ ratings of applicants in a less cognitively demanding context.

A third reason for failing to find an influence of information diagnosticity on the quality of interviewer ratings of applicants may be differences in the dependent variables studied. In the Hilton and Fein (1987) and Skowronski and Carlston (1989) studies, the impact of information diagnosticity on the level of rating that was assigned to the target was assessed. In the current study, the effect of information diagnosticity on the quality of interviewer ratings of applicants was assessed. The information gathering model predicted that ratings of applicants would become more accurate as the level of information diagnosticity increased. Because diagnostic information could be positive or negative in nature, there was no theoretical reason to expect that interviewers would rate applicants higher as the information they gathered increased in diagnosticity. Therefore, the relationship between information diagnosticity and applicant rating (examined in other research addressing diagnosticity) was of minimal interest in this study.
One issue to consider when evaluating research on diagnosticity is that this variable has generally been measured subjectively. That is, information diagnosticity (i.e., the degree to which information can help categorize targets) has been estimated by subjects or by experimenters: the actual, objective degree to which the information correlates with a particular trait or with category membership is unknown. However, a more objective operationalization of diagnosticity has been developed by Martin (1987). She compared the perceived and actual diagnosticity of various personality traits taken from the Bem Sex Role Inventory (Bem, 1974). Where discrepancies existed between the diagnostic value that a group of subjects thought was associated with a particular characteristic and its actual diagnosticity, its perceived diagnosticity was more extreme than its actual diagnostic value. This indicates that perceptions of the diagnosticity of information are not always accurate.

In this study, the fact that only the perceived diagnosticity of applicant information was measured may help explain the results. It would be interesting to compare interviewers' perceptions of the diagnosticity of the information that they gathered from applicants against its actual diagnosticity. The objective diagnosticity of applicant information would be a more meaningful measure of information diagnosticity than its perceived diagnosticity, and might assist in creating a model that accurately reflects the information gathering processes of interviewers more than does the perceived diagnosticity of the information. Unfortunately, not all interviewers were employed at the target organization at the time of interview coding, so their perceptions of information diagnosticity were inaccessible.
Another reason that the quality of applicant information did not affect the quality of interviewer ratings may be that the interviewers were collecting, but not using, diagnostic information provided by applicants in the interview. Social cognition researchers (e.g., Gilbert & Hixon, 1991; Schnieder, 1991) have found that information that is accessed or gathered is not always used. Perhaps interviewers primarily based applicant evaluations on pre-interview information or on only a small amount of the information that the interview yielded. Since interviewers in this study were in control of both the content and length of the interview, this proposition does not appear to make sense. Why would interviewers spend time gathering information that they do not intend to use? Johns' (1975) work may shed some light on this issue. Johns found that interviewers who controlled the order in which information was received used less information to make their applicant evaluations and took less time to make the evaluations compared to interviewers who received applicant information in a random order. In this study, interviewers were free to ask whatever questions they wished in whatever order they chose. Perhaps this explains why the interviewers gathering the most diagnostic information were not those making the most valid applicant ratings.

Another possibility is that the interviewers did not believe that the interviewees were answering their questions truthfully. It is generally expected that both interviewers and interviewees engage in some impression management during interviews (Dipboye, 1992). Literature advising interviewers on interviewing techniques describes how interviewers should "read into" applicant responses and how to assess their truthfulness (e.g., Fear, 1984; Half, 1985). If interviewers in this study were attempting to interpret applicant information and questioned its
veracity, information that would ordinarily be diagnostic (assuming that it was true), was useless. Therefore, the interviewers may have based their ratings on factors other than interviewee responses, such as verifiable pre-interview information.

A critical limitation in this study was the low statistical power involved. Even at the aggregate interviewer level, the number of subjects required for a 50% chance of detecting a true correlation of .09 (the highest validity coefficient calculated at the aggregate interviewer level) was not available. At the individual interviewer level, no more than 50 subjects were available for each analysis. To have a 50% chance of detecting a true correlation of .25 (the highest validity coefficient calculated at the individual interviewer level), 62 subjects are required. Therefore, we recommend that researchers use a more adequate sample size for future work in this area.

One goal of this study was to examine the relationship between interviewer questioning behavior and the accuracy of interviewers' evaluations of applicants. Both lab and field studies have shown that aspects of the interview, such as interview structure and job-relatedness, can affect interviewers' evaluations of applicants (Arvey & Campion, 1982; Dipboye & Gaugler, 1992; Jablin & Miller, 1990; McDaniel et al., 1991; Wiesner & Cronshaw, 1988). Other question characteristics, such as openness, directedness, and secondary nature, as well as whether or not an actual job situation is posed to the applicant, have also been shown to influence the quality of interviewers' evaluations of applicants (Jablin & Miller, 1990; Latham & Saari, 1984; Latham, Saari, Pursell, & Campion, 1980; Motowidlo et al., 1991; Weckley & Gier, 1987; Wiesner & Cronshaw, 1988). The results in the present investigation are inconsistent
with the previous research: there does not appear to be a relationship between questioning behavior of interviewers and their evaluation accuracy. In fact, neither questioning behavior nor information diagnosticity bore a statistically significant relationship with interviewer accuracy at the aggregate interviewer level. There are some critical issues regarding the way some of the question characteristics were operationalized in this study that may explain this result.

For example, Jablin and Miller (1990) found that interviewers who used structured interviews made more accurate and less biased applicant evaluations than those who used unstructured approaches. Yet, there was no statistically significant relationship between interview structure and the accuracy of interviewers’ evaluations in the present investigation.

Consider that structured interviews generally refer to those in which the same questions are asked of all applicants in the same order. Questioning structure was measured in this study according to the extent that interviewers asked the same questions across interviews. However, in many structured interviews, both the questioning process and the applicant scoring procedures are regulated by predetermined rules. The interviewers in this study received only vague written or oral guidelines regarding applicant response scoring. In highly structured interviews, examples of question responses that would be scored in a particular way are provided to interviewers. It is possible that response scoring guidelines exert critical influences on the effect of interview structure on the quality of interviewers’ judgments. If this is the case, then there is no reason to expect that questioning structure would affect the quality of interviewers’ evaluations of applicants in this study.
A second hypothesized influence on interview validity is the job-relatedness of the interview questions. McDaniel and colleagues (1991) found that job-related interviews produce applicant evaluations of higher validity than psychological interviews. Similarly, Weisner and Cronshaw (1988) found higher interview validity when questions were based upon a formal job analysis of the target position than when they were constructed in a less systematic way. In this study, interview questions were considered job-related if they inquired about one or more of the dimensions identified by a job analysis of the target position.

However, in job-related interviews, response scoring guidelines are typically drawn from critical incidents recorded in the job analysis phase of interview development. The interviewers in this study relied upon no such guidelines. Perhaps this deficit explains why the use of job-related questions did not affect interviewer rating accuracy.

The general lack of support for the information gathering model may also be due to the poor predictive validity of both the interview and test scores. In other research conducted on the target organization (Kohn, 1993), pre-interview information accounted for a much larger proportion of variance in interview scores (47%) than did the interview information (1%). Given that interviewers were trained to rely heavily upon pre-interview information (e.g., test scores, military background, arrest record) when rating the interviewees, Kohn's (1993) finding is hardly surprising. Correlational tests among the interview dimensions and the pre-interview tests revealed that the interviewers were generally relying on the tests in the way that they had been instructed (during training) when rating the applicants. Unfortunately, neither test scores nor interview ratings significantly predicted subsequent job performance. Since even the
true score estimates based on this information did not adequately predict employee performance, one might question the wisdom of instructing interviewers to rely on this information in evaluating applicants. Regardless, relationships proposed in the information gathering model may hold only when the interviewers' evaluations have adequate criterion-related validity.

The quality of the performance measure utilized in the target organization may have substantially restricted the potential validity of interview ratings and pre-interview test scores. When this study was designed, it was expected that performance ratings based on a well-constructed performance appraisal instrument would be available for the employees. Unfortunately, this was not the case: only data from a severely flawed performance appraisal instrument were available. This instrument had not been based upon a job analysis of the target position, nor had the supervisors utilizing it been trained in its use. The target organization had indicated that the instrument was not good for measuring performance, and therefore, the ratings it produced were generally disregarded. Since the target organization found little use for information gathered with this instrument, it is not surprising that the pattern of predictor scores bore little resemblance to the pattern of performance scores displayed on this measure. Furthermore, the pattern of performance scores taken at the first appraisal period did not significantly relate to the pattern of those taken at the second performance period ($r = -.06, \text{n.s.}$). Both of these performance ratings were made during what is called the "honeymoon period", where initial novelty and challenge of a new job keeps new employees motivated and keeps their output high. Therefore, the variability in the performance
scores taken during this time should be reduced, thus attenuating the correlation between the interview and performance scores.

This suggests that as a minimum, performance scores taken over a longer period of time should be used to estimate the validity of this interview. However, the attrition rate in the target organization precluded use of more than two performance scores per subject in this study. Perhaps because the interviewers' ratings did not significantly correlate with the performance measure, changes in evaluation accuracy corresponding to changes in interview questioning behavior were concealed. Researchers interested in investigating interview information gathering processes might consider restricting their efforts to interviews with adequate criterion-related validity.

However, there is merit in studying interviews leading to applicant evaluations of poor validity in order to learn how not to go about the interview process. This information would allow for comparison and contrast of the information gathering processes that more and less valid interviewers use. Then, individuals who use the effective techniques can be selected as interviewers, or interviewers can be trained to avoid the procedures that lead to applicant evaluations of poor validity.

To investigate possible differences in information gathering and use of information among individual interviewers, analyses were performed at the individual interviewer level. Research indicates that interviewers differ with respect to a multitude of factors, including their predictive validity (e.g., Dipboye et al., 1990; Dreher et al., 1988; Zedeck et al., 1983). At the individual level of analysis, none of the proposed relationships in the information gathering model were statistically significant. However, the information gathering process of one of the interviewers (Interviewer 2)
was somewhat consistent with the model predictions. Questioning behavior significantly predicted rating accuracy for this interviewer. Although only the conformity factor predicted evaluation accuracy at the $p < .05$ level of significance, the openness factor predicted evaluation accuracy at the $p < .06$ level of significance for this interviewer. This finding is consistent with research suggesting that interviewers who ask open-ended and situational questions make better quality interview ratings than do interviewers who do not use questions of this nature (Dipboye, 1992; Jablin & Miller, 1990).

Moreover, this interviewers’ applicant ratings predicted job performance better than those of the other interviewers’ for the first performance appraisal. Furthermore, although not statistically significant, the relationship between information diagnosticity and evaluation accuracy was highest for this interviewer.

These findings provide some support for research that has found individual interviewer differences in validity (Dipboye et al., 1990; Dreher et al., 1988; Zedeck et al., 1983). It also sheds some light on what may cause these differences---some aspects of questioning behavior. Future research should investigate the causes of these differences. This information would give practitioners a guide for improving the utility of their interviews: the organization might select interviewers who ask interview questions that provide highly diagnostic applicant information. Alternatively, the organization could provide training on how to question applicants such that maximally diagnostic information is elicited during interviews. Researchers might also investigate whether structuring the interview process can correct for interviewers who would produce less valid applicant ratings if left to their own devices during their interviews.
One of the most interesting findings of this study was the strong positive relationship between interviewer questioning style and interviewee response diagnosticity. This relationship suggests that open-ended, situational, secondary, standardized, and job-related questions may be superior to other question types in their ability to elicit diagnostic information. These results are consistent with those of Motowidlo and colleagues (1992), who found that interviewers’ questioning skill (e.g., the degree to which the interviewer asked for job-related information, the degree to which the interviewer elicited information about job situations) significantly and positively correlated with the amount of relevant job information that was elicited from job applicants. In other words, interviewers with good questioning skills gathered more job-relevant information from interviewees than interviewers with poorer questioning skills.

However, Motowidlo et al. (1992) found that applicant qualifications moderated this relationship. When applicants excelled on the interview dimensions, more relevant information could be gathered than when they performed poorly on the dimensions. The applicant information subsequently influenced the quality of the interviewers' ratings: the more job-relevant information that the interviewers gathered, the more accurate the employment judgments made of their applicants. This suggests that more accurate judgments can be made about applicants stronger in the interview dimensions than those weaker in the dimensions. Consistent with these findings, the interviewers’ overall rating of applicants correlated moderately and positively with interview rating accuracy in this study ($r = .23, p < .01$).
Conversely, Jackson and Dipboye (1990) found that interviewers use more diagnostic information gathering processes with more poorly qualified applicants than with those higher in qualifications. The information gathering model predicts that use of more diagnostic information gathering processes should result in better quality evaluations of applicants, regardless of applicant qualifications.

Future research might address the possible moderating role of applicant qualifications in the information gathering processes of interviewers. It is possible that interviewers only use the information gathering processes described in the model when applicants appear qualified for the job. Interviewers may not bother to use complex evaluation processes if they have identified a "knock-out" factor (e.g., skill deficit) for the applicant. Although all interviewees in this study were hired, their interview dimension scores ranged from the lowest to the highest possible score for all but one of the interview dimensions (interpersonal skill). Their mean overall interview score was 65.0 ($SD = 6.5$) of a possible 80 points (range: 42 - 80 points).

The hiring of interviewees with low scores on some of the interview dimensions may be due to the rapid expansion of the target organization during data collection. The organization was forced to hire a relatively high proportion of the applicants who interviewed in order to meet staffing needs. It may be more appropriate to test the information gathering model under conditions more favorable to the organization, when interviewers would be able to meaningfully discriminate among good and poor applicants to a greater degree. In other words, another reason for the lack of support for the model in this study may be lack of relevance for the
interviewers' use of the processes described in the information gathering model when rating applicants.

The most prevalent and consistent finding in this study was the relationship between interviewer questioning behavior and applicant information diagnosticity. One opportunity for future research might be to examine interviewers' use of applicant information. Previous policy capturing research (e.g., Dipboye et al., 1990; Spychalski, 1994) has shown that interviewers do not use applicant information in the same way when rating applicants. Like the effectiveness of individual interviewer questioning behavior, the effectiveness of interviewer use of applicant information may be a meaningful criterion for selecting interviewers. Alternatively, organizations could train interviewers to make the best use of the applicant information that they collected.

One pragmatic suggestion based on our findings is to consider the interview as a selection, evaluation, or training tool for interviewers. Since interviewers are not equally effective, the efficacy of their information gathering processes and the quality of their evaluations should be heavily weighed hiring, appraisal, or training criteria. An applicant rating task in the context of an interview could be a useful behavioral test for training, selecting, or appraising the performance of interviewers. Interviewer rating performance on this task could help employers with one of their most critical ongoing needs: staffing their organizations with individuals (e.g., interviewers) who perform well.
Footnotes

1 Although interviews containing fewer directed questions may result in evaluations of higher validity than those in which more directed questions are asked, there is no reason to expect to find this relationship in this study.

2 Under other conditions, we would propose a third hypothesis (i.e., The more interviewers use directed questions, the less diagnostic the information that they gather from the interview will be, whereas the less interviewers use directed questions, the more diagnostic the information that they gather from the interview will be). However, there is no reason to expect to find this relationship in this investigation.

3 Another method of interview validation has been developed by Kenny and La Voie (1985). These authors designed a validation method that, unlike the traditional technique, does not assume independence of observations in the predictor or criterion. In this method, non-independence in the predictor and criterion scores is estimated with the intra-class correlation among these variables (i.e., the extent to which the scores correlate among one another). This estimate is used to adjust the validity coefficient accordingly. The result is a correlation between predictor and criterion scores with the interviewer and/or rater component partialled out. This method was not adopted in this study because the intra-class correlation for the interview scores was not significant ($r = -.01$, n.s.), indicating that the scores within interviewers were independent. It was not possible to estimate non-independence in the criterion scores due to lack of rater identification information. Under these conditions, it is acceptable to use the traditional $r$ as an estimate of validity.
The Dreher et al. technique cannot be used when turnover is the criterion because it requires a mean and standard deviation of the criterion for calculation. In this study, turnover is a dichotomous variable.
References


Appendix A

Coding Conventions and Coding Forms

- Interview Questions
- Applicant Responses
QUESTION AND RESPONSE CODING CONVENTIONS

IF THE QUESTION CAN BE IDENTIFIED BY A NUMBER FROM THE STANDARDIZED LISTS OF INTERVIEWER QUESTIONS, ONLY CODE THE FOLLOWING VARIABLES: DIRECTED, OPEN-ENDED, AND SECONDARY.

CODE ALL QUESTIONS OR COMMENTS THAT FUNCTION AS QUESTIONS IN THE INTERVIEW. NOTE COMMENTS BY MARKING THE "C" BOX. WRITE, BUT DO NOT CODE, COMMENTS THAT DO NOT FUNCTION AS QUESTIONS IN THE INTERVIEW.

1-DIRECTED: - suggest a bias for a particular type of response
- imply that specific answer is desired

For example: So you have done drugs, haven’t you?
You have been arrested before, right?
You don’t have a short fuse or anything of that nature, do you?
You would never hit an inmate, would you?

Scoring:
Mark directed questions with a “1”; mark questions which are not directed with a “0”

2-OPEN-ENDED: - broad in nature
- allow considerable freedom in determining content, latitude, and amount of information disclosure
- potentially allow applicant maximal opportunity to speak
- not answerable with a simple “yes” or “no”
- no definite limit to the number of responses which could be provided

For example: Why do you want to be a correctional officer?
What do you feel are your strongest character traits?

Examples of closed ended questions:
Have you worked here before?
How old are you?
What is your name?

Scoring:
Mark open-ended questions with a “1”; mark questions which are not open-ended with a “0”
3-SECONDARY: seek additional information about issues or topics alluded to in responses to either preceding primary or secondary questions.

For example: (applicant)...then I went to Georgia.  
(interviewer) What did you do in Georgia?  
(applicant) Well, I'd write him up.  
(interviewer) What do you mean, "write him up"? To whom would you write and what would you say?

Scoring:
Mark secondary questions with a "1"; mark questions which are not secondary with a "0".

4-SITUATIONAL: applicants are asked how they would deal with specific situations on the job.

For example: What would you do if an inmate bit you and he had AIDS?  
What would you do if an inmate refused to do something you asked him to do?

Scoring:
Mark situational questions with a "1"; mark questions which are not situational with a "0".

5-STANDARDIZATION: to what extent does the interviewer ask the same questions of each applicant?

Scoring:
Mark a "1" if the question appears on the standardized list. Mark a "0" if the question does not appear on the list.

If the question is slightly different from its corresponding standardized question, please note the difference in parentheses next to the question number.
a) **Attitude/motivation:**

- Strives to put forth their best effort at all times.
- Potential to maintain uniform and equipment in good condition and personal grooming to acceptable standards.
- Potential to work diligently and conscientiously in carrying out assignments rather than merely "putting in time."
- Takes the initiative to gain more knowledge of correctional duties and improve performance.

<table>
<thead>
<tr>
<th>Questions regarding:</th>
<th>Conception of correctional duties</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Why applicant wants the job</td>
</tr>
<tr>
<td></td>
<td>How applicant feels about him/her self</td>
</tr>
<tr>
<td></td>
<td>Why applicant left former job</td>
</tr>
<tr>
<td></td>
<td>Problems/activities at school</td>
</tr>
<tr>
<td></td>
<td>Disciplinary problems</td>
</tr>
</tbody>
</table>

b) **Learning ability:**

- Potential to quickly learn how to perform all job tasks and to continually improve performance of these tasks.
- Ability to understand written procedures and rules as well as verbal instructions of supervisors.
- Ability to learn and be proficient in the use of equipment necessary to perform correctional officer tasks.

<table>
<thead>
<tr>
<th>Questions regarding:</th>
<th>School (grades, problems)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Job history (progression)</td>
</tr>
</tbody>
</table>

c) **Physical Ability:**

- Potential to adequately control inmates by physical force when necessary.
- Ability to defend oneself and other correctional officers against physical attack by inmates.
- Possess adequate physical strength and stamina necessary to withstand prolonged hours of work activity and retain the strength necessary to perform all job tasks.

<table>
<thead>
<tr>
<th>Questions regarding:</th>
<th>Medical status (limiting condition or disability)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Exercise schedule</td>
</tr>
<tr>
<td></td>
<td>Physical exertion required on the job</td>
</tr>
<tr>
<td></td>
<td>Absenteeism due to medical condition</td>
</tr>
<tr>
<td></td>
<td>Self-defense training</td>
</tr>
<tr>
<td></td>
<td>Getting to work on time (due to physical problems)</td>
</tr>
</tbody>
</table>
(d) **Emotional stability/stress tolerance:**

- potential to perform all duties in an even-tempered manner
- potential to perform duties when under time pressures, emotional stress, or difficult and unpleasant circumstances without losing control of the situation or him or herself

questions regarding: temper
history of emotional problems
reason for job change (due to pressure, change)
drug use
situational reasoning question
military-related disciplinary problems

(e) **Interpersonal skills:**

- ability to earn respect from and get along with other correctional officers and supervisors
- ability to communicate effectively with inmates so as not to incite discontent and create misunderstandings
- ability to become and participate as a team player and contribute to the social and professional order of correctional officers
- willingness to be helpful and courteous to visitors to the unit

questions regarding: getting along with others
biases (racial/sexual)
number of friends/social activities
cooperation
reason for being fired/problems at work/quitting
problems at school
most situational reasoning questions

(f) **Judgment/common sense:**

- ability to use good judgment and common sense when performing many tasks
- when faced with options on verbal action to take in various job situations, takes action which is the most effective for the given situation and is consistent with policies and rules
- ability to see a problem in depth and react quickly with a well planned and thought-out action
- ability to manage personal affairs and avoid excessive use of alcohol and illegal drugs

questions regarding: alcohol/drug abuse
financial affairs
situational reasoning question
conviction record
disciplinary problems
quality of friends
g) **Responsibility:**
- Ability to conform to the rules of society
- Willingness to accept responsibility in previous employment and follow the rules established by employers and to perform required job tasks
- Accepts and fulfills family responsibilities and contributes to the community
- Observes civil laws

  - Questions regarding:
    - Work history (how many changes, times fired)
    - Difficulty with employer
    - Employment status---if not employed, seeking job?
    - Success in military
    - Court-ordered child support/alimony
    - Number of traffic violations
    - Membership in community organization---office held?
    - Ability/willingness to handle guns
    - Getting to work on time
    - Disciplinary problems
    - Drug use
    - Conviction record

h) **Personal Integrity:**
- Potential for complying with criminal laws of society
- Potential for honesty in personal affairs and employment history
- Potential for obeying traffic laws
- Truthfulness

  - Questions regarding:
    - Conviction record
    - Drug use
    - Lying on application
    - Court-ordered child support/alimony
    - Compliance with general, moral laws
    - Compliance with traffic laws

i) **Other job issues:**
- Regarding administrative issues that may affect hiring decision

  - Questions regarding:
    - Location preference
    - Willingness to use weapons
    - Previous application to TDCJ
Note: Questions regarding the following content areas typically address the job dimensions noted. However, there are exceptions to these guidelines: consider them on a case-by-case basis.

<table>
<thead>
<tr>
<th>Question subject</th>
<th>Relevant job dimension(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) drug use</td>
<td>emotional stability, judgment/common sense, personal integrity, responsibility</td>
</tr>
<tr>
<td>2) situational reasoning</td>
<td>emotional stability, interpersonal skills, judgment/common sense</td>
</tr>
<tr>
<td>3) conviction/arrest record</td>
<td>judgment/common sense, personal integrity, responsibility</td>
</tr>
<tr>
<td>4) work problems</td>
<td>interpersonal skills, responsibility</td>
</tr>
<tr>
<td>5) disciplinary problems</td>
<td>judgment/common sense, responsibility, attitude/motivation</td>
</tr>
<tr>
<td>6) school problems</td>
<td>attitude/motivation, learning ability, interpersonal skills</td>
</tr>
</tbody>
</table>

Scoring:

Mark job-related questions with a “1”; mark questions which are not job-related with a “0”.

---

7- **Diagnosticity:**

- to what extent is the interviewee’s answer helpful in rating him across typical social judgment domains
- to what extent does the applicant’s answer assist the interviewer in deciding how good of an employee that individual will make
- to what extent does the answer help the interviewer decide if the interviewee is “high”, “medium”, or “low” on each of the interview dimensions?

This variable has to do with the answer that the interviewee provides. Assume that the interviewee is telling the truth in providing the answer, but remember that he or she is constrained by the interview context (i.e., the applicant is motivated to present a positive image to the interviewer). Usually, a typical or “average” response will be less diagnostic than one that is more uncommon.

When rating information diagnosticity, only rate information that is new. If an applicant repeats information in a subsequent answer, disregard it in rating the information in the second response.

If the applicant provides information completely redundant with a previous answer, do not code it. Write “RED” in the columns where you normally would code it.

If an applicant asks a question or makes a comment (other than a response to a question), mark the “C” box.

Do not record or code questions after the “Do you have any questions?” point unless they provide diagnostic information. Please listen to the entire interview to be sure that no questions provide diagnostic information. Code questions the applicant asks before the “Do you have any questions?” point if applicable.

Each answer will be rated on three types of diagnosticity.
1) **Typical diagnosticity (TD):** helpful in making decisions about individuals across a large number of social judgment domains.

For example: "I had one felony conviction in 1988 for armed robbery."

This information is helpful in categorizing that this individual may not be trustworthy or honest, which is relevant to many social judgment domains. Furthermore, it is helpful in determining that this individual may not be a good candidate for a job that involves enforcing the law since he or she has demonstrated disregard for it. Therefore, this information would be given a high rating on the following scale:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>useful in categorizing across no domains</td>
<td>useful in categorizing across a few domains</td>
<td>useful in categorizing across some domains</td>
<td>useful in categorizing across many domains</td>
<td>useful in categorizing across all domains</td>
</tr>
</tbody>
</table>

*Note:* In making this rating, consider how the average person in the average population would respond.

2) **Interview diagnosticity (ID):** helpful in making decisions about applicants across a large number of job-related domains.

For example: "I use marijuana."

This information is helpful to the interviewer in deciding how the applicant should be rated on the dimensions of personal integrity, emotional stability/stress tolerance, and judgment/common sense. This information is useful in categorizing across three job-related domains. Therefore, it would be given a 3.

*Scoring:* count the number of job-related domains to which the information applies (number of non-"1"s).

3) **Interview judgment-specific diagnosticity (LISD):**

-very helpful in making decisions about an applicant in a specific job-related domain.

-how helpful is that information in the particular domain to which it applies

For example: "I was fired from my past two jobs."

In trying to determine the applicant's attitude/motivation, this information is more helpful in determining what rating to give this applicant than the information that she has a general understanding of the duties of a correctional officer. Therefore, this first piece of information would be given a high rating on the following scale:

<table>
<thead>
<tr>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>no help in determining rating</td>
<td>somewhat helpful in determining rating</td>
<td>moderately helpful in determining rating</td>
<td>very helpful in determining rating</td>
<td>provides all info needed to determine rating</td>
</tr>
</tbody>
</table>

*Scoring:*

For each job-related domain (identified in the ID rating), rate the information in terms of interview judgment-specific diagnosticity.
Note:
1) The information that the applicant is a former TDCJ correctional officer should be given a "3" on all IUSD dimensions, and an "8" on ID.

2) In making this rating, consider how the average TDCJ job applicant would respond.

3) A court martial can be granted for something as minor as not shining one's shoes or as severe as killing one's bunk mate. Unless you know why the action was issued, this information is not particularly diagnostic.

Examples:

1) (app): I use illegal drugs

   TD: 4
   ID (=3): personal integrity 4
           emotional stability/stress tolerance 3
           judgment/common sense 3

2) (app): It takes a lot to make me angry.

   TD: 3
   ID (=2): emotional stability/stress tolerance 3
            interpersonal skill 2

3) (app): My location preference is Huntsville.

   TD: 2
   ID: 0
   IUSD: 0

Additional examples of diagnosticity ratings of frequently asked questions are provided on the following pages.
Appendix B

Coding Conventions and Coding Forms

True Score Estimates
JOB FACTOR DEFINITIONS

Attitude/Motivation: Strives to put forth their best effort at all times; potential to maintain uniform and equipment in good condition and personal grooming to acceptable standards; potential to work diligently and conscientiously in carrying out assignments rather than merely "putting in time"; takes the initiative to gain more knowledge of correctional duties and improve performance; recognizes own deficiencies and corrects them. This dimension is important in performing job tasks related to enforcing inmate dress and grooming standards (e.g. task 88); it also relates to willingness to improve performance in all job tasks.

Perceptiveness/Alertness: Observes details and demonstrates alertness and the ability to perceive potential problems with inmates and visitors to the unit; possesses the ability to quickly comprehend oral and written communication and formulate a quick and effective response. Perceptiveness and alertness is essential to the performance of many correctional officer duties, such as inspecting to ensure that only those inmates who are assigned to a cellblock or dorm area are allowed into those areas and that no inmates are allowed on the runs during count (Task 54 and 34) and investigating infractions of rules and regulations and incidents and occurrences involving inmates (Task 88 and 89).

Communication Skills: Ability to speak and write clearly and concisely; potential for speaking to inmates in strong and commanding tone; ability to prepare written reports which accurately reflect incidents; ability to effectively communicate orally with inmates, to include being able to listen and pass on information. Communication skills are important in performing many correctional officer tasks, to include: writing necessary incident, offense and use of force reports (Task 6); issuing verbal commands to inmates (Task 73); calling out to gain the attention of inmates, employees or other officers and notifying supervisor or other officers of emergencies or potentially threatening situations (Task 2 and 3); and, answering inmate questions regarding unit/department operations and procedures.

Learning Ability: Potential to quickly learn how to perform all job tasks and to continually improve performance of these tasks; ability to understand written procedures and rules as well as verbal instructions of supervisors; and, ability to learn and be proficient in the use of equipment necessary to perform correctional officer tasks. The ability to learn unit rules, procedures, policy bulletins and directives (Task 94 and 97) is important to any correctional officer. He/she must also learn to effectively use walkie talkies and two-way radios in vehicles (Task 20 and 21); use inmate restraining devices (Task 74); use weapons and riot gear (Task 18 and 110); and, to administer first aid to inmates in emergency situations (Task 112).
Physical Ability: Potential to adequately control inmates by physical force when necessary; ability to defend oneself and other correctional officers against physical attack by inmates; possess adequate physical strength and stamina necessary to withstand prolonged hours of work. Activity and retain the strength necessary to perform all job tasks. This dimension is important in the performance of several correctional officer duties, such as: putting restraining devices on inmates (Task 74); removing any inmate causing a serious disturbance from work area (Task 76); using physical force to protect self (Task 78); settling or mediating inmate disputes (Task 91); and, restraining; overpowering, and/or subduing inmates (Task 105).

Emotional Stability/Stress Tolerance: Potential to perform all duties in an even-tempered manner; potential to perform duties when under time pressures, emotional stress, or difficult and unpleasant circumstances without losing control of the situation or himself. This dimension is reflected in many aspects of correctional officer duties, such as securing crime scenes (Task 71); using physical force to protect self (Task 78); using deadly force when necessary (Task 85); controlling life threatening situations (Task 107); and, evacuating inmates or staff during emergencies (Task 135).

Interpersonal Skills: Ability to earn the respect and get along with other correctional officers and supervisors; ability to communicate effectively with inmates so as not to incite discontent and create misunderstandings; ability to become and participate as a team player and contribute to the social and professional order of correctional officers; and, willingness to be helpful and courteous to visitors to the unit. Interpersonal skills are important in the performance of all tasks involving correctional officer interaction with other correctional officers and inmates. These include: informing relief officers of all information pertaining to the day which is vital to maintaining the security of the unit (Task 27); explaining unit rules and procedures to new officers (Task 95); supervising inmate work details outside of the unit (Task 65); answering inmate questions regarding unit operations and procedures (Task 29); supervising support service inmates during their work inside the unit (Task 64); issuing verbal commands to inmates (Task 73); settling or mediating inmate disputes (Task 91); explaining rules and procedures to new inmates (Task 94); and, controlling life threatening situations.

Judgement/Common Sense: Ability to use good judgement and common sense when performing many tasks; when faced with options on verbal action to take in various job situations, takes action which is the most effective for the given situation and is consistent with policies and rules; ability to see a problem in depth and react quickly with a well planned and thought-out action; and, ability to manage personal affairs and avoid excessive use of alcohol and illegal drugs. This dimension is reflected in many aspects of correctional officer duties, such as: obtaining emergency attention for inmates (Task 28); pat searching visitors to the unit (Task 57); using deadly force when necessary (Task 79); assessing severity of incidents before taking appropriate action (Task 84); settling or mediating inmate disputes (Task 91); and, controlling life threatening situations.
Responsibility: Ability to conform to the rules of society; willingness to accept responsibility in previous employment and follow the rules established by employers and to perform required job tasks; accepts family and fulfills family responsibilities and contributes to the community; and, observes civil laws. Responsibility is an important dimension in successfully performing many correctional officer tasks. These include: securing crime scenes (Task 71); storing and safeguarding inmate property (Task 72); escorting inmates to and from specified destinations outside the confines of the unit (Task 62); using deadly force when necessary (Task 79); controlling life-threatening situations (Task 107); and, handling and using weapons (Task 108).

Personal Integrity: Potential for complying with criminal laws of society; potential for honesty in personal affairs and employment history; potential for obeying traffic laws; and, truthfulness. Personal Integrity is an important dimension which applies to the performance of most correctional officer tasks, such as in: writing necessary incident, offense and use of force reports (Task 6); maintaining inventories of weapons, equipment and other supplies (Task 8); informing relief officers of all information pertaining to the day which is vital to maintaining the security of inmates (Task 27); taking inmate count (Task 39); investigating infractions of rules and regulations (Task 88); and, recognizing inmate schemes to gain testimony in disciplinary hearings.
Correctional Officer Applicant

**SCORING GUIDELINES**

**Attitude/Motivation**

*Low* - Does not care if he gets the job. Uncooperative and unenthusiastic. Sloppy, poorly groomed, unkempt, poor posture. Has no idea of what correctional duties involve.

*Medium* - Needs the job. Thinks he will like it. Cooperative. Acceptable dress, grooming, posture, etc. Some understanding of correctional duties in TDC.

*High* - Great desire for the job. Very cooperative and enthusiastic. Feels good about himself. Clean, neatly dressed, well groomed, good posture. Has researched correctional duties and has a good understanding of them.

**Perceptiveness/Alertness**

*Low* - Does not respond initially when called for interview. Not alert. Does not understand interview questions and responds slowly.

Observation and Memory (Part I) score 45 to 60.

*Medium* - Normal alertness. Understands interview questions and provides average responses.

Observation and Memory (Part I) score 65 to 80.

*High* - Very alert. Fully understands interview questions and responds quickly.

Observation and Memory (Part I) score 85 to 100.
Communication Skills

Low - Difficulty in speaking and understanding, improper grammar, gives rambling answers to questions. Difficulty in understanding and completing employment application and other documents. Weak voice, does not speak with authority.

Reading Comprehension (Part III) score 40 to 55.
Vocabulary (Part IV) score 55 to 65.
Arithmetic (Part V) score 65 to 75.

Medium - Speaks and understands with little difficulty. Mostly good grammar. Essentially understood and completed employment application and other documents but made errors. Average force of voice.

Reading Comprehension (Part III) score 60 to 75.
Vocabulary (Part IV) score 70 to 80.
Arithmetic (Part V) score 80 to 85.

High - No difficulty in understanding questions. Speaks clearly, concisely and forcefully. Good grammar. Easy to understand. Employment application and other documents understood and completed without errors.

Reading Comprehension (Part III) score 80 to 100.
Vocabulary (Part IV) score 85 to 100.
Arithmetic (Part V) score 90 to 100.

Learning Ability

Low - Evidence of difficulty in school, poor grades, non-progressive job history.

Reading Comprehension (Part III) score 40 to 55.
Arithmetic (Part V) score 65 to 75.

Medium - Average grades in school, progressive job history.

Reading Comprehension (Part III) score 60 to 75.
Arithmetic (Part V) score 80 to 85.

High - Excellent school record, high grades. High progressive job history.

Reading Comprehension (Part III) score 80 to 100.
Arithmetic (Part V) score 90 to 100.
Physical Ability


Medium - Some medical problems. Some disability rating. Some absenteeism from the job for medical reasons. Medical condition is likely to result in some limitations to job assignment. Average physical condition. Regular program of moderate exercise or employed in a job requiring moderate physical demands. Some fat but not near weight limit.

High - No medical problems, disability rating or job limitations. No record excessive absenteeism from past job for medical reasons. Good physical condition. Regular program of strenuous exercise or employed in a very physically demanding job. No excess fat. Well within weight standard.

NOTES:

1) Height or body frame size of the applicant is not a consideration in the evaluation process.

2) Applicants will not be scored lower if they have not had self-defense training. However, if they have completed a self-defense training program, they may be rated higher than otherwise.

Emotional Stability/Stress Tolerance

Low - Gets rattled quickly. Quick temper. History of emotional problems. Quit jobs because he could not cope with changes or pressure.

Medium - Even tempered. Does not get rattled easily.

High - Responds to question forcefully. No history of emotional problems or job changes due to inability to cope with change or pressure.

Interpersonal Skills

Low - Problem with getting along with other people. Uncooperative. Loner. Racially or sexually biased. Discourteous. Tends to annoy others.

Medium - Gets along with other people. Average cooperation. Can work effectively with other people. Courteous.

High - Gets along great with others. Very cooperative. Not socially or sexually biased. Very courteous. Team player.
Judgement/Common Sense

Low - Completely inappropriate responses to situational reasoning questions. Inappropriately dressed for interview. Strong record of alcohol or drug abuse (especially recent or current) or difficulty in managing personal financial affairs.

Situational Reasoning (Part II) score 80 to 85.

Medium - Average response to situational reasoning questions. Some record of alcohol or drug abuse. No special effort to look sharp and dress appropriately for interview.

Situational Reasoning (Part II) score 90.

High - Good response to situational reasoning questions. No record or no recent record of alcohol or drug abuse. No difficulty in managing personal financial affairs or other indication of poor judgement. Very appropriately dressed for interview.

Situational Reasoning (Part III) score 95 to 100.

Responsibility

Low - Frequent job changes, fired more than once. Evidence of difficulty with employers and failure to follow rules. Lack of active job hunting or productive work if not presently working. Uncooperative. Incomplete application. Does not provide direct response to questions. Has not or successfully worked in structured environment (i.e. military). Failure to provide Court ordered child support. Seven or more traffic violations within past 5 years.

Medium - Some movement between jobs. Fired once. Some indication that applicant is actively seeking employment, if not presently working. Cooperative. Worked some in structured environment. Four to 6 traffic violations within past 5 years.


NOTES:

1) Traffic violations do not include the following, which will be considered as misdemeanors for the purpose of rating under Personal Integrity:
   °Driving while License suspended
   °Assault with a Motor Vehicle
   °Failure to stop and render aid
   °Violate promise to appear
Personal Integrity

Disqualified - Intentional concealment of conviction in Question 13 or other important facts in employment application. Less than 15 years since completion of sentence for felony conviction.

Low - More than one misdemeanor conviction. Felony conviction over 15 years ago. Any hard drug use and/or conviction from 1 to 5 years ago. Marijuana use within past year.

Medium - Hard drug use and/or conviction more than 5 years ago. One misdemeanor conviction. Marijuana use from 1 to 5 years ago.

High - No criminal history. Marijuana use more than 5 years ago. No hard drug use.

NOTES:

1) Non-traffic offenses listed on the DL check will be counted as misdemeanors for the purposes of scoring Personal Integrity. However, dispositions will not be required for these offenses, unless they are also listed in the TCIC/NCIC check.

2) Do require dispositions on all arrests or convictions reported in the application, pre-interview questionnaire, interview or any other source.

3) For the purposes of scoring, forfeiture of bond in lieu of prosecution or restitution in lieu of prosecution are to be considered as misdemeanor convictions.

4) Arrests or convictions reported on pre-interview questionnaire, interview or other source should be noted on evaluation form.

5) Disqualification criteria:

- If you learn of a conviction not listed in Question 13 from any source after the application is submitted, the applicant will be disqualified. These sources include the pre-interview questionnaire, the applicant in the interview, TCIC/NCIC, Fingerprint Card, or disposition.

- Do not disqualify an applicant who fails to report offenses which are listed only on the DL check, except DWI.

- Do not disqualify an applicant solely for not listing an arrest or other important information on the pre-interview questionnaire. However, where an applicant has not listed such information on the pre-interview questionnaire, he or she should be given one lesser rating in Personal Integrity.
6) Military Record

An applicant's military record should be carefully considered and heavily weighted in the evaluation process. Applicants without military service will not be penalized.

Indicators of successful military service are: HONORABLE Discharge, no Article 15's or Court Martial, continued military and civilian education, regular promotions, and achievement, commendation and good conduct medals.

Indicators of unsuccessful military service are: GENERAL, UNDER HONORABLE CONDITIONS, UNDER OTHER THAN HONORABLE CONDITIONS, DISHONORABLE, or BAD CONDUCT DISCHARGE, Article 15's or Court Martial, lack of promotion, discharge at a pay grade of E-1 or E-2.

Under the Uniform Code of Military Justice, discipline is administered by: Article 15, Summary Court Martial, Special Court Martial and General Court Martial. Article 15's or even Court Martial may be administered for something less than a misdemeanor (i.e. late for work) or for an offense which equates to a misdemeanor. A court martial conviction may be for an offense which equates to a misdemeanor or felony. When evaluating applicants, look at the offenses provided by Article 15 or court martial and try to equate civil offenses.

Military offenses which equate to a civil felony conviction do not necessarily disqualify an applicant for 15 years. Flag these cases and discuss with a supervisor.
<table>
<thead>
<tr>
<th></th>
<th></th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>Motivation</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>2</td>
</tr>
<tr>
<td>2.</td>
<td>Perceptiveness/Alertness</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1, 2, 4, 6</td>
</tr>
<tr>
<td>3.</td>
<td>Communication Skills</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>1, 2, 6</td>
</tr>
<tr>
<td>4.</td>
<td>Learning Ability</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1, 2, 6</td>
</tr>
<tr>
<td>5.</td>
<td>Physical Ability</td>
<td>1</td>
<td>2</td>
<td>4</td>
<td>1, 2, 6</td>
</tr>
<tr>
<td>6.</td>
<td>Emotional Stability/</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>1, 2</td>
</tr>
<tr>
<td></td>
<td>Stress Tolerance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7.</td>
<td>Interpersonal Skills</td>
<td>2</td>
<td>4</td>
<td>8</td>
<td>1, 2, 4</td>
</tr>
<tr>
<td>8.</td>
<td>Judgment/Common Sense</td>
<td>3</td>
<td>6</td>
<td>12</td>
<td>1, 2, 3, 4, 5, 6</td>
</tr>
<tr>
<td>9.</td>
<td>Responsibility</td>
<td>3</td>
<td>6</td>
<td>12</td>
<td>1, 2, 3, 4, 5</td>
</tr>
<tr>
<td>10.</td>
<td>Personal Integrity</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>1, 2, 3, 4, 5</td>
</tr>
</tbody>
</table>

Sources:
1. Application
2. Interview
3. TCIC/NCIC
4. Reference letters
5. FP card
6. Test Scores
Appendix C

Intercorrelations Among Study Variables
Appendix C

Intercorrelations Among Study Variables

Aggregate Interviewer Level (N = 145-149)

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Attitude /motivation</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Perception</td>
<td>.11</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Communication skill</td>
<td>.09</td>
<td>.15</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Learning ability</td>
<td>.08</td>
<td>.06</td>
<td>.63**</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Physical ability</td>
<td>.19*</td>
<td>.12</td>
<td>.03</td>
<td>.07</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Emotional stability /stress tolerance</td>
<td>.06</td>
<td>-.11</td>
<td>-.15</td>
<td>-.08</td>
<td>-.04</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Interpersonal skill</td>
<td>.15</td>
<td>-.02</td>
<td>.03</td>
<td>.05</td>
<td>-.03</td>
<td>.17*</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Judgement/ common sense</td>
<td>-.09</td>
<td>.06*</td>
<td>.03</td>
<td>.13</td>
<td>-.09</td>
<td>.05</td>
<td>.02</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9. Responsibility</td>
<td>.02</td>
<td>-.19*</td>
<td>-.06</td>
<td>-.13</td>
<td>-.06</td>
<td>.13</td>
<td>.16</td>
<td>.09</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10. Personal integrity</td>
<td>-.13</td>
<td>-.07</td>
<td>.05</td>
<td>-.03</td>
<td>-.10</td>
<td>.00</td>
<td>.14</td>
<td>.05</td>
<td>-.05</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Interview score</td>
<td>.34**</td>
<td>.18*</td>
<td>.47**</td>
<td>.42**</td>
<td>.14</td>
<td>.21**</td>
<td>.35**</td>
<td>.43**</td>
<td>.40**</td>
<td>.33**</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>12. Performance appraisal I</td>
<td>-.01</td>
<td>.02</td>
<td>.01</td>
<td>.02</td>
<td>.06</td>
<td>-.03</td>
<td>-.03</td>
<td>-.06</td>
<td>-.02</td>
<td>.15</td>
<td>.01</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations Among Study Variables**

### Aggregate Interviewer Level

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Performance appraisal II</td>
<td>-.06</td>
<td>.00</td>
<td>-.16*</td>
<td>.02</td>
<td>-.14</td>
<td>.05</td>
<td>.01</td>
<td>-.08</td>
<td>.09</td>
<td>-.01</td>
<td>-.07</td>
<td>.09</td>
</tr>
<tr>
<td>14. Turnover</td>
<td>.01</td>
<td>-.03</td>
<td>.03</td>
<td>-.01</td>
<td>.11</td>
<td>-.11</td>
<td>.02</td>
<td>-.04</td>
<td>-.09</td>
<td>-.02</td>
<td>-.03</td>
<td>-.09</td>
</tr>
<tr>
<td>15. True attitude/motivation</td>
<td>.18*</td>
<td>.15</td>
<td>.19*</td>
<td>.16</td>
<td>-.04</td>
<td>-.01</td>
<td>.04</td>
<td>.09</td>
<td>.00</td>
<td>.00</td>
<td>.17*</td>
<td>.00</td>
</tr>
<tr>
<td>16. True perception</td>
<td>.09</td>
<td>.81**</td>
<td>.24**</td>
<td>.12</td>
<td>.05</td>
<td>-.09</td>
<td>.04</td>
<td>.06</td>
<td>-.09</td>
<td>-.02</td>
<td>.21**</td>
<td>.04</td>
</tr>
<tr>
<td>17. True communication skill</td>
<td>.09</td>
<td>.15</td>
<td>.58**</td>
<td>.49**</td>
<td>-.07</td>
<td>-.07</td>
<td>.07</td>
<td>.11</td>
<td>-.04</td>
<td>.10</td>
<td>.39**</td>
<td>.02</td>
</tr>
<tr>
<td>18. True learning ability</td>
<td>.17*</td>
<td>.03</td>
<td>.38**</td>
<td>.43**</td>
<td>.01</td>
<td>.18*</td>
<td>.01</td>
<td>.12</td>
<td>.04</td>
<td>.06</td>
<td>.39**</td>
<td>.03</td>
</tr>
<tr>
<td>19. True physical ability</td>
<td>.10</td>
<td>.16*</td>
<td>.00</td>
<td>-.01</td>
<td>.29**</td>
<td>-.03</td>
<td>.09</td>
<td>-.17*</td>
<td>-.08</td>
<td>.00</td>
<td>-.01</td>
<td>.07</td>
</tr>
<tr>
<td>20. True emotional stability/stress tolerance</td>
<td>.20*</td>
<td>.06</td>
<td>-.09</td>
<td>.01</td>
<td>.11</td>
<td>-.03</td>
<td>.08</td>
<td>.04</td>
<td>.00</td>
<td>.12</td>
<td>.09</td>
<td>.05</td>
</tr>
<tr>
<td>21. True interpersonal skill</td>
<td>-.01</td>
<td>-.14</td>
<td>.07</td>
<td>-.02</td>
<td>-.15</td>
<td>.01</td>
<td>.01</td>
<td>-.03</td>
<td>.22**</td>
<td>-.08</td>
<td>.01</td>
<td>-.06</td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations Among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>22. True judgement/common sense</td>
<td>-.09</td>
<td>.23**</td>
<td>.06</td>
<td>.06</td>
<td>-.09</td>
<td>.04</td>
<td>.43**</td>
<td>-.16*</td>
<td>.09</td>
<td>.13</td>
<td>-.09</td>
<td></td>
</tr>
<tr>
<td>23. True responsibility</td>
<td>-.05</td>
<td>-.09</td>
<td>.06</td>
<td>.07</td>
<td>-.09</td>
<td>.20**</td>
<td>.16*</td>
<td>.14**</td>
<td>.12</td>
<td>.35**</td>
<td>-.01</td>
<td></td>
</tr>
<tr>
<td>24. True personal integrity</td>
<td>-.15</td>
<td>.14</td>
<td>.11</td>
<td>.06</td>
<td>-.07</td>
<td>.04</td>
<td>.06</td>
<td>-.02</td>
<td>-.12</td>
<td>.56**</td>
<td>.18*</td>
<td>.06</td>
</tr>
<tr>
<td>25. True interview score</td>
<td>.13</td>
<td>.37**</td>
<td>.42**</td>
<td>.35**</td>
<td>-.02</td>
<td>.04</td>
<td>.14</td>
<td>.19*</td>
<td>.06</td>
<td>.19*</td>
<td>.47**</td>
<td>.02</td>
</tr>
<tr>
<td>26. Test scores</td>
<td>.02</td>
<td>.59**</td>
<td>.70**</td>
<td>.64**</td>
<td>.03</td>
<td>-.09</td>
<td>-.02</td>
<td>.26**</td>
<td>-.20*</td>
<td>-.06</td>
<td>.42**</td>
<td>-.04</td>
</tr>
<tr>
<td>27. Accuracy</td>
<td>.23**</td>
<td>.19*</td>
<td>.02</td>
<td>-.13</td>
<td>.04</td>
<td>.00</td>
<td>.11</td>
<td>.13</td>
<td>.22**</td>
<td>.02</td>
<td>.23**</td>
<td>.04</td>
</tr>
</tbody>
</table>

**Question variables:**

| 28. Directednessa                 | -.06 | -.08 | .09  | .16  | -.02 | -.16*| -.03 | .03  | -.13 | .12  | -.03 | -.02 |
| 29. Open-endedness                | -.13 | -.05 | .14  | .16* | .01  | -.21**| -.14 | .02  | -.03 | -.01 | -.02 | -.12 |
| 30. Secondary                     | .03  | .02  | .09  | .09  | .01  | -.26**| -.19*| -.04 | -.16*| .12  | -.07 | .04  |
| 31. Situational                   | -.15 | -.07 | .05  | .08  | -.07 | -.12 | -.06 | .05  | .06  | .15  | .04  | -.06 |
| 32. Attitude/motivation           | .00  | .03  | .09  | .09  | .08  | -.21**| -.15 | -.11 | -.04 | -.16 | -.07 | -.14 |
Appendix C (continued)

**Intercorrelations Among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>33. Learning ability</td>
<td>.04</td>
<td>-.06</td>
<td>-.16*</td>
<td>-.20*</td>
<td>.06</td>
<td>.14</td>
<td>.08</td>
<td>-.06</td>
<td>.02</td>
<td>-.12</td>
<td>-.11</td>
<td>.06</td>
</tr>
<tr>
<td>34. Physical ability</td>
<td>.06</td>
<td>.08</td>
<td>-.03</td>
<td>-.07</td>
<td>-.02</td>
<td>.13</td>
<td>.16*</td>
<td>.02</td>
<td>.06</td>
<td>-.14</td>
<td>.03</td>
<td>.13</td>
</tr>
<tr>
<td>35. Emotional stability/stress tolerance</td>
<td>-.14</td>
<td>-.07</td>
<td>.06</td>
<td>.09</td>
<td>.07</td>
<td>-.23*</td>
<td>-.14</td>
<td>.01</td>
<td>.01</td>
<td>-.09</td>
<td>-.02</td>
<td>-.07</td>
</tr>
<tr>
<td>36. Interpersonal skill</td>
<td>-.14</td>
<td>-.05</td>
<td>.06</td>
<td>.00</td>
<td>.01</td>
<td>-.26**</td>
<td>-.17*</td>
<td>.01</td>
<td>.09</td>
<td>.09</td>
<td>-.01</td>
<td>-.10</td>
</tr>
<tr>
<td>37. Judgment/common sense</td>
<td>-.02</td>
<td>-.02</td>
<td>.01</td>
<td>.13</td>
<td>.08</td>
<td>-.19*</td>
<td>-.28**</td>
<td>-.02</td>
<td>-.10</td>
<td>-.15</td>
<td>-.15</td>
<td>-.07</td>
</tr>
<tr>
<td>38. Responsibility</td>
<td>.11</td>
<td>-.03</td>
<td>-.03</td>
<td>.05</td>
<td>.11</td>
<td>-.19*</td>
<td>-.23**</td>
<td>-.13</td>
<td>-.14</td>
<td>-.27**</td>
<td>-.23**</td>
<td>-.06</td>
</tr>
<tr>
<td>39. Personal integrity</td>
<td>-.03</td>
<td>-.07</td>
<td>-.06</td>
<td>.03</td>
<td>.14</td>
<td>-.07</td>
<td>-.26**</td>
<td>-.02</td>
<td>-.05</td>
<td>-.25**</td>
<td>-.19*</td>
<td>.06</td>
</tr>
<tr>
<td>40. Standardization</td>
<td>-.23**</td>
<td>-.08</td>
<td>-.09</td>
<td>-.07</td>
<td>-.06</td>
<td>-.07</td>
<td>.06</td>
<td>.09</td>
<td>.13</td>
<td>.09</td>
<td>.02</td>
<td>-.07</td>
</tr>
</tbody>
</table>

**Diagnosticity variables:**

<p>| 41. Typical diagnosticity            | -.07| .02| -.04| -.05| .05| -.25**| -.01| .05| .00| .01| -.06| .00 |
| 42. Attitude/motivation              | -.06| .03| .02| .07| .06| -.07| -.14| -.02| -.07| .00| -.05| -.09 |
| 43. Learning ability                 | .13| .01| -.06| -.06| .04| .12| .08| .00| -.03| .06| .04| .01 |
| 44. Physical ability                 | .09| -.03| -.05| -.13| .00| .10| .11| -.07| .08| -.07| .00| .15 |</p>
<table>
<thead>
<tr>
<th>Variable</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>45. Emotional stability/sense of stress tolerance</td>
<td>-.15</td>
<td>-.02</td>
<td>.15</td>
<td>.6*</td>
<td>.03</td>
<td>-.15</td>
<td>-.01</td>
<td>-.01</td>
<td>-.02</td>
<td>-.04</td>
<td>.01</td>
<td>-.12</td>
</tr>
<tr>
<td>46. Interpersonal skill/common responsibility</td>
<td>-.02</td>
<td>-.02</td>
<td>.13</td>
<td>.12</td>
<td>.03</td>
<td>-.22</td>
<td>-.05</td>
<td>-.04</td>
<td>-.04</td>
<td>-.07</td>
<td>.00</td>
<td>-.04</td>
</tr>
<tr>
<td>47. Judgment/understanding</td>
<td>-.08</td>
<td>-.08</td>
<td>-.06</td>
<td>-.06</td>
<td>.00</td>
<td>-.08</td>
<td>-.06</td>
<td>-.11</td>
<td>-.02</td>
<td>-.09</td>
<td>-.12</td>
<td>-.08</td>
</tr>
<tr>
<td>48. Responsibility/specific diagnosis</td>
<td>.06</td>
<td>.06</td>
<td>.02</td>
<td>.13</td>
<td>.25**</td>
<td>-.15</td>
<td>-.08</td>
<td>-.14</td>
<td>-.14</td>
<td>-.16</td>
<td>-.12</td>
<td>-.16</td>
</tr>
<tr>
<td>49. Personal integrity/specific diagnosis</td>
<td>-.04</td>
<td>-.04</td>
<td>.02</td>
<td>.16*</td>
<td>-.25**</td>
<td>-.19*</td>
<td>-.25**</td>
<td>.01</td>
<td>-.06</td>
<td>-.12</td>
<td>-.16</td>
<td>-.08</td>
</tr>
<tr>
<td>50. Interview judgment/specific diagnosis</td>
<td>-.02</td>
<td>-.02</td>
<td>.02</td>
<td>.08</td>
<td>.09</td>
<td>-.22**</td>
<td>-.14</td>
<td>-.09</td>
<td>-.08</td>
<td>-.07</td>
<td>-.13</td>
<td>-.04</td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations Among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>13. Performance appraisal II</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14. Turnover</td>
<td>-.14</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. True attitude /motivation</td>
<td>-.06</td>
<td>-.03</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16. True perception</td>
<td>.05</td>
<td>-.02</td>
<td>.32**</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. True communication-skill</td>
<td>-.05</td>
<td>.02</td>
<td>.24**</td>
<td>.28**</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18. True learning ability</td>
<td>.11</td>
<td>.08</td>
<td>.17*</td>
<td>.16</td>
<td>.38**</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. True physical ability</td>
<td>-.02</td>
<td>.09</td>
<td>-.02</td>
<td>.17*</td>
<td>-.01</td>
<td>-.04</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20. True emotional stability/stress tolerance</td>
<td>.17*</td>
<td>.01</td>
<td>.29**</td>
<td>.06</td>
<td>-.01</td>
<td>.09</td>
<td>.05</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21. True interpersonal skill</td>
<td>.06</td>
<td>-.04</td>
<td>.28**</td>
<td>.00</td>
<td>-.08</td>
<td>.19*</td>
<td>-.13</td>
<td>.19*</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note: * indicates significance at the .05 level; ** indicates significance at the .01 level.*
### Intercorrelations Among Study Variables

#### Aggregate Interviewer Level

<table>
<thead>
<tr>
<th>Variable</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>22. True judgement/common sense</td>
<td>.06</td>
<td>-.03</td>
<td>.05</td>
<td>.13</td>
<td>.08</td>
<td>.02</td>
<td>-.12</td>
<td>.04</td>
<td>-.08</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23. True responsibility</td>
<td>.05</td>
<td>-.21**</td>
<td>.22**</td>
<td>-.04</td>
<td>.14</td>
<td>.18*</td>
<td>-.18*</td>
<td>.06</td>
<td>.25**</td>
<td>-.02</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>24. True personal integrity</td>
<td>.02</td>
<td>-.06</td>
<td>.04</td>
<td>.09</td>
<td>.18*</td>
<td>.03</td>
<td>.00</td>
<td>.00</td>
<td>-.09</td>
<td>.14</td>
<td>.07</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>25. True interview score</td>
<td>.08</td>
<td>-.05</td>
<td>.59**</td>
<td>.53**</td>
<td>.61**</td>
<td>.57**</td>
<td>.13</td>
<td>.34**</td>
<td>-.04</td>
<td>.27**</td>
<td>.42**</td>
<td>.29**</td>
<td>1.0</td>
</tr>
<tr>
<td>26. Test scores</td>
<td>-.08</td>
<td>.01</td>
<td>.20**</td>
<td>.59**</td>
<td>.59**</td>
<td>.39**</td>
<td>.00</td>
<td>-.01</td>
<td>.42**</td>
<td>.26**</td>
<td>.02</td>
<td>.12</td>
<td>.55**</td>
</tr>
<tr>
<td>27. Accuracy</td>
<td>.00</td>
<td>.08</td>
<td>.19*</td>
<td>.29**</td>
<td>.09</td>
<td>.21*</td>
<td>.09</td>
<td>.39**</td>
<td>.32**</td>
<td>.12</td>
<td>.11</td>
<td>-.02</td>
<td>.41**</td>
</tr>
</tbody>
</table>

**Question variables:**

| 28. Directednessa                     | .03 | .02  | .14  | -.09 | -.11| .03  | -.08| .12  | .00  | .12  | -.01 | -.06 | .00  |
| 29. Open-endedness                    | -.08| .14  | .15  | -.12 | -.04| -.04 | -.11| .09  | .02  | .03  | .06  | .02  | .00  |
| 30. Secondary                         | -.07| .03  | -.02 | .00  | -.05| .08  | .07 | .13  | -.01| .04  | -.09 | -.02 | .01  |
| 31. Situational                       | .10 | -.05 | .09  | -.10 | -.08| .06  | -.03| .13  | .02  | .11  | .13  | .15  | .09  |
| 32. Attitude/motivation               | -.07| .19* | -.05 | .04  | .00 | -.02 | -.11| -.11 | .02  | -.08 | -.16*| -.07 | -.11 |
Appendix C (continued)

Intercorrelations Among Study Variables

Aggregate Interviewer Level

<table>
<thead>
<tr>
<th>Variable</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question variables:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. Learning ability</td>
<td>.03</td>
<td>.03</td>
<td>.07</td>
<td>-.03</td>
<td>-.18*</td>
<td>-.12</td>
<td>.05</td>
<td>-.08</td>
<td>-.11</td>
<td>.02</td>
<td>-.09</td>
<td>-.10</td>
<td>-.14</td>
</tr>
<tr>
<td>34. Physical ability</td>
<td>.00</td>
<td>-.05</td>
<td>-.05</td>
<td>.10</td>
<td>.09</td>
<td>-.09</td>
<td>-.14</td>
<td>-.07</td>
<td>-.15</td>
<td>.00</td>
<td>-.10</td>
<td>-.05</td>
<td></td>
</tr>
<tr>
<td>35. Emotional stability/stress tolerance</td>
<td>.10</td>
<td>.08</td>
<td>.08</td>
<td>-.08</td>
<td>-.04</td>
<td>.01</td>
<td>-.03</td>
<td>.15</td>
<td>.05</td>
<td>.08</td>
<td>.01</td>
<td>.10</td>
<td>.05</td>
</tr>
<tr>
<td>36. Interpersonal skill</td>
<td>.01</td>
<td>.07</td>
<td>.09</td>
<td>-.08</td>
<td>-.12</td>
<td>.01</td>
<td>-.10</td>
<td>.12</td>
<td>.04</td>
<td>.02</td>
<td>.08</td>
<td>.07</td>
<td>.01</td>
</tr>
<tr>
<td>37. Judgment/common sense</td>
<td>.07</td>
<td>.09</td>
<td>-.06</td>
<td>-.08</td>
<td>-.09</td>
<td>.00</td>
<td>.04</td>
<td>.13</td>
<td>.00</td>
<td>.02</td>
<td>-.12</td>
<td>-.18*</td>
<td>-.09</td>
</tr>
<tr>
<td>38. Responsibility</td>
<td>.01</td>
<td>.09</td>
<td>-.10</td>
<td>-.01</td>
<td>-.03</td>
<td>-.04</td>
<td>-.02</td>
<td>.00</td>
<td>-.01</td>
<td>-.13</td>
<td>-.23**</td>
<td>-.30**</td>
<td>-.19*</td>
</tr>
<tr>
<td>39. Personal integrity</td>
<td>.02</td>
<td>.00</td>
<td>-.06</td>
<td>.00</td>
<td>-.09</td>
<td>-.07</td>
<td>.04</td>
<td>.03</td>
<td>-.09</td>
<td>-.06</td>
<td>-.23**</td>
<td>-.27**</td>
<td>-.19*</td>
</tr>
<tr>
<td>40. Standardization</td>
<td>-.03</td>
<td>.02</td>
<td>.05</td>
<td>-.06</td>
<td>-.09</td>
<td>-.12</td>
<td>-.05</td>
<td>-.08</td>
<td>.07</td>
<td>.01</td>
<td>.16*</td>
<td>.21**</td>
<td>.02</td>
</tr>
<tr>
<td><strong>Diagnosticity variables:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. Typical diagnosticity</td>
<td>-.08</td>
<td>.05</td>
<td>.22**</td>
<td>.03</td>
<td>-.14</td>
<td>-.09</td>
<td>-.16*</td>
<td>.29**</td>
<td>.15</td>
<td>-.12</td>
<td>.03</td>
<td>-.09</td>
<td>.01</td>
</tr>
<tr>
<td>42. Attitude/motivation</td>
<td>-.13</td>
<td>.12</td>
<td>.12</td>
<td>.00</td>
<td>-.06</td>
<td>.00</td>
<td>-.19*</td>
<td>.16*</td>
<td>.13</td>
<td>.06</td>
<td>-.09</td>
<td>.04</td>
<td>.03</td>
</tr>
<tr>
<td>43. Learning ability</td>
<td>.05</td>
<td>-.08</td>
<td>.09</td>
<td>.07</td>
<td>.03</td>
<td>.00</td>
<td>.01</td>
<td>.07</td>
<td>.04</td>
<td>-.08</td>
<td>.13</td>
<td>-.04</td>
<td>.08</td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations Among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>13</th>
<th>14</th>
<th>15</th>
<th>16</th>
<th>17</th>
<th>18</th>
<th>19</th>
<th>20</th>
<th>21</th>
<th>22</th>
<th>23</th>
<th>24</th>
<th>25</th>
</tr>
</thead>
<tbody>
<tr>
<td>44. Physical ability</td>
<td>.03</td>
<td>-.05</td>
<td>-.05</td>
<td>.02</td>
<td>.09</td>
<td>-.01</td>
<td>.09</td>
<td>-.19</td>
<td>-.01</td>
<td>-.23</td>
<td>-.03</td>
<td>-.17</td>
<td>-.07</td>
</tr>
<tr>
<td>Diagnosticity variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>45. Emotional stability/stress tolerance</td>
<td>.01</td>
<td>.07</td>
<td>.16</td>
<td>.03</td>
<td>.02</td>
<td>-.01</td>
<td>-.09</td>
<td>.06</td>
<td>.13</td>
<td>-.01</td>
<td>.17</td>
<td>.03</td>
<td>.12</td>
</tr>
<tr>
<td>46. Interpersonal skill</td>
<td>-.02</td>
<td>-.04</td>
<td>.22</td>
<td>.01</td>
<td>-.01</td>
<td>.00</td>
<td>-.07</td>
<td>.16</td>
<td>.07</td>
<td>-.05</td>
<td>.09</td>
<td>-.01</td>
<td>.09</td>
</tr>
<tr>
<td>47. Judgment/common sense</td>
<td>.12</td>
<td>.03</td>
<td>.11</td>
<td>-.02</td>
<td>.02</td>
<td>-.03</td>
<td>-.03</td>
<td>.22</td>
<td>.13</td>
<td>-.10</td>
<td>.05</td>
<td>-.10</td>
<td>.05</td>
</tr>
<tr>
<td>48. Responsibility</td>
<td>-.03</td>
<td>.10</td>
<td>.00</td>
<td>.02</td>
<td>-.06</td>
<td>-.10</td>
<td>-.01</td>
<td>.13</td>
<td>.09</td>
<td>-.12</td>
<td>-.12</td>
<td>-.25</td>
<td>-.09</td>
</tr>
<tr>
<td>49. Personal integrity</td>
<td>.15</td>
<td>.00</td>
<td>.06</td>
<td>-.04</td>
<td>-.06</td>
<td>-.04</td>
<td>.00</td>
<td>.29</td>
<td>.03</td>
<td>.07</td>
<td>-.08</td>
<td>-.16</td>
<td>.00</td>
</tr>
<tr>
<td>50. Interview judgment-specific diagnosticity</td>
<td>.03</td>
<td>.06</td>
<td>.18</td>
<td>.01</td>
<td>-.01</td>
<td>-.06</td>
<td>-.09</td>
<td>.23</td>
<td>.18</td>
<td>-.11</td>
<td>.05</td>
<td>-.16</td>
<td>.05</td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations Among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>37</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>26. Test scores</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27. Accuracy</td>
<td>.08</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Question variables:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>28. Directedness&lt;sup&gt;a&lt;/sup&gt;</td>
<td>.05</td>
<td>.13</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>29. Open-endedness</td>
<td>.11</td>
<td>.10</td>
<td>.45</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>30. Secondary</td>
<td>.08</td>
<td>.06</td>
<td>.17</td>
<td>.27</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>31. Situational</td>
<td>.03</td>
<td>.06</td>
<td>.31</td>
<td>.64</td>
<td>.37</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>32. Attitude/motivation</td>
<td>.09</td>
<td>-.03</td>
<td>.04</td>
<td>.23</td>
<td>-.08</td>
<td>-.11</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>33. Learning ability</td>
<td>-.24</td>
<td>.01</td>
<td>-.15</td>
<td>-.13</td>
<td>-.29</td>
<td>-.31</td>
<td>-.08</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>34. Physical ability</td>
<td>-.04</td>
<td>.07</td>
<td>-.33</td>
<td>-.42</td>
<td>-.46</td>
<td>-.49</td>
<td>-.07</td>
<td>.43</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>35. Emotional stability/stress</td>
<td>-.07</td>
<td>-.03</td>
<td>.31</td>
<td>.61</td>
<td>.25</td>
<td>.71</td>
<td>.15</td>
<td>-.26</td>
<td>-.46</td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>stress tolerance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>36. Interpersonal skill</td>
<td>.01</td>
<td>.03</td>
<td>.29</td>
<td>.62</td>
<td>.18</td>
<td>.73</td>
<td>.11</td>
<td>-.19</td>
<td>-.35</td>
<td>.82</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations Among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>37</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>Question variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>37. Judgment/common sense</td>
<td>.06</td>
<td>-.06</td>
<td>.24**</td>
<td>.57**</td>
<td>.41**</td>
<td>.57**</td>
<td>.22**</td>
<td>-.26**</td>
<td>-.51**</td>
<td>.65**</td>
<td>.51**</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>38. Responsibility</td>
<td>-.06</td>
<td>.09</td>
<td>-.09</td>
<td>-.10</td>
<td>-.03</td>
<td>-.41**</td>
<td>.29**</td>
<td>.11</td>
<td>.06</td>
<td>-.18*</td>
<td>-.26**</td>
<td>-.27**</td>
<td>1.0</td>
</tr>
<tr>
<td>39. Personal integrity</td>
<td>.00</td>
<td>.00</td>
<td>-.05</td>
<td>.03</td>
<td>.01</td>
<td>-.19</td>
<td>.07</td>
<td>.06</td>
<td>-.09</td>
<td>-.04</td>
<td>-.09</td>
<td>.43**</td>
<td>.63**</td>
</tr>
<tr>
<td>40. Standardization</td>
<td>-.09</td>
<td>-.14</td>
<td>.11</td>
<td>.15</td>
<td>-.44</td>
<td>.17*</td>
<td>.06</td>
<td>-.07</td>
<td>-.04</td>
<td>.28**</td>
<td>.29**</td>
<td>.01</td>
<td>-.18*</td>
</tr>
<tr>
<td>Diagnosticity variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. Typical diagnosticity</td>
<td>.09</td>
<td>.33**</td>
<td>.42**</td>
<td>.08</td>
<td>.25**</td>
<td>.14</td>
<td>-.11</td>
<td>.28**</td>
<td>.35**</td>
<td>.36**</td>
<td>.21**</td>
<td>.00</td>
<td></td>
</tr>
<tr>
<td>42. Attitude/motivation</td>
<td>.12</td>
<td>.02</td>
<td>.25**</td>
<td>.34**</td>
<td>.16*</td>
<td>.16*</td>
<td>.35**</td>
<td>-.13</td>
<td>-.36**</td>
<td>.29**</td>
<td>.24**</td>
<td>.21**</td>
<td>.00</td>
</tr>
<tr>
<td>43. Learning ability</td>
<td>-.06</td>
<td>.04</td>
<td>-.04</td>
<td>-.24**</td>
<td>.00</td>
<td>-.29**</td>
<td>-.13</td>
<td>.34**</td>
<td>.17*</td>
<td>-.29**</td>
<td>-.29**</td>
<td>-.29**</td>
<td>.03</td>
</tr>
<tr>
<td>44. Physical ability</td>
<td>-.08</td>
<td>.04</td>
<td>-.18*</td>
<td>-.45**</td>
<td>-.29**</td>
<td>-.45**</td>
<td>-.09</td>
<td>.25**</td>
<td>.65**</td>
<td>-.44**</td>
<td>-.39**</td>
<td>-.49**</td>
<td>.00</td>
</tr>
<tr>
<td>45. Emotional stability/</td>
<td>.12</td>
<td>-.07</td>
<td>.26**</td>
<td>.55**</td>
<td>.13</td>
<td>.49**</td>
<td>.09</td>
<td>-.25**</td>
<td>-.41**</td>
<td>.52**</td>
<td>.51**</td>
<td>.39**</td>
<td>-.19*</td>
</tr>
<tr>
<td>stress tolerance</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>46. Interpersonal skill</td>
<td>.06</td>
<td>-.01</td>
<td>.34**</td>
<td>.53**</td>
<td>.22**</td>
<td>.49**</td>
<td>.14</td>
<td>-.18*</td>
<td>-.42**</td>
<td>.49**</td>
<td>.56**</td>
<td>.38**</td>
<td>-.19*</td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations Among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>26</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>33</th>
<th>34</th>
<th>35</th>
<th>36</th>
<th>37</th>
<th>38</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosticity variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>47. Judgment/common sense</td>
<td>.00</td>
<td>-.01</td>
<td>.24**</td>
<td>.45**</td>
<td>.20**</td>
<td>.41**</td>
<td>.09</td>
<td>-.19*</td>
<td>-.36**</td>
<td>.45**</td>
<td>.37**</td>
<td>.58**</td>
<td>.05</td>
</tr>
<tr>
<td>48. Responsibility</td>
<td>-.03</td>
<td>.09</td>
<td>.10</td>
<td>.03</td>
<td>.02</td>
<td>-.29**</td>
<td>.25**</td>
<td>-.01</td>
<td>-.12</td>
<td>-.10</td>
<td>-.15</td>
<td>.16*</td>
<td>.68**</td>
</tr>
<tr>
<td>49. Personal integrity</td>
<td>.02</td>
<td>.04</td>
<td>.09</td>
<td>.14</td>
<td>.09</td>
<td>-.02</td>
<td>.01</td>
<td>-.11</td>
<td>-.24**</td>
<td>.16</td>
<td>.01</td>
<td>.35**</td>
<td>.36**</td>
</tr>
<tr>
<td>50. Interview judgment- specific diagnosticity</td>
<td>.05</td>
<td>.02</td>
<td>.31**</td>
<td>.45**</td>
<td>.17*</td>
<td>.25</td>
<td>.23**</td>
<td>-.16*</td>
<td>-.38**</td>
<td>.39**</td>
<td>.33**</td>
<td>.45**</td>
<td>.19*</td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations Among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>39</th>
<th>40</th>
<th>41</th>
<th>42</th>
<th>43</th>
<th>44</th>
<th>45</th>
<th>46</th>
<th>47</th>
<th>48</th>
<th>49</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Question variables:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>39. Personal integrity</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>40. Standardization</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.03</td>
<td></td>
<td></td>
<td>1.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Diagnosticity variables:</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>41. Typical diagnosticity</td>
<td></td>
<td>-.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>42. Attitude/motivation</td>
<td></td>
<td>-.07</td>
<td>.08</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>.52**</td>
</tr>
<tr>
<td>43. Learning ability</td>
<td></td>
<td>-.13</td>
<td>-.19</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>44. Physical ability</td>
<td></td>
<td>-.20**</td>
<td>-.14</td>
<td>-.25**</td>
<td>-.27**</td>
<td>-.41**</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>45. Emotional stability/stress tolerance</td>
<td></td>
<td>-.02</td>
<td>-.24**</td>
<td>-.52**</td>
<td>-.23**</td>
<td>-.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>-.37**</td>
<td>1.0</td>
</tr>
<tr>
<td>46. Interpersonal skill</td>
<td></td>
<td>-.06</td>
<td>-.12</td>
<td>-.50**</td>
<td>-.26**</td>
<td>-.01</td>
<td>-.31**</td>
<td>-.70**</td>
<td></td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>47. Judgment/common sense</td>
<td></td>
<td>.15</td>
<td>.06</td>
<td>.34**</td>
<td>.17*</td>
<td>.02</td>
<td>-.27**</td>
<td>-.51**</td>
<td>-.53**</td>
<td></td>
<td></td>
<td>1.0</td>
</tr>
<tr>
<td>48. Responsibility</td>
<td></td>
<td>.38**</td>
<td>-.05</td>
<td>.38**</td>
<td>-.26**</td>
<td>-.19*</td>
<td>.01</td>
<td>.00</td>
<td>.01</td>
<td></td>
<td></td>
<td>.23**</td>
</tr>
</tbody>
</table>
Appendix C (continued)

**Intercorrelations among Study Variables**

**Aggregate Interviewer Level**

<table>
<thead>
<tr>
<th>Variable</th>
<th>39</th>
<th>40</th>
<th>41</th>
<th>42</th>
<th>43</th>
<th>44</th>
<th>45</th>
<th>46</th>
<th>47</th>
<th>48</th>
<th>49</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diagnosticity variables:</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>49. Personal integrity</td>
<td>.61**</td>
<td>.03</td>
<td>.30**</td>
<td>.17*</td>
<td>.11</td>
<td>-.22**</td>
<td>.22**</td>
<td>.17*</td>
<td>.33**</td>
<td>.55**</td>
<td>1.0</td>
<td></td>
</tr>
<tr>
<td>50. Interview judgment-specific diagnosticity</td>
<td>.18*</td>
<td>.09</td>
<td>.64**</td>
<td>.50**</td>
<td>.24**</td>
<td>-.13</td>
<td>.67**</td>
<td>.67**</td>
<td>.74**</td>
<td>.56**</td>
<td>.56**</td>
<td>1.0</td>
</tr>
</tbody>
</table>

* \( p < .05 \)  ** \( p < .01 \).

a directedness was reverse-scored
Appendix D

Hierarchical Regression of Performance Appraisals on Interview Score
Appendix D

Hierarchical Regression of Performance Appraisal I on Interview Score

Aggregate Interviewer level (N = 149)

Whole model: F (7, 141) = .77, R^2 = .03, n.s.

<table>
<thead>
<tr>
<th>Source/Step</th>
<th>b</th>
<th>F</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test Scores</td>
<td>-.01</td>
<td>.65</td>
<td>.42</td>
</tr>
<tr>
<td>2. Interview score</td>
<td>-.07</td>
<td>.87</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>-10.25</td>
<td>1.86</td>
<td>.17</td>
</tr>
<tr>
<td>Interviewer vector 1</td>
<td>-8.9</td>
<td>1.55</td>
<td>.21</td>
</tr>
<tr>
<td>Interviewer vector 2</td>
<td>.17</td>
<td>2.23</td>
<td>.14</td>
</tr>
<tr>
<td>Interaction 1</td>
<td>.14</td>
<td>1.7</td>
<td>.19</td>
</tr>
<tr>
<td>Interaction 2</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note. b = regression weight after all variables have been entered into the equation*
Hierarchical Regression of Performance Appraisal II on Interview Score

Aggregate Interviewer level (N = 149)

Whole model: F (7, 141) = .46, R² = .01, n.s.

<table>
<thead>
<tr>
<th>Source/Step</th>
<th>b</th>
<th>F</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test Scores</td>
<td>-.01</td>
<td>.58</td>
<td>.45</td>
</tr>
<tr>
<td>2. Interview score</td>
<td>-.12</td>
<td>.95</td>
<td>.33</td>
</tr>
<tr>
<td>Interview vector 1</td>
<td>-13.74</td>
<td>1.08</td>
<td>.29</td>
</tr>
<tr>
<td>Interview vector 2</td>
<td>-3.92</td>
<td>.09</td>
<td>.77</td>
</tr>
<tr>
<td>Interaction 1</td>
<td>.22</td>
<td>1.22</td>
<td>.27</td>
</tr>
<tr>
<td>Interaction 2</td>
<td>.06</td>
<td>.10</td>
<td>.75</td>
</tr>
</tbody>
</table>

*Note.* b = regression weight after all variables have been entered into the equation.
Appendix D, cont.

Hierarchical Regression of Turnover Behavior on Interview Score

Aggregate Interviewer level (N = 149)

Whole model: F (7, 141) = .41, R^2 = .02, n.s.

<table>
<thead>
<tr>
<th>Source/Step</th>
<th>b</th>
<th>F</th>
<th>Significance level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Test Scores</td>
<td>.00</td>
<td>.24</td>
<td>.63</td>
</tr>
<tr>
<td>2. Interview score</td>
<td>.00</td>
<td>.10</td>
<td>.75</td>
</tr>
<tr>
<td>Interviewer vector 1</td>
<td>1.0</td>
<td>1.9</td>
<td>.17</td>
</tr>
<tr>
<td>Interviewer vector 2</td>
<td>.09</td>
<td>.02</td>
<td>.89</td>
</tr>
<tr>
<td>Interaction 1</td>
<td>-.02</td>
<td>1.96</td>
<td>.16</td>
</tr>
<tr>
<td>Interaction 2</td>
<td>-.01</td>
<td>.03</td>
<td>.87</td>
</tr>
</tbody>
</table>

Note. b = regression weight after all variables have been entered into the equation.