Journal of Applied Psychology
2001, Vol. 86, No. 5, 897-913

0021-9010/01/$5.00 DOI: 10.1037//0021-9010.86.5.897

Identification and Meta-Analytic Assessment of Psychological Constructs Measured in Employment Interviews

Allen I. Huffcutt
Bradley University

James M. Conway
Central Connecticut State University

Philip L. Roth
Clemson University

Nancy J. Stone
Creighton University

There has been a growing interest in understanding what constructs are assessed in the employment interview and the properties of those assessments. To address these issues, the authors developed a comprehensive taxonomy of 7 types of constructs that the interview could assess. Analysis of 338 ratings from 47 actual interview studies indicated that basic personality and applied social skills were the most frequently rated constructs in this taxonomy, followed by mental capability and job knowledge and skills. Further analysis suggested that high- and low-structure interviews tend to focus on different constructs. Taking both frequency and validity results into consideration, the findings suggest that at least part of the reason why structured interviews tend to have higher validity is because they focus more on constructs that have a stronger relationship with job performance. Limitations and directions for future research are discussed.

Much of the employment interview research published in the past 10–15 years has focused on interview validity. There have been a number of primary studies (e.g., Campion, Campion, & Hudson, 1994; Johnson, 1991; Pulakos & Schmitt, 1995; Walters, Miller, & Ree, 1993), several meta-analyses of interview validity (e.g., Huffcutt & Arthur, 1994; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Schmidt & Rader, 1999; Wiesner & Cronshaw, 1988; Wright, Lichtenfels, & Pursell, 1989), and a meta-analysis of interview reliability (Conway, Jako, & Goodman, 1995). These works collectively suggest that interviews can predict performance on the job.

Although a great deal is now known about interview reliability and validity, much less is known about the constructs captured by employment interviews (Schmidt & Rader, 1999). A number of possible constructs have been suggested in the interview literature, including cognitive ability (Campion, Pursell, & Brown, 1988), motivation (Ulrich & Trumbo, 1965), social skills (Arvey & Campion, 1982), and person–organization fit (Harris, 1999). However, the extent to which most of these constructs are actually assessed in interviews is unclear. The only construct that has been investigated empirically to any real extent is cognitive ability, and meta-analytic research suggests that, on average, it represents less than 20% of the variance in interview ratings (Huffcutt, Roth, & McDaniel, 1996).

Allen I. Huffcutt, Department of Psychology, Bradley University; James M. Conway, Department of Psychology, Central Connecticut State University; Philip L. Roth, Department of Management, Clemson University; Nancy J. Stone, Department of Psychology, Creighton University.

Correspondence concerning this article should be addressed to Allen I. Huffcutt, Department of Psychology, Bradley University, Peoria, Illinois 61625. Electronic mail may be sent to huffcutt@bradley.edu.

The lack of research on the constructs assessed in employment interviews is not surprising. Unlike many other psychological tests, such as those for personality and mental ability, interviews are not designed to assess specific constructs (Bobko, Roth, & Potosky, 1999; Campion, Palmer, & Campion, 1997). Rather, they are individually designed to assess work-related characteristics for a given position, and various constructs may be embedded in or associated with those work-related characteristics. Given the diversity of jobs for which interviews are developed, it would not be surprising to see considerable variation across interviews in both the number and the type of constructs assessed (Klimoski, 1993).

Understanding the constructs involved in employment interviews is potentially important. Significant advances have been made in interviewing methodology, including the situational interview (Latham, Saari, Pursell, & Campion, 1980) and behavior description interview (Janz, 1982) formats, but many continued advances are likely to come from analysis of what constructs interviews measure, not from methodological variants. In particular, analysis of constructs may provide greater insight into why formats such as the situational interview predict performance and may allow interviews to be optimally designed to achieve specific outcomes such as high incremental validity and minimal impact on protected groups.

It is our belief that four fundamental issues (or steps) are involved in the important process of understanding the constructs captured by employment interviews. First, a taxonomy of possible constructs that could be assessed in employment interviews should be constructed. Such a taxonomy would provide a common and systematic framework for identifying and classifying interview constructs. Second, the constructs that interviewers attempt to assess should be identified, and information on the relative frequency of those attempts should be compiled. Such an analysis would indicate which constructs in the taxonomy are actually rated

in employment interviews and, perhaps more important, which constructs are the most commonly rated. Third, the degree to which the ratings for these constructs reflect the intended characteristic should be evaluated. Some ratings may represent a more accurate measurement of the intended construct than other ratings, possibly due to the nature of the construct itself (e.g., extroversion vs. creativity) or to differential influence from general factors such as mental ability, personality, job experience, or the type of questions (see Huffcutt, Weekley, Wiesner, DeGroot, & Jones, in press). Fourth, the general properties of these ratings should be considered, including validity, incremental validity, and impact on protected groups. Some construct ratings may be a stronger predictor of job performance than other ratings and may differ in regard to incremental validity as well as the level of group differences.

The purpose of this investigation was to begin the process of identifying and evaluating the constructs assessed in employment interviews. The task of completing all four of the steps outlined above is substantial and cannot be accomplished in a single study; rather, it probably will take decades to complete these analyses. Our intent was to thoroughly address the first two steps described above and then to initiate work on the third and fourth steps. To address the first two steps, we constructed a comprehensive taxonomy of possible interview constructs based on current psychological literature and then used this taxonomy to identify constructs commonly rated in interview studies. To initiate work on the third and fourth steps, we accumulated and meta-analytically summarized the empirical data that are currently available for ratings of the various constructs. As evident later in this article, data on the degree to which construct ratings reflect the intended characteristics tend to be very sparse. In fact, there was only one construct (general mental ability) for which we found enough data to analyze. Fortunately, there was a much more reasonable amount of data on the properties of the construct ratings, specifically for validity and group differences, although the number of data points was small for some of the constructs.

Development of a Taxonomy of Possible Interview Constructs

As a first step in understanding employment interview constructs, we created a taxonomy of possible constructs. To help develop this taxonomy, we examined the psychological literature in areas such as cognitive, social, personality, and industrial–organizational psychology and identified constructs relevant to the employment interview. In particular, we looked for established constructs that have a long history in applied psychology, including direct applications in personnel selection and placement (e.g., mental ability, personality, interests and preferences); for other constructs that could be measured in an interview (e.g., organizational fit, social skills); and for characteristics in which many employers are routinely interested (e.g., job knowledge). This search required judgment on our part and was guided by our background and experience with interview studies and the interview literature. Our collection of relevant constructs fell into seven major categories. The categories, the constructs within each category, and even subfactors for some of the constructs are presented next.

The first category, and probably the best place to start a discussion of established psychological constructs, is mental capability. Performing mental operations is an important part of most jobs (Hunter & Hunter, 1984), and many employers are no doubt interested in how well applicants can do these operations.

The first individual construct that we found in this category was "general intelligence," also called "general mental ability" or "general cognitive ability" (see Herrnstein & Murray, 1994; Schmidt & Hunter, 1998), and it reflects the overall ability to learn and process information. Earlier in the 20th century, Spearman (1927) proposed the idea that a central processing capability, which he called *g*, underlies much of our common mental functioning (see also Thurstone, 1938). The existence of *g* is empirically supported by the finding that primary areas of mental functioning such as math, verbal, spatial, perceptual, and mechanical skills correlate moderately with one another and form a single, psychometrically meaningful factor. Not surprisingly, measures of general intelligence have been found to be related to performance across a wide range of jobs (Hunter & Hunter, 1984; Schmidt & Hunter, 1998).

Although a variety of psychological measures are available to assess general intelligence (or one of its primary areas), a number of employers still appear to use the employment interview for its assessment. Spychalski (1994), for example, developed an interview for corrections officers that included assessment of the ability to learn procedures. In a similar manner, Huse (1962) evaluated managerial applicants on intellectual capacity in his interview. Although ability tests may be superior psychometrically, many employers continue to assess mental traits in interviews for a variety of reasons, including logistical considerations, habit, legal concerns, and even a basic belief in the accuracy of human judgment (see Dipboye, 1992).

The second construct that we found under mental capability can be called "applied mental skills," and it reflects the application of mental ability to solve organizational problems and address various organizational issues. Specific areas of this construct include judgment, decision making, problem solving, and planning. The roots of this construct go back to the first half of the 20th century, including the use of a practical judgment subtest in the U.S. Army Alpha Examination during World War I (Taylor, 1949; Terman, 1918) and Cardall's (1942) development of an industrial screening test to measure practical judgment.[1] Prominent ability tests such as the Wechsler and the Stanford-Binet continue to include items in which a specific situation is described (e.g., fire in a movie theater) and test takers must indicate what they would do in that situation (see Murphy & Davidshofer, 1998), and the concept of applying mental capability to real-world contexts continues to be discussed in the literature (Sternberg & Kaufman, 1998).

Given that most jobs have at least some problems to solve and some decisions to make, it is not surprising that a number of employers appear to use the interview for evaluation of applied mental skills. Examples of such applications include Pulakos and Schmitt's (1995) assessment of the ability to evaluate information and to make decisions from that information in their interview for investigative agents and Berkley's (1984) assessment of the ability to solve problems in his interview for a corrections position.

[1] Cardall (1942) defined practical judgment as "recognition of possible alternatives of action and the ability to select the best" (p. 1).

The third construct that we found under mental capability was "creativity," and it reflects the capability to generate innovative ideas and solutions. Creativity is different from the traditional conceptualization of mental capability in applied psychology (i.e., learning, retaining, and processing information) because it requires flexibility of thought, originality, and the ability to see beyond current structures and operations (Cohen & Swerdlik, 1999). Nonetheless, many believe it still constitutes a mental operation and thus is best placed in the mental capability category.[2] Given the sparsity of paper-and-pencil tests of creativity, it is not surprising that at least some employers use the interview for its assessment. Examples of such assessment include Hoffman and Holden's (1993) evaluation of innovation in their study of a management position in a gas company and Chapman and Rowe's (1998) evaluation of creativity in their study of cooperative education workers.

The second category of psychological constructs relevant to the interview is knowledge and skills. Rather than focusing on the capability to process information, these constructs revolve around information already stored in long-term memory and include both declarative (i.e., terms, values, names, and dates) and procedural (i.e., actions, skills, and operations) components (see Winograd, 1975). Not surprisingly, knowledge and skills appear to have at least some relationship with general intelligence in that people with higher intelligence often can retain more information (Borman, White, Pulakos, & Oppler, 1991), although other factors such as exposure, experience, and interest can influence knowledge retention as well.

The usefulness of knowledge constructs is supported by evidence that direct measures of job knowledge and skills are predictive of job performance (Hunter & Hunter, 1984; Schmidt & Hunter, 1992). Some employers use background credentials (e.g., education, training, or experience) as indirect measures of knowledge and skills, although such credentials do not have as strong of a record for predicting performance (Hunter & Hunter, 1984) and are often used more in the initial screening process (Riggio, 1996). Assessment of knowledge and skills has been included in at least some interviews, including Adorno, Binning, Srinivasagam, and Williams's (1997) evaluation of "technical knowledge" in their interview study of assembly workers and Landy's (1976) evaluation of "experience" in his interview study of police officers.

The third category of selected psychological constructs is basic personality tendencies. Commonly called "traits," these tendencies reflect long-term predispositions to act in certain ways. Although research on personality has been ongoing for most of the past century (e.g., Allport, 1937), there appears to be a growing consensus that there are five main personality dimensions. Collectively known as the "Big Five," these dimensions include Extroversion, Conscientiousness, Agreeableness, Openness to Experience, and Emotional Stability. Descriptions of each of these traits is as follows (see Costa & McCrae, 1992; Digman, 1990; Oliver, 1989). Extroversion is the basic tendency to be socially active and includes elements of both basic sociability (high need for and enjoyment of social activity) and power-related tendencies (assertiveness and dominance). People high on this trait are typically described as warm, gregarious, energetic, assertive, dominant, driven, and competitive. Conscientiousness reflects the drive to accomplish assigned tasks and duties to the best of one's ability and to do so within the confines of established procedures and protocols. People high on this trait are often described as responsible, dependable, competent, punctual, deliberate, and respectful of authority. Agreeableness is the basic desire to be liked by and to fit in with other people. People high on this trait are commonly described as likable, friendly, warm, caring, polite, tactful, and helpful. Openness to Experience reflects the tendency to be open to new ideas and flexible in one's thinking. People high on this trait are usually described as open, curious, flexible, and imaginative. Finally, Emotional Stability reflects the regulation and management of one's emotions, including doing so in stressful conditions. People high on this trait are typically described as calm, poised, composed, confident, and stable. There is a growing body of literature that suggests that personality tendencies can predict job performance (Barrick & Mount, 1991; Mount & Barrick, 1995; Tett, Jackson, & Mitchell, 1991).

This category is important to study in relation to the interview because, in addition to mental capability and accumulated knowledge, many employers also seem interested in how potential employees would typically act on the job. Employers appear to use interviews, similar to use of ability tests, for assessment of personality traits for reasons such as logistical considerations and habit even though established and psychometrically superior measures are available. Personality has been evaluated in a number of interview studies, including Zedeck, Tziner, and Middlestadt's (1983) assessment of "self-confidence" in their interview for female military officers; Dipboye, Gaugler, Hayes, and Parker's (1992) assessment of "responsibility" in their interview for an entry-level corrections position; and Chapman and Rowe's (1998) evaluation of "friendliness" in their interview for cooperative education workers.

The fourth category of relevant psychological constructs is applied social skills, a category related to basic personality tendencies. This category reflects the ability to function effectively in social situations, the skills for which may be influenced both by underlying personality structure and by acquired competencies. Historically, the roots of this category go back to the first half of the 20th century. In 1920, Thorndike proposed the concept of "social intelligence," a concept that he defined as the ability to understand others and to act wisely in human relations (see also Thorndike, 1921). More recent derivations of his concept include Gardner's (1983) concept of "personal intelligences" (i.e., interpersonal and intrapersonal skills) and Sternberg, Conway, Ketron, and Bernstein's (1981) concept of "social competence." Compared with basic personality, research and literature relating to applied social skills are much more sparse and underdeveloped. In fact, we were able to find only one established measure relating to these constructs—Stevens and Campion's (1999) test for teamwork skills—and that measure was designed specifically for selection in autonomous and semiautonomous work teams. It is possible that their test may be predictive of performance in more individual-based jobs as well, although that would have to be established in future research.

Given the importance of social components to many jobs and the lack of established measures, it is not surprising that many employers use the interview for assessment of skills in social situations. We were able to identify four specific social skills that

[2] We thank one of the reviewers for making this suggestion.

employers have evaluated in an interview: oral communication skills, interpersonal skills, leadership, and persuasiveness. Oral communication reflects the ability to express (and receive) ideas and information clearly, accurately, and convincingly, and it has been assessed in studies such as Robertson, Gratton, and Rout's (1990) interview for financial services representatives in England. Interpersonal skills refer to the ability to relate to, understand, work with, and develop rapport with others, and they have been assessed in studies such as Dougherty, Ebert, and Callender's (1986) interview for entry-level positions at a large energy corporation. Leadership is the ability to direct and motivate others, and it has been assessed in studies such as Wiesner, Latham, Bradley, and Okros's (1992) interview for Canadian Naval officers. Finally, persuasiveness is the ability to change other people's opinion in important matters, and it has been assessed in studies such as Hoffman and Holden's (1993) interview for a management position in a gas company.

The fifth category of psychological constructs is interests and preferences, and it represents an inclination toward certain areas or activities. Items in this category would include a preference for a particular type of work or profession, a preference for a specific company or geographical area, involvement in related hobbies, and an interest in certain topics or subjects. Research on interests and preferences goes back to the turn of the last century, including G. Stanley Hall's development of a questionnaire to assess recreational interests in 1907 (see Cohen & Swerdlik, 1999). In a similar manner, the Strong–Campbell Interest Inventory, a popular instrument today, was first published in 1928 (see Cohen & Swerdlik, 1999).

Although standard measures of interests and preferences do not have a strong record for predicting job performance (Hunter & Hunter, 1984), some employers still appear to include characteristics related to interests and preferences in their interviews. Johnson (1991), for example, evaluated "interest in medicine" in his interview for medical residents, and Roth and Campion (1992) assessed "interest" in the position in their interview for refinery technicians.

More recently, "organizational fit" has emerged as a unique and potentially important concept related to organizations, and it is our sixth category. The idea behind organizational fit is that each organization takes on its own unique culture or climate, defined by characteristics such as values, goals, norms, and attitudes (Cable & Judge, 1997; Kristof, 1996). The closer the values and attitudes of an individual correspond to those of the organization, the better the fit is between them. Measures of organizational fit have been shown to correlate with a number of criteria, including work attitude, organizational tenure, prosocial behaviors, and work performance (Kristof, 1996; see also Rynes & Gerhart, 1990).

At least some employers appear to use the interview for evaluation of organizational fit. Such instances include DeGroot's (1997) assessment of "appreciation of diversity" in his interview study of first-level managers and Bradley, Bernthal, and Thomas's (1998) assessment of "quality orientation" in their interview study of process operators. Some have even argued that the ability to assess organizational fit is a key reason for the continued popularity of the interview (Karren & Graves, 1994).

Finally, we felt it necessary to include a "physical attributes" category of constructs because some employers appear to use the interview to assess physical characteristics. Some of the physical characteristics that employers assess are general in nature, such as Raza and Carpenter's (1987) evaluation of attractiveness. Other physical characteristics that employers assess are more job-related, such as Grove's (1981) evaluation of stamina and agility. Although physical attributes are not really psychological constructs in a true sense, we included them in our construct framework to be thorough because they are rated in at least some interviews.

Collectively, the aforementioned categories and constructs provided a workable framework from which to begin the process of identifying the constructs rated in employment interviews. As we noted earlier, these constructs were selectively chosen and represent only a fraction of the constructs available in the psychological literature. There are other popular constructs that we did not use, such as Festinger's (1957) concept of cognitive dissonance and Rotter's (1966) concept of locus of control. Moreover, there are alternative ways to look at some of these constructs, such as Gardner's (1983) theory that there are seven largely independent types of intelligence. Regardless, it is our opinion that the constructs presented above are the ones most relevant for the task at hand, namely, identifying the constructs captured in employment interview evaluations.

Degree of Structure

The interview process can be influenced by a number of factors, including use of an interview panel, availability of background information, type of questions asked, and type of job analysis (see Campion et al., 1997; Dipboye, 1992). Among these factors, the degree of structure is generally considered to be the most important, not only because of its effects on the interview process itself but also because of its impact on reliability and validity (Campion et al., 1997; Huffcutt & Arthur, 1994; Wiesner & Cronshaw, 1988).

There is a real possibility that structure could also influence the constructs that are captured in interviews. High-structure interviews differ from low-structure interviews in a number of ways, many of which could influence construct measurement. For instance, the dimensions (i.e., constructs) rated in structured interviews are more likely to be based on a job analysis than are the dimensions rated in unstructured interviews. Consequently, it would not be surprising to see measurement of more general, impression-based constructs in low-structure interviews (e.g., motivation, ability to think) and measurement of more specific, job-related constructs in high-structure interviews (e.g., job knowledge, ability to solve problems or come up with creative solutions). In addition to tapping different constructs, unstructured interviews tend to have considerably lower reliability (Conway et al., 1995), so their ratings may represent less accurate measurement of the intended constructs.

Accordingly, when analyzing the frequency and properties (e.g., validity) of interview constructs, we broke as many of our analyses down by structure as possible. Analyzing structural differences was not always easy given limitations in sample size for some of the constructs. Nonetheless, we felt it was important to include analyses of structure given its centrality to the interview process and the outcomes of this process.

Method

Search for Primary Data

We conducted an extensive search to locate usable employment interview studies for our investigation. Interview studies included in previous interview meta-analyses were reviewed (Conway et al., 1995; Huffcutt & Arthur, 1994; McDaniel et al., 1994; Wiesner & Cronshaw, 1988), more recent issues (since 1994) of the *Journal of Applied Psychology* and *Personnel Psychology* were examined, the databases *PsycLIT* and *ABI-INFORM* were searched, and recent conference programs were checked. Supplemental inquiries were also made to prominent researchers in the interview area to obtain any additional studies not included in the aforementioned sources. Finally, authors of several recent studies in which the desired information was not reported were contacted to see if that information was available.

Four main criteria guided our search. First, a study had to list the specific dimensions (i.e., characteristics) assessed in that interview. Studies not listing the specific dimensions were excluded, such as that of Orpen (1985), who developed a structured interview around six behaviorally defined dimensions but did not indicate what those dimensions were.

Second, a study had to involve a real position in business and industry (either applicants or incumbents) or a position in a professional training program that included duties in a real setting (e.g., medical school resident). Here, we excluded Walsh's (1975) study because the subjects were college football players and also several other studies involving high school students being interviewed as part of the general admissions process to college.

Third, a study had to represent a typical interview. We dropped Hilton, Bolin, Parker, Taylor, and Walker's (1955) study because the interview ratings were actually made by psychologists who read notes made by the original interviewers. (The original interviewers did not make any ratings.) We also dropped the two older studies from Martin (1972) because the main purpose of those interviews was to interpret and augment psychological test results.

Fourth, a study had to provide at least some supplemental information for the individual dimensions in the interview. Consistent with our second objective to begin exploring and establishing relationships for the constructs, we looked for any possible data relating to their properties. This exploratory information included correlations with job performance (i.e., validity coefficients), correlations with psychological tests, group differences (race or sex), and interrater reliability.

In total, we were able to locate 47 employment interview studies that met the four aforementioned search criteria.[3] Citations for these studies are included in the general list of references, identified by an asterisk. They included a wide range of job types, interview designs, companies, and products. A total of 338 assessment characteristics (i.e., dimensions) were identified from these studies. The mean number of dimensions per study was 7.2, with an actual range from 3 to 18.

Mapping of Interview Constructs

Using the framework of psychological constructs presented earlier, Allen I. Huffcutt, James M. Conway, and Philip L. Roth attempted to code each of the 338 interview dimensions as to the construct that it reflected (first the category and then the individual construct). We based our coding on both the label of the dimensions and any definitions provided in the studies. Our general strategy was to independently code the dimensions and then to discuss and resolve any differences.

To assess the consistency of the initial coding process, we compiled data on the level of agreement among the three of us. Specifically, we recorded each agreement as a "+" and each disagreement as a "−." In total, Allen I. Huffcutt had an 86% agreement rate with James M. Conway and an 85% agreement rate with Philip L. Roth. James M. Conway, in turn, had an 82% agreement rate with Philip L. Roth. These results suggest that the initial process of identifying constructs was sufficiently consistent.

To further ensure that all coding was as accurate and free from bias as possible, we had a fourth individual independently code all interview dimensions. This person was an experienced applied psychologist who kept current with the general selection literature but was not familiar with our present study prior to being contacted. We gave her a copy of our construct framework and a list of the dimensions from the interview studies and then asked her to code the dimensions as to the construct they reflected. The subsequent agreement between the consensus codings by the first three authors and those of the independent rater was 88%, again suggesting that the construct codings were accurate and unlikely to be biased.

Recording of Exploratory Information

We recorded whatever supplemental information was available for each of the 338 interview dimensions in our data set. It is important to note that the level of analysis with this information was for each individual rating characteristic and not for the overall interview scores. As we expected, not every study reported all of the supplemental information for the individual dimensions being rated, which reduced the overall amount of information available.

For the validity coefficients, we recorded the uncorrected correlation between interviewer ratings on a given characteristic and evaluation of overall job performance by a supervisor or manager. For the ability correlations, we recorded the uncorrected correlation between interviewer ratings on a characteristic and scores on some type of mental ability test.

For racial group differences, we recorded the mean and the standard deviation of the ratings for a characteristic for White and Black subjects, respectively. From this information we computed a *d* score, which reflected the number of within-group standard deviations that the mean of the White subjects was different from the mean of the Black subjects (see Hunter & Schmidt, 1990). In a similar manner, for sex group differences we recorded the mean and the standard deviation of the ratings for male and female subjects, respectively, and computed a *d* score from this information. Here, the *d* score indicated the number of standard deviations that the mean of the male subjects was different from the mean of the female subjects. The *d* values for both race and sex were computed such that a positive value indicated that the unprotected group (Whites and male subjects, respectively) received higher ratings, on average, than the protected group (Blacks and female subjects, respectively) whereas a negative value indicated that the protected group received higher ratings.

For analysis of construct validity, we recorded the uncorrected correlation between interviewer ratings and scores on an established psychological measure of the same construct. We found a fair amount of this information for mental ability, but corresponding correlations for personality, applied social skills, and the other constructs in the framework were minimal. Therefore, we were forced to limit this analysis to mental ability and leave analysis of the remaining constructs for future research.

Data on interrater reliability were similarly sparse. There were some data on interrater reliability for overall interview scores but considerably less data for the individual characteristics assessed. Moreover, what data there were at the characteristic level varied in format, with some representing the reliability of multiple ratings from the same interview and others representing the reliability of ratings from different interviews. Accordingly, we were forced to leave this issue for future research as well.

---

[3] There are actually only 45 studies identified in the reference list. The reason is that Johnson (1991) and U.S. Office of Personnel Management (1987) each had two usable studies.

Analysis of Exploratory Information and Artifact Corrections

For each construct for which data were available, we computed the mean of the exploratory statistics, namely, the mean validity, the mean cognitive ability correlation, the mean *d* value for race, and the mean *d* value for sex. These means were computed by giving each study coefficient equal weight. We decided to compute means in this manner out of concern that a handful of studies with a relatively large sample size could dominate the results for a number of constructs. This was particularly important for a study such as ours in which a number of cells had comparatively few studies, because a single large sample-size study would be extremely influential. For example, if we used sample weighting, the 2 largest studies for interpersonal skills (see Table 3, which is presented later) would count almost as much as the other 17 studies combined. Research in this area suggests that equal weighting provides more stable estimates than sample weighting when a large-sample study is present (Fuller & Hester, 1999; Osburn & Callender, 1992).

To provide information on the practical significance of our results, we computed a 90% confidence interval for each of our mean estimates.[4] Some concern has been noted in the literature about confidence interval formulas that are based on the assumption that the effect is consistent in the population and not moderated by features of the situation or the study methodology (see Erez, Bloom, & Wells, 1996; Hedges & Vevea, 1998). Accordingly, we used a formula offered by Osburn and Callender (1992), specifically their Equation 5, to form confidence intervals for the mean correlations between interview ratings and performance ratings and for the mean correlations between interview ratings and mental ability test scores. As Osburn and Callender noted, this formula is appropriate for both homogeneous and heterogeneous situations. We used a similar equation provided by Hunter and Schmidt (1990, p. 430) to form confidence intervals for mean race and sex effect sizes, and this equation is appropriate for both homogeneous and heterogeneous situations as well. In both cases, the confidence intervals were formed by taking the uncorrected mean estimate plus or minus 1.65 times the square root of the mean sampling variance.

Finally, as is typical in selection research, we corrected the mean observed validity estimates for the influence of various statistical artifacts. Given the sparsity of artifact data in the studies in our database, we used the artifact distribution approach outlined by Hunter and Schmidt (1990). Specifically, we corrected all mean validity estimates (a) for range restriction in the interview by using the average ratio of .74 found by Huffcutt and Arthur (1994) and (b) for measurement error in job performance evaluations by using the average interrater reliability value of .52 found by Viswesvaran et al. (1996). There appears to be some debate over the accuracy of the .52 value (Murphy & DeShon, 2000), but at the present time, we believe it to be the best estimate available.

In a similar manner, we corrected the mean observed correlations between interview construct ratings and mental ability test scores by using the artifact distribution approach. Here, we corrected all mean estimates for range restriction in the interview by using the .74 value noted above and for measurement error in the ability tests by using a value of .90 for their reliability (see Huffcutt et al., 1996). Because ability testing is typically done before selection decisions are made, the .74 value for range restriction in the interview may have resulted in a slight overcorrection to these mean effect sizes. No corrections for artifacts were made to the group differences *d* values for race and sex, although artifacts undoubtedly had at least some influence on their magnitude as well.

Identification and Analysis of Interview Structure

Huffcutt and Arthur (1994) established a framework whereby interviews can be classified along four distinct levels of structure. Because the number of coefficients available for some of the constructs was relatively low, we decided to simplify our structural classification by collapsing this framework into two more general levels. Specifically, we classified studies as high structure if a majority of the questions were specified beforehand and as low structure if the interviewers had fairly wide discretion in terms of choosing what topics to cover or at least what questions to ask. Then we recomputed and corrected the mean exploratory statistics as described above separately for high- and low-structure interviews.

Results

Frequency of Construct Measurement

Results of the mapping process between psychological constructs and interview dimensions are shown in Table 1. The data on the left show the number (and percentage) of dimensions in each of the major construct categories, and the data in the middle show the number of dimensions associated with each construct. Common labels for the interview dimensions are shown on the right.

As indicated in Table 1, the largest number of dimensions reflected basic personality tendencies (35%), followed by applied social skills (28%). Mental capability (16%) and knowledge and skills (10%) were the next most frequently rated constructs. Constructs in the remaining three categories were rated less frequently, including interests and preferences (4%), physical attributes (4%), and organizational fit (3%).

Differences in the frequency of construct measurement between low- and high-structure interviews are presented in Table 2. As shown, when expressed as percentages, there were noticeable differences in the frequency of measurement for a number of the constructs. For mental capability, general intelligence was assessed almost five times more often in low-structure interviews, whereas applied mental skills were assessed more than twice as often in high-structure interviews. Among knowledge constructs, direct knowledge and skills were assessed more than twice as often in high-structure interviews, whereas education and training as well as experience and general work history were assessed roughly three times as often in low-structure interviews.

For personality, Agreeableness was rated more than three times more often in low-structure interviews than in high-structure interviews, and Emotional Stability was rated more than twice as often. Openness to Experience was rated more frequently in high-structure interviews, although a ratio could not be estimated because it was not rated at all in the low-structure interviews. There did not appear to be substantial differences between low- and high-structure frequencies for Extroversion and Conscientiousness.[5]

For applied social skills, communication and interpersonal skills were rated approximately twice as often in high-structure inter-

[4] A variety of intervals have been used in the literature. Viswesvaran, Ones, and Schmidt (1996), for example, used 80% intervals because those intervals isolated points at which 10% of values would be higher and 10% would be lower. Wiesner and Cronshaw (1988) formed 95% confidence intervals, which isolated the outer 2.5% points. We chose to use 90% intervals because that seemed reasonable given the other intervals just described and because they isolated the outer 5% of values on either side.

[5] As shown in Table 2, there were 36 conscientiousness-type constructs rated in the high-structure studies even though there were only 28 high-structure studies in the data set. The reason why there were more conscientiousness coefficients than high-structure studies is that most of these studies included several conscientiousness-type dimensions. For example, the interviewers in Roth and Campion's (1992) study evaluated candidates on both reliability and dependability.

views, and leadership was rated more than three times as often. Persuading and negotiating appeared to be rated with about the same frequency in both low- and high-structure interviews. Among the remaining categories, organizational fit was assessed about one-and-one-half times more often in high-structure interviews, physical attributes (both general and job-related) were assessed more than four times as often in low-structure interviews, and the frequency of ratings was fairly close for interests and preferences.

Overall, these data suggest a potentially important effect, namely, a tendency for high- and low-structure interviews to emphasize somewhat different constructs. High-structure interviews appear to be focused more on applied mental skills, direct job knowledge, applied social skills (communication, interpersonal skills, and leadership), and organizational fit, whereas low-structure interviews appear to be focused more on general mental ability, background credentials (education, training, and experience), some aspects of personality (agreeableness and emotional stability), and physical attributes. In short, these results suggest that structure affects not only the conduct of the interview but also what constructs are rated. We next examine relevant validity evidence.

Validity of Construct Ratings

Mean correlations between ratings of the various constructs and overall performance evaluations are presented in Table 3. To enhance the stability and generalizability of the results, we did not attempt to interpret the mean value for any individual construct with fewer than four studies (see Viswesvaran et al., 1996).

For our overall analysis of all studies, the mean corrected validity coefficients ranged from .24 to .58 (with a mean of .36). We noted in the introduction that some construct ratings may be a stronger predictor of job performance than other ratings, and these results provide support for that idea. The highest mean corrected validities were observed for ratings of creativity (.58), agreeableness (.51), organizational fit (.49), leadership (.47), emotional stability (.47), job knowledge (.42), and interpersonal skills (.39). The lowest mean validities were observed for ratings of interests and preferences (.24), general intelligence (.24), communication skills (.26), and applied mental skills (.28).

Four things should be noted in regard to the aforementioned findings. First, some of these mean correlations were based on a fairly small number of studies (e.g., there were four studies for creativity). As a result, these mean estimates should be interpreted with some caution because they could include sampling error. Moreover, use of mean artifact distribution values for interview range restriction and performance unreliability could have resulted in over- or undercorrection of these estimates.

Second, these mean validity estimates included both low-structure and high-structure studies. Structured interviews generally have higher validity than unstructured interviews (Wiesner & Cronshaw, 1988), so mean estimates based on a higher proportion of structured studies would be expected to have higher mean validity. For example, five of the six studies in which job knowledge was rated were of high structure.

Third, these values reflect the mean validity across different types of jobs. It is possible that some constructs may be predictive of performance in some jobs but not in other jobs, whereas other constructs may be more universal predictors of performance. That constructs such as organizational fit, emotional stability, interpersonal skills, and creativity had high validity across various types of positions may suggest that they are somewhat universal predictors of performance. Alternatively, it may be that some of these high-validity constructs are easier to observe in interview situations, the "signs" versus "samples" distinction brought up earlier by Wernimont and Campbell (1968). Interpersonal skills and some elements of emotional stability (e.g., maturity, stress tolerance), for example, may be fairly salient during an interview.

Finally, these results suggest that ratings of creativity have higher validity than ratings of general intelligence, something that appears to be inconsistent with existing selection literature. There are at least two possible explanations. One is that the mean validity estimate for creativity was based on only four total studies and thus may have contained sampling error. The other is that the nature of the questions asked to evaluate creativity may have made it easier to evaluate than general intelligence. For example, some of the studies in which creativity was assessed used behavior description questions (e.g., Hoffman & Holden, 1993; Huffcutt et al., in press). Having candidates describe a past situation in which they came up with a creative solution to an organizational problem might be fairly easy to evaluate. In contrast, general intelligence is more abstract and general, which could make it more difficult to assess in an interview (unless the questions posed mathematical or verbal problems like those found on typical ability tests).

Results of the structure analyses are also presented in Table 3. As we expected, the mean validities for constructs rated in more structured interviews were higher overall than the mean validities for constructs rated in less structured interviews. In particular, the mean corrected validity across all constructs rated in low-structure interviews was .24, whereas the mean corrected validity across all constructs rated in high-structure interviews was .39. There were three constructs that had four or more studies for both low-structure and high-structure interviews, and they allowed a more direct comparison of validity. These constructs were applied mental skills, conscientiousness, and interpersonal skills. The mean validities for structured ratings were higher in all three cases, although the difference was somewhat smaller for interpersonal skills. This smaller difference for interpersonal skills could reflect sampling error, or it could suggest that interpersonal skills can be rated reasonably well in low-structure interviews.

Among constructs rated in high-structure interviews, there were some clear differences in terms of validity. Emotional stability and organizational fit had the highest mean corrected validities (.56 and .58, respectively), whereas interests and preferences and communication skills had the lowest corrected validities (.26 and .31, respectively). Differences among constructs rated in low structure were less pronounced, with interpersonal skills having the highest mean corrected validity (.31) and applied mental skills having the lowest corrected validity (.13). As we noted earlier, some of these estimates were based on a limited number of studies and should be viewed as tentative.

Correlation Between Construct Ratings and Mental Ability Test Scores

Mean correlations between interview construct ratings and mental ability test scores for all studies are shown in the first part of Table 4. As we noted earlier, all of the mean correlations in our study, including those for low and high structure, were computed

Table 1
Interview Constructs and Associated Dimension Labels

N_c	%	Category and construct	n	Common interview dimension label
55	16.3	*Mental capability*—ability to learn, organize, process, and evaluate information *Major theme*: assessing how well candidates can think		
		General intelligence	20	Intellectual capacity, intelligence, mental ability, ability to learn, learning the work, analytical ability, mental alertness, ability to think quickly, perceptiveness
		Verbal ability	1	Support for arguments
		Applied mental skills	28	Problem solving, problem assessment, judgment, decision making, critical thinking, planning, organizing
		Creativity and innovation	6	Creativity, creativeness, innovation
33	9.8	*Knowledge and skills*—accumulated knowledge, skills, and abilities *Major theme*: assessing what candidates know and what they can do		
		Job knowledge and skills	14	Knowledge, technical knowledge, job knowledge, product knowledge, use of tools, budgeting
		Education and training	6	Education, academic achievement, grades in school
		Experience and general work history	13	Experience, work history, exposure
118	34.9	*Basic personality tendencies*—predispositions to act in certain ways *Major theme*: assessing how candidates are likely to act in the workplace		
		Extroversion	21	Assertiveness, dominance, ability to control situations, drive, energy, decisiveness, ambition, positive outlook
		Conscientiousness	55	Dependability, responsibility, reliability, timeliness, sense of duty, need for achievement, motivation, willingness to work hard, initiative, persistence, time management, moral character, integrity, ethics, professionalism
		Agreeableness	10	Friendliness, likability, empathy, concern for others, attitude, general attitude
		Openness to experience	6	Adaptability, flexibility, openness to change
		Emotional stability	21	Emotional stability, stress tolerance, performance under stress, poise, social adjustment, self-control, coping, maturity, self-confidence, ego strength
		Other personality traits	5	Independence, self-reliance, self-understanding
94	27.8	*Applied social skills*—ability to function effectively in social situations *Major theme*: assessing how well candidates can deal with other people		
		Communication skills	26	Oral communication, communication skills, expression, ability to present ideas, conversation ability, voice and speech, listening
		Interpersonal skills	43	Interpersonal skills, interpersonal relations, social skills, social sensitivity, working with others, ability to relate to people, rapport, tact, ability to deal with people, adapting to people, teamwork, cooperation, team focus, team building
		Leadership	20	Leadership, vision, coaching, developing people, delegation, maintaining control, directing others, activating others, developing teamwork in others, building morale, discipline
		Persuading and negotiating	5	Persuasiveness, ability to negotiate
15	4.4	*Interests and preferences*—inclination toward certain areas or activities *Major theme*: assessing what candidates like to do		
		Occupational interests	13	Job interest, interest in position, investment, commitment to a career
		Subject and topic interests	1	Extracurricular activities
		Hobbies and general interests	1	Hobbies

Table 1 *(continued)*

N_c	%	Category and construct	n	Common interview dimension label
11	3.3	*Organizational fit*—compatibility of attitudes and beliefs with those of the organization		
		Major theme: assessing what things the candidates really believe in		
		Values and moral standards	11	Quality orientation, safety orientation, appreciation for diversity, acceptance of company mission, customer service, customer focus, belief in product value, pride in the organization
12	3.6	*Physical attributes*		
		General physical attributes	8	Health, appearance, attractiveness
		Job-related physical skills	4	Physical requirements, physical ability, stamina, agility

Note. A total of 338 dimensions were coded. N_c = the number of dimensions in each of the major construct categories; % = the percentage of dimensions in each of the major construct categories; n = the number of dimensions associated with each individual construct.

by equally weighting the respective correlations, and that practice was maintained for the ability correlations. Across construct categories, the mean corrected correlation with mental ability test scores was .17. The low magnitude of this mean correlation suggests that *g* does not saturate interview ratings at the construct level, a finding that is consistent with previous meta-analytic research involving total interview scores (Huffcutt et al., 1996).

There did appear to be at least some differences across constructs in terms of the correlation with mental ability test scores. In particular, ratings of general intelligence, job knowledge, and experience had the highest mean correlations, whereas ratings of applied social skills (communication, interpersonal skills, and leadership), applied mental skills, and interests and preferences had the lowest mean correlations.

Table 2
Differences in Construct Frequency Between Low- and High-Structure Interviews

	Low structure		High structure	
Category and construct	n	%	n	%
Mental capability	25	19.2	30	14.4
General intelligence	15	11.5	5	2.4
Specific ability	1	0.8	0	0.0
Applied mental skills	6	4.6	22	10.6
Creativity and innovation	3	2.3	3	1.4
Knowledge and skills	15	11.5	18	8.7
Job knowledge and skills	3	2.3	11	5.3
Education and training	4	3.1	2	1.0
Experience and general work history	8	6.2	5	2.4
Basic personality tendencies	48	36.9	70	33.7
Extroversion	8	6.2	13	6.3
Conscientiousness	19	14.6	36	17.3
Agreeableness	7	5.4	3	1.4
Openness to experience	0	0.0	6	2.9
Emotional stability	12	9.2	9	4.3
Other personality traits	2	1.5	3	1.4
Applied social skills	23	17.7	71	34.1
Communication skills	7	5.4	19	9.1
Interpersonal skills	11	8.5	32	15.4
Leadership	3	2.3	17	8.2
Persuading and negotiating	2	1.5	3	1.4
Interests and preferences	7	5.4	8	3.8
Occupational interests	5	3.8	8	3.8
Hobbies and extracurricular activities	2	1.5	0	0.0
Organizational fit	3	2.3	8	3.8
Values and moral standards	3	2.3	8	3.8
Physical attributes	9	6.9	3	1.4
General physical attributes	6	4.6	2	1.0
Job-related physical skills	3	2.3	1	0.5

Note. There were 19 studies in the low-structure category with a total of 130 characteristics rated and 28 studies in the high-structure category with a total of 208 characteristics rated.

Table 3
Mean Validity Correlations for Ratings of Interview Constructs

Category and construct	Overall					Low structure					High structure				
	k	*TSS*	$\bar{r}_{xy}$	90% CI	$\bar{r}_{xy}$(c)	*k*	*TSS*	$\bar{r}_{xy}$	90% CI	$\bar{r}_{xy}$(c)	*k*	*TSS*	$\bar{r}_{xy}$	90% CI	$\bar{r}_{xy}$(c)
Mental capability															
General intelligence	8	1,916	.13	.04–.22	.24	6	1,699	.14	.06–.22	.26	2	217	.11	—	—
Applied mental skills	13	5,027	.15	.08–.22	.28	4	1,160	.07	−.02–.16	.13	9	3,867	.19	.11–.27	.35
Creativity and innovation	4	296	.32	.21–.33	.58	2	152	.28	—	—	2	144	.36	—	—
Knowledge and skills															
Job knowledge and skills	6	2,617	.23	.11–.35	.42	1	31	.49	—	—	5	2,586	.18	.07–.29	.33
Education and training	1	312	.05	—	—	1	312	.05	—	—	0	—	—	—	—
Experience and general work history	3	495	.27	.11–.43	.49	2	380	.17	—	—	1	115	.47	—	—
Basic personality tendencies															
Extroversion	8	1,055	.18	.12–.24	.33	3	650	.12	.09–.15	.22	5	405	.22	.14–.30	.40
Conscientiousness	22	3,532	.18	.13–.23	.33	6	1,656	.13	.05–.21	.24	16	1,876	.20	.15–.35	.37
Agreeableness	4	344	.28	.26–.30	.51	1	68	.25	—	—	3	276	.29	.27–.31	.53
Openness to experience	2	527	.16	—	—	0	—	—	—	—	2	527	.16	—	—
Emotional stability	6	917	.26	.20–.32	.47	2	539	.18	—	—	4	378	.31	.27–.35	.56
Other personality traits	1	102	.08	—	—	0	—	—	—	—	1	102	.08	—	—
Applied social skills															
Communication skills	9	2,963	.14	.06–.22	.26	2	783	.05	—	—	7	2,180	.17	.07–.27	.31
Interpersonal skills	19	3,620	.21	.16–.26	.39	5	1,191	.17	.08–.26	.31	14	2,429	.22	.16–.28	.40
Leadership	8	633	.26	.18–.34	.47	2	152	.40	—	—	6	481	.22	.13–.31	.40
Persuading and negotiating	3	245	.13	.10–.16	.24	2	152	.11	—	—	1	93	.18	—	—
Interests and preferences															
Occupational interests	9	914	.13	.05–.21	.24	2	380	.08	—	—	7	534	.14	.07–.21	.26
Organizational fit															
Values and moral standards	5	912	.27	.17–.37	.49	1	537	.07	—	—	4	375	.32	.23–.41	.58
Physical attributes															
General physical attributes	1	312	−.18	—	—	1	312	−.18	—	—	0	—	—	—	—
Job-related physical skills	2	935	.15	—	—	1	471	.10	—	—	1	464	.19	—	—

Note. Confidence intervals (CI) and corrected mean validity correlations are not shown for constructs with fewer than three coefficients, indicated by those cells with dashes. *k* = the number of dimensions in each construct category that had provided that information; *TSS* = the total number of subjects associated with those dimensions; $\bar{r}_{xy}$ = the uncorrected correlation between the interview ratings and job performance evaluations; $\bar{r}_{xy}$(c) = the correlation corrected for range restriction in the interview and measurement error in performance evaluations.

Results for low-structure and high-structure interviews are shown in the remaining part of Table 4. Across constructs, the mean corrected correlation with mental ability test scores was .10 for high-structure interviews and .31 for low-structure interviews. This analysis suggests greater saturation of mental ability in ratings for low-structure interviews, a finding that has direct implication for racial group differences. Racial group differences are discussed in the next section.

On a slightly different note, these results also provide at least some information on the construct validity of interview ratings. Ratings of general intelligence correspond directly to mental ability test scores at a construct level and should have had a higher correlation with the test scores than the other constructs. Although ratings of general intelligence did in fact have the highest mean correlation among all of the constructs, the low magnitude of this mean correlation (.32 uncorrected, .44 corrected) suggests that these two measurement devices are far from equivalent. More than that, this finding casts doubt on the ability to evaluate general intelligence with an employment interview.

Racial Group Differences in Construct Ratings

Mean effect sizes for racial differences in the various construct ratings are shown in Table 5. Across all studies and constructs, the mean *d* value was .30. This value suggests at least some racial group differences in interview construct ratings. These differences are considerably less than those for mental ability tests, which can approach a difference of one standard deviation (see Hunter & Hunter, 1984). However, it is important to note that the 1.00 value applies to ability tests as a whole, whereas our values apply to the average group differences resulting from a single rating.

As indicated in Table 5, there appeared to be some racial group differences across constructs. The highest mean effect size was for ratings of general intelligence and ratings of experience and general work history (both had a mean *d* value of .49). These values are high, especially because they reflect racial group differences for a single rating, and may suggest some caution in the evaluation of these characteristics with interviews. The lowest mean value was for ratings of applied mental skills, which had a mean *d* of .13.

Racial group differences for low-structure and high-structure interviews, respectively, are also shown in Table 5. As indicated in Table 5, racial group differences were considerably higher overall for low-structure interviews. In particular, the average *d* value across all constructs rated in low-structure interviews was .51, whereas the average *d* value across all constructs rated in high-structure interviews was .13. It would appear that less formal use of a job analysis and giving interviewers more discretion in terms

Table 4
Mean Correlations Between Interview Construct Ratings and Mental Ability Test Scores

Category and construct	Overall					Low structure					High structure				
	k	*TSS*	$\bar{r}_g$	90% CI	$\bar{r}_g$(c)	*k*	*TSS*	$\bar{r}_g$	90% CI	$\bar{r}_g$(c)	*k*	*TSS*	$\bar{r}_g$	90% CI	$\bar{r}_g$(c)
Mental capability															
General intelligence	5	795	.32	.22–.42	.44	4	623	.34	.22–.46	.46	1	172	.25	—	—
Applied mental skills	8	3,273	.05	.00–.10	.07	1	197	.13	—	—	7	3,076	.04	−.02–.10	.06
Creativity and innovation	2	200	.20	—	—	1	107	.29	—	—	1	93	.10	—	—
Knowledge and skills															
Job knowledge and skills	6	2,616	.27	.19–.35	.37	2	134	.34	—	—	4	2,482	.24	.22–.26	.33
Education and training	0	—	—	—	—	0	—	—	—	—	0	—	—	—	—
Experience and general work history	4	688	.26	.23–.29	.36	3	516	.27	.23–.31	.37	1	172	.25	—	—
Basic personality tendencies															
Extroversion	4	678	.15	.02–.28	.21	2	413	.24	—	—	2	265	.07	—	—
Conscientiousness	16	2,495	.11	.05–.17	.16	6	1,036	.17	.07–.27	.24	10	1,449	.07	−.01–.15	.10
Agreeableness	4	688	.16	.10–.22	.23	3	516	.17	.10–.24	.24	1	172	.13	—	—
Openness to experience	1	76	−.02	—	—	0	—	—	—	—	1	76	−.02	—	—
Emotional stability	6	1,437	.15	.08–.22	.21	3	516	.19	.08–.30	.27	3	921	.11	.05–.17	.16
Other personality traits	1	101	−.03	—	—	1	101	−.03	—	—	0	—	—	—	—
Applied social skills															
Communication skills	5	2,318	.08	.03–.13	.11	1	103	.11	—	—	4	2,035	.07	.01–.13	.10
Interpersonal skills	12	2,240	.00	−.06–.06	.00	2	138	−.13	—	—	10	2,101	.02	−.04–.08	.03
Leadership	5	783	.07	.05–.09	.10	1	107	.04	—	—	4	676	.08	.05–.11	.11
Persuading and negotiating	2	200	.05	—	—	1	107	.06	—	—	1	93	.04	—	—
Interests and preferences															
Occupational interests	7	1,058	.09	.03–.15	.13	2	413	.21	—	—	5	645	.05	.01–.09	.07
Organizational fit															
Values and moral standards	2	489	.05	—	—	0	—	—	—	—	2	489	.05	—	—
Physical attributes															
General physical attributes	1	103	.11	—	—	1	103	.11	—	—	0	—	—	—	—
Job-related physical skills	0	—	—	—	—	0	—	—	—	—	0	—	—	—	—

Note. Confidence intervals (CI) and corrected mean correlations are not shown for constructs with fewer than three coefficients, indicated by those cells with dashes. *k* = the number of dimensions in each construct category that had provided that information; *TSS* = the total number of subjects associated with those dimensions; $\bar{r}_g$ = the uncorrected correlation between the interview ratings and job performance evaluations; $\bar{r}_g$(c) = the correlation corrected for range restriction in the interview and measurement error in performance evaluations.

of the interview process tended to result in higher racial group differences in the subsequent ratings.

Finally, it should be noted that racial group differences for ratings of personality characteristics overall were at last three times higher than those found for paper-and-pencil tests. For example, Ones and Viswesvaran (1998) found that racial group differences for personality dimensions such as integrity tended to be below an effect size of .10. However, we note that self-reports of personality and other ratings of personality often are only slightly correlated (Mount, Barrick, & Strauss, 1994). Interviews may represent a situation in which a different method of measurement (using others' ratings) provides different personality-related information that is associated with somewhat larger racial group differences.

Sex Group Differences in Construct Ratings

Mean effect sizes for sex differences in the various construct ratings are shown in Table 6. Across all constructs, the mean *d* value was .06, which suggests negligible sex differences in interview construct ratings overall. There were some differences across constructs, such as the positive .13 mean effect size (i.e., favoring male subjects) for ratings of general intelligence and the −.13 mean effect size (i.e., favoring female subjects) for ratings of applied social skills.

Mean effect sizes for low-structure and high-structure studies, respectively, are also shown in Table 6. The mean *d* value for all constructs rated in high-structure interviews was .00, suggesting that these interviews overall had little or no impact on female subjects. The mean *d* value across all constructs rated in low-structure interviews was .23, suggesting that these interviews overall had some impact on female subjects. There was one construct for which a direct comparison could be made between low-structure and high-structure interviews, and that construct was conscientiousness. Results for conscientiousness suggested higher group differences for low-structure interviews than for high-structure interviews (mean *d*s of .34 and .12, respectively).

Discussion

The first purpose of this investigation was to develop a taxonomy of possible constructs that employment interviews could measure. Using literature from a number of areas in psychology, we constructed a comprehensive taxonomy with seven different categories of constructs: mental capability, knowledge and skills, basic personality tendencies, applied social skills, interests and

Table 5
Mean Effect Sizes for Racial Group Differences in Interview Construct Ratings

Category and construct	Overall				Low structure				High structure			
	k	*TSS*	$\bar{d}_{WB}$	90% CI	*k*	*TSS*	$\bar{d}_{WB}$	90% CI	*k*	*TSS*	$\bar{d}_{WB}$	90% CI
Mental capability												
General intelligence	6	2,331	.49	.25–.73	5	1,564	.58	.35–.81	1	767	.03	—
Applied mental skills	6	1,566	.13	.00–.26	1	471	.49	—	5	1,095	.06	−.02–.14
Creativity and innovation	0	—	—	—	0	—	—	—	0	—	—	—
Knowledge and skills												
Job knowledge and skills	1	103	1.07	—	1	103	1.07	—	0	—	—	—
Education and training	0	—	—	—	0	—	—	—	0	—	—	—
Experience and general work history	3	1,208	.49	.28–.70	2	441	.63	—	1	767	.21	—
Basic personality tendencies												
Extroversion	3	1,333	.18	.11–.25	1	338	.28	—	2	995	.14	—
Conscientiousness	15	5,443	.30	.20–.40	8	2,554	.41	.26–.56	7	2,889	.17	.11–.23
Agreeableness	4	1,313	.33	.18–.48	2	441	.52	—	2	872	.15	—
Openness to experience	1	267	.05	—	0	—	—	—	1	267	.05	—
Emotional stability	6	2,359	.34	.23–.45	3	912	.45	.33–.57	3	1,447	.24	.12–.36
Other personality traits	1	103	.52	—	1	103	.52	—	0	—	—	—
Applied social skills												
Communication skills	7	1,864	.39	.22–.56	2	574	.87	—	5	1,290	.21	.11–.31
Interpersonal skills	6	1,733	.22	.11–.33	2	652	.37	—	4	1,081	.14	−.03–.25
Leadership	2	463	.25	—	0	—	—	—	2	463	.25	—
Persuading and negotiating	0	—	—	—	0	—	—	—	0	—	—	—
Interests and preferences												
Occupational interests	2	1,105	.33	—	1	338	.39	—	1	767	.26	—
Organizational fit												
Values and moral standards	3	568	−.12	−.15–−.09	0	—	—	—	3	568	−.12	−.15–−.09
Physical attributes												
General physical attributes	1	103	.48	—	1	103	.48	—	0	—	—	—
Job-related physical skills	3	919	.21	.08–.34	2	652	.31	—	1	267	.02	—

Note. A positive effect size indicates that White candidates had higher mean interview ratings than Black candidates. Confidence intervals (CI) are not shown for constructs with fewer than three coefficients, indicated by those cells with dashes. *k* = the number of dimensions in each construct category that provided racial group information; *TSS* = the total number of subjects associated with those dimensions; $\bar{d}_{WB}$ = the mean effect size for White–Black group differences.

preferences, organizational fit, and physical attributes. The development of such a taxonomy is obviously a critical component in the process of understanding interview constructs, because it guided the classification and interpretation of data in this study and may influence future research in this area. We believe that our taxonomy provides a comprehensive and meaningful basis on which interview construct research can build and, as such, represents a contribution to the interview literature in and of itself.

The second purpose of this investigation was to evaluate which constructs in this framework are actually rated in employment interviews and, perhaps more important, which are the most commonly assessed. To make this assessment, we compiled a database of 338 characteristics that were rated in 47 actual interview studies. These 47 studies included a diverse mixture of companies, products, format (e.g., level of structure, type of questions), job complexity, and sources (e.g., journal articles, technical reports, conference papers, and dissertations). We believe this database was reasonably representative of employment interviews in general and provided a sound basis on which to analyze interview constructs.

Our results suggest that personality traits and applied social skills are rated more often in employment interviews than are any other type of construct. These constructs reflect behavioral tendencies and provide employers with an idea of how potential employees are likely to act on the job and how well they can interact with other employees. Given the frequency with which they appeared to be rated (combined, they accounted for more than 60% of all the rated characteristics), it would seem that many employers are interested in behavioral tendencies and that they are an important part of many jobs. Among these characteristics, conscientiousness was the single most commonly rated construct, as it accounted for more than 16% of all ratings and appeared under labels such as *responsibility, dependability, initiative,* and *persistence.* Interpersonal skills was the next most frequently rated construct, as it accounted for approximately 13% for all ratings and appeared under labels such as *interpersonal relations, social skills, team focus,* and *the ability to work with people.*

Mental capability and knowledge and skills were the next most frequently rated constructs after behavioral tendencies. These constructs reflect either what candidates already know or how well they can process new information, and combined, they accounted for more than 25% of all the characteristics rated. In fact, behavioral tendencies and mental-type constructs together accounted for almost 90% of all interview ratings, whereas interests, organizational fit, and physical attributes accounted for the remaining ratings.

A key finding from our frequency analyses was that low-structure and high-structure interviews do not tend to measure the same constructs. In particular, low-structure interviews often focus more on constructs such as general intelligence, education and training, experience, and interests, whereas high-structure inter-

Table 6
Mean Effect Sizes for Sex Group Differences in Interview Construct Ratings

Category and construct	Overall				Low structure				High structure			
	k	*TSS*	$\bar{d}_{MF}$	90% CI	*k*	*TSS*	$\bar{d}_{MF}$	90% CI	*k*	*TSS*	$\bar{d}_{MF}$	90% CI
Mental capability												
General intelligence	4	1,250	.13	−.17–.43	3	697	.08	−.31–.47	1	823	.28	—
Applied mental skills	6	1,598	−.13	−.22–−.04	0	—	—	—	5	1,598	−.13	−.22–−.04
Creativity and innovation	0	—	—	—	0	—	—	—	0	—	—	—
Knowledge and skills												
Job knowledge and skills	1	89	−.08	—	0	—	—	—	0	—	—	—
Education and training	0	—	—	—	0	—	—	—	0	—	—	—
Experience and general work history	2	1,168	.59	—	1	345	.49	—	1	823	.69	—
Basic personality tendencies												
Extroversion	4	1,496	.30	.16–.44	1	345	.38	—	3	1,151	.27	.09–.45
Conscientiousness	12	4,041	.19	.06–.32	4	1,052	.34	.17–.51	8	2,989	.12	−.03–.27
Agreeableness	4	1,458	.04	−.27–.35	2	516	.06	—	2	942	.02	—
Openness to experience	1	464	−.01	—	0	—	—	—	1	464	−.01	—
Emotional stability	3	1,490	.16	−.29–.61	1	345	.54	—	2	1,145	−.03	—
Other personality traits	0	—	—	—	0	—	—	—	0	—	—	—
Applied social skills												
Communication skills	4	911	−.26	−.43–−.09	0	—	—	—	4	911	−.26	−.43–−.09
Interpersonal skills	8	1,490	−.12	−.23–−.01	1	181	−.41	—	7	1,309	−.08	−.18–.06
Leadership	2	219	−.02	—	0	—	—	—	2	219	−.02	—
Persuading and negotiating	0	—	—	—	0	—	—	—	0	—	—	—
Interests and preferences												
Occupational interests	2	1,168	.41	—	1	345	.34	—	1	823	.47	—
Organizational fit												
Values and moral standards	3	325	−.28	−.43–−.13	0	—	—	—	3	325	−.28	−.43–−.13
Physical attributes												
General physical attributes	1	171	−.77	—	1	171	−.77	—	0	—	—	—
Job-related physical skills	2	645	.67	—	1	181	1.38	—	1	464	−.05	—

Note. A positive effect size indicates that male candidates had higher mean interview ratings than female candidates. Confidence intervals (CI) are not shown for constructs with fewer than three coefficients, indicated by those cells with dashes. *k* = the number of dimensions in each construct category that provided that sex group information; *TSS* = the total number of subjects associated with those dimensions; $\bar{d}_{MF}$ = the mean effect size for male–female group differences.

views often focus more on constructs such as job knowledge and skills, organizational fit, interpersonal and social skills, and applied mental skills (e.g., problem solving, decision making). These differences are most likely due to the more frequent (and stringent) use of formal job analysis methodology in the development of high-structure interviews. Thus, it would appear that structure influences not only the procedural conduct of the interview (e.g., consistency in the asking of questions and the manner in which responses are scored) but also what constructs are rated. We are unaware of the issue of different constructs being empirically substantiated in any other place in the literature and believe this represents a contribution in and of itself.

Another purpose of this investigation was to begin to accumulate information on the properties of interview construct ratings, including validity and group differences. We were able to identify a number of constructs that had high validity across jobs, including job knowledge, interpersonal skills, creativity, agreeableness, job knowledge, emotional stability, leadership, and organizational fit. We were also able to identify some constructs that had fairly low validity across jobs, including interests and preferences and communication skills.

What is particularly interesting is that many of the high-validity constructs are those that tend to be assessed more often in high-structure interviews. Although one would expect higher validity for constructs rated more often in high-structure interviews, such a correspondence could suggest another explanation for why structured interviews have higher validity than low-structure interviews. It is generally accepted that structured interviews have higher validity in part because they represent more reliable assessment of responses (Conway et al., 1995). Taken as a whole, our evidence suggests that structured interviews also have higher validity because the constructs rated more frequently in them (e.g., job knowledge, interpersonal skills, organizational fit) tend to be better predictors of performance than the constructs rated more frequently in low-structure interviews (e.g., interests, education, experience).

What lends further support to the above notion is that it is largely consistent with existing selection literature. Hunter and Hunter (1984), for example, found that standard measures of job knowledge had high validity whereas standard measures of education, experience, and interests had much lower validity. The one exception was general mental ability, as existing selection literature suggests it is a strong predictor of performance (Hunter & Hunter, 1984), yet interview ratings of it did not correlate strongly with job performance. However, as noted a little later, the results of our construct analysis suggest that interviewer ratings of general intelligence do not tend to have a strong association with the mental ability construct.

On a more practical note, we were able to identify several constructs that appear to have a desirable combination of proper-

ties for structured interviews. In particular, applied mental skills (e.g., decision making, problem solving, judgment), interpersonal skills (e.g., interpersonal relations, rapport, tact, cooperation), and conscientiousness (e.g., reliability, dependability, persistence) all appear to provide reasonable validity with low racial differences and very low sex differences (which often favor women). So, when there is a choice of which constructs to focus on with a structured interview (based on the job analysis), our results suggest that the above constructs are a better choice than other constructs like interests and communication skills. These recommendations should be viewed as tentative of course because several of the findings are based on a relatively small number of studies. In addition, we should note that several other constructs, particularly creativity, organizational fit (e.g., appreciation for diversity, quality orientation, pride in the organization) and emotional stability (e.g., stress tolerance, poise, self-control), look very promising in terms of validity, but there were insufficient data available to assess their group differences.

In contrast, identifying desirable constructs among less structured interviews was more difficult. The ratings made from these interviews were generally less valid and had higher race and sex differences. Among the constructs for which we had sufficient data to analyze, conscientiousness-related characteristics were probably the most desirable because they provided at least some validity with only moderate levels of race and sex group differences. In comparison, ratings of general intelligence had similar validity but higher levels of race and sex differences. Ratings of interpersonal skills actually had higher validity than either ratings of conscientiousness or ratings of general intelligence, but we did not have sufficient data to analyze their group differences. Again, we do not advocate use of low-structure interviews but want to provide at least some guidance for them in case an organization insists on using them.

One other purpose of this investigation was to begin the process of determining the degree to which interview ratings actually reflect the intended constructs. This is an important issue and one that should not be taken for granted, especially because it has been raised in other areas of selection. For example, a common finding in the assessment center literature (Fleenor, 1996; Sackett & Dreher, 1982) is that individual ratings tend to align more with the exercise from which they came (e.g., in-basket) than with the specific characteristic being rated (e.g., leadership).

Although data on the correspondence between interview ratings and paper-and-pencil scores of the same constructs were generally sparse, we did find sufficient information to make a partial assessment with general intelligence. Our results did not provide strong support for the construct validity of these ratings, at least not for low-structure interviews. Rather, these results suggest that assessment of general intelligence attributes (e.g., ability to learn, intellectual capacity) is probably best left to more traditional paper-and-pencil instruments. As more data become available, it may be possible to determine if the relationship between ratings of general intelligence and mental ability test scores is stronger for structured interviews.

An interesting question is why ratings of general intelligence do not correspond very well to mental ability test scores in low-structure interviews. It could be a matter of reliability, because the reliability of low-structure interviews in general is not very good (Conway et al., 1995). Alternatively, it may be that ratings of general intelligence are picking up on other constructs, such as general impressions or characteristics that are salient during the interview process. Extroversion, for example, has been found to influence interviewer evaluations in at least one study (Caldwell & Burger, 1998).

So what have we learned about employment interview constructs? Four major conclusions emerge from the above discussion. First, a variety of constructs can be and are rated in employment interviews. Although personality, social skills, mental capability, and knowledge and skills are the most commonly rated, other constructs such as organizational fit, interests, and physical attributes are rated as well. Second, high-structure interviews tend to focus on different constructs than low-structure interviews, most likely due to the more stringent use of job analysis methodology in the former. Third, part of the reason for the higher validity of structured interviews may be that they focus more on constructs that have a stronger relationship with performance (e.g., job knowledge vs. education). Finally, group differences (both race and sex) should be taken into account when employment interviews are being developed because the extent of these differences varies across constructs (even for constructs with comparable validity).

Where do we go from here? Probably one of the most important needs for future research is to assess the degree to which interviewers can accurately assess personality and social characteristics. Specifically, we need correlations between interviewer ratings and independent assessments of the same characteristics. In addition, we call for more research on the relative contribution of reliability and specific construct measurement to high-structure interview validity. Finally, we need a lot more data to flesh out our understanding of many of these constructs and, in particular, to allow better control for moderators such as the degree of structure. There does appear to be a tendency for researchers to report information only on total interview scores, including relationships with ability, personality tests, or both. We call for interview researchers to thoroughly report their results for each dimension and not just for the total interview scores. We also call for more job-specific interview research to tease out which constructs are more universal and which are more specific to given classes of jobs (e.g., sales, managerial).

As always, limitations of this investigation should be noted. First, although the total number of dimensions collected for construct identification and mapping was reasonable ($N = 338$), the number of coefficients for some of the analyses was low. Small sample sizes made these results more tentative and did not allow for analysis of other potential moderator variables in addition to structure. Second, we did not find data on the correlation between interviewer ratings of personality traits and social skills and corresponding paper-and-pencil measures of the same traits. This information would allow determination of the degree to which interviewers are successful in assessing these characteristics. Third, we did not find enough interrater reliability data to analyze, because again most of these data were reported only for total interview scores. Fourth, we could not correct the validity and ability correlations individually for artifacts, which may have led to some inconsistencies in the corrected mean values. Finally, constructing the construct taxonomy required judgment on our part, and it is quite possible that other researchers might have organized it in a different way.

References

References marked with an asterisk indicate studies included in the meta-analysis.

*Adorno, A. J., Binning, J. F., Srinivasagam, N., & Williams, K. B. (1997, April). *Incremental validity of structured panel interview ratings and ability measures.* Paper presented at the 12th Annual Conference of the Society for Industrial and Organizational Psychology, St. Louis, MO.

*Albrecht, P. A., Glaser, E. M., & Marks, J. (1964). Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology, 48,* 351–360.

Allport, G. W. (1937). *Personality.* New York: Holt.

Arvey, R. D., & Campion, J. E. (1982). The employment interview: A summary and review of recent research. *Personnel Psychology, 35,* 281–322.

Barrick, M. R., & Mount, M. K. (1991). The Big Five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44,* 1–26.

*Berkley, S. (1984). *VII. Validation report corrections officer trainee.* Commonwealth of Pennsylvania, State Civil Service Commission.

Bobko, P., Roth, P. L., & Potosky, D. (1999). Derivation and implications of a meta-analytic matrix incorporating cognitive ability, alternative predictors and job performance. *Personnel Psychology, 52,* 561–589.

*Bolanvich, D. J. (1994). Selection of female engineering trainees. *Journal of Educational Psychology, 35,* 545–553.

Borman, W. C., White, L. A., Pulakos, E. D., & Oppler, S. H. (1991). Models of supervisory job performance ratings. *Journal of Applied Psychology, 76,* 863–872.

*Bosshardt, M. J. (1992). *Situational interviews versus behavior description interviews: A comparative validity study.* Unpublished doctoral dissertation, University of Minnesota, Twin Cities Campus.

*Bradley, K. M., Bernthal, P., & Thomas, J. (1998, April). *Factor structure and predictive validity of structured employment interviews: An application of industrial psychology in a Kuwaiti petrochemical corporation.* Paper presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

Cable, D. M., & Judge, T. A. (1997). Interviewers' perceptions of person–organization fit and organizational selection decisions. *Journal of Applied Psychology, 82,* 546–561.

Caldwell, D. F., & Burger, J. M. (1998). Personality characteristics of job applicants and success in screening interviews. *Personnel Psychology, 51,* 119–136.

Campion, M. A., Campion, J. E., & Hudson, J. P., Jr. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology, 79,* 998–1002.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1997). A review of structure in the selection interview. *Personnel Psychology, 50,* 655–702.

Campion, M. A., Pursell, E. D., & Brown, B. K. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. *Personnel Psychology, 41,* 25–42.

Cardall, A. J. (1942). *Preliminary manual for the Test of Practical Judgment.* Chicago: Science Research Associates.

*Chapman, D. S., & Rowe, P. M. (1998). *The impact of videoconference media, interview structure, and interviewer sex on interviewer evaluations in the employment interview: A field experiment.* Unpublished manuscript.

Cohen, R. J., & Swerdlik, M. E. (1999). *Psychological testing and assessment: An introduction to tests and measurement* (4th ed.). Mountain View, CA: Mayfield.

*Conard, M. A. (1988). The contribution of the interview to the prediction of job performance. (Doctoral dissertation, University of Connecticut, 1988). *Dissertation Abstracts International, 49,* 5555.

Conway, J. M., Jako, R. A., & Goodman, D. F. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80,* 565–579.

*Conway, J. M., & Peneno, G. M. (1999). Comparing structured interview question types: Construct validity and applicant reactions. *Journal of Business and Psychology, 13,* 485–506.

Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO-PI–R) and NEO Five-Factor Inventory (NEO-FFI) professional manual.* Odessa, FL: Psychological Assessment Resources.

*Dalessio, A. T., & Silverhart, T. A. (1994). Combining biodata, test, and interview information: Predicting decisions and performance criteria. *Personnel Psychology, 47,* 303–315.

*DeGroot, T. (1997). *The impact of managerial nonverbal cues on the reactions of subordinates.* Unpublished doctoral dissertation, University of Florida.

*Delaney, E. C. (1954). *Teacher selection and evaluation: With special attention to the validity of the personal interview and the National Teacher Examinations as used in one selected community.* Unpublished doctoral dissertation, Columbia University.

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology, 41,* 417–440.

Dipboye, R. L. (1992). *Selection interviews: Process perspectives.* Cincinnati, OH: South-Western.

*Dipboye, R. L., Gaugler, B. B., Hayes, T. L., & Parker, D. S. (1992). *Individual differences in the incremental validity of interviewers' judgments.* Unpublished manuscript.

*Dougherty, T. W., Ebert, R. J., & Callender, J. C. (1986). Policy capturing in the employment interview. *Journal of Applied Psychology, 71,* 9–15.

Erez, A., Bloom, M. C., & Wells, M. T. (1996). Using random rather than fixed effects models in a meta-analysis: Implications for situational specificity and validity generalization. *Personnel Psychology, 49,* 275–306.

Festinger, L. (1957). *A theory of cognitive dissonance.* Stanford, CA: Stanford University Press.

Fleenor, J. W. (1996). Constructs and developmental assessment centers: Further troubling empirical findings. *Journal of Business and Psychology, 10,* 319–335.

Fuller, J. B., & Hester, K. (1999). Comparing the sample-weighted and unweighted meta-analysis: An applied perspective. *Journal of Management, 25,* 803–828.

*Gaines, L. K., & Lewis, B. R. (1982). Reliability and validity of the oral interview board in police promotions: A research note. *Journal of Criminal Justice, 10,* 403–419.

Gardner, H. (1983). *Frames of mind: The theory of multiple intelligences.* New York: Basic Books.

*Gordon, M. J., & Lincoln, J. A. (1976). Family practice resident selection: Value of the interview. *Journal of Family Practice, 3,* 175–177.

*Green, P. C., Alter, P., & Carr, A. F. (1993). Development of standard anchors for scoring generic past-behavior questions in structured interviews. *International Journal of Selection and Assessment, 1*(4), 203–212.

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

Harris, M. M. (1999). What is being measured? In R. W. Eder & M. M. Harris (Eds.), *The employment interview handbook* (pp. 143–157). Thousand Oaks, CA: Sage.

Hedges, L. V., & Vevea, J. L. (1998). Fixed- and random-effects models in meta-analysis. *Psychological Methods, 3,* 486–504.

Herrnstein, R. J., & Murray, C. (1994). *The bell curve: Intelligence and class structure in American life.* New York: Free Press.

Hilton, A. C., Bolin, S. F., Parker, J. W., Jr., Taylor, E. K., & Walker, W. B. (1955). The validity of personnel assessments by professional psychologists. *Journal of Applied Psychology, 39,* 287–293.

*Hoffman, C. C., & Holden, L. M. (1993). Dissecting the interview: An application of generalizability analysis. In D. L. Denning (Chair), *Psy-*

chometric analysis of the structured interview. Symposium conducted at the 8th Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Huffcutt, A., & Arthur, W., Jr. (1994). Hunter and Hunter (1984) revisited: Interview validity for entry-level jobs. *Journal of Applied Psychology, 79,* 184–190.

Huffcutt, A., Roth, P., & McDaniel, M. (1996). A meta-analytic investigation of cognitive ability in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81,* 459–473.

*Huffcutt, A., Weekley, J., Wiesner, W., DeGroot, T., & Jones, C. (in press). Evaluation and comparison of situational and behavior description interview questions. *Personnel Psychology.*

Hunter, J. E., & Hunter, R. F. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96,* 72–98.

Hunter, J. E., & Schmidt, F. L. (1990). *Methods of meta-analysis: Correcting error and bias in research findings.* Newbury Park, CA: Sage.

*Huse, E. F. (1962). Assessments of higher-level personnel: IV. The validity of the assessment techniques based on systematically varied information. *Personnel Psychology, 15,* 195–205.

Janz, T. (1982). Initial comparisons of patterned behavior description interviews versus unstructured interviews. *Journal of Applied Psychology, 67,* 577–580.

*Johnson, E. K. (1991). The structured interview: Manipulating structuring criteria and the effects on validity, reliability, and practicality. (Doctoral dissertation, Tulane University, 1990). *Dissertation Abstracts International, 51,* 5622.

*Jones, D. E., Jr. (1978). *Predicting teacher processes with the Teacher Perceiver Interview.* Unpublished doctoral dissertation, Virginia Polytechnic Institute and State University.

Karren, R. J., & Graves, L. M. (1994). Assessing person–organization fit in personnel selection: Guidelines for future research. *International Journal of Selection and Assessment, 2*(3), 146–156.

*Kinicki, A. J., Lockwood, C. A., Hom, P. W., & Griffeth, R. W. (1990). Interviewer predictions of applicant qualifications and interviewer validity: Aggregate and individual analyses. *Journal of Applied Psychology, 75,* 477–486.

Klimoski, R. J. (1993). Predictor constructs and their measurement. In N. Schmitt & W. C. Borman (Eds.), *Personnel selection in organizations* (pp. 99–134). San Francisco: Jossey-Bass.

Kristof, A. L. (1996). Person–organization fit: An integrative review of its conceptualizations, measurement, and implications. *Personnel Psychology, 49,* 1–49.

*Landy, F. J. (1976). The validity of the interview in police officer selection. *Journal of Applied Psychology, 61,* 193–198.

Latham, G. P., Saari, L. M., Pursell, E. D., & Campion, M. A. (1980). The situational interview. *Journal of Applied Psychology, 65,* 422–427.

*Little, J. P., Shoenfelt, E. L., & Brown, R. D. (2000, April). *The situational versus the patterned-behavior-description interview for predicting customer-service performance.* Paper presented at the 15th Annual Conference of the Society for Industrial and Organizational Psychology, New Orleans, LA.

*Martin, M. A. (1972). *Reliability and validity of Canadian Forces selection interview procedures* (Report No. 72-4). CFB Toronto, Downsview, Ontario, Canada: Canadian Forces Personnel Applied Research Unit.

McDaniel, M. A., Whetzel, D. L., Schmidt, F. L., & Maurer, S. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79,* 599–616.

*Moreano, A. G., & Sproule, C. F. (1976). *Reliability and other data on structured oral examinations.* Unpublished manuscript, Commonwealth of Pennsylvania, State Civil Service Commission.

Mount, M. K., & Barrick, M. R. (1995). The Big Five personality dimensions: Implications for research and practice in human resources management. *Research in Personnel and Human Resources Management, 13,* 153–200.

Mount, M. K., Barrick, M. R., & Strauss, J. P. (1994). Validity of observer ratings of the Big Five personality factors. *Journal of Applied Psychology, 79,* 272–280.

Murphy, K. R., & Davidshofer, C. O. (1998). *Psychological testing: Principles and applications* (4th ed.). Upper Saddle River, NJ: Prentice Hall.

Murphy, K. R., & DeShon, R. (2000). Inter-rater correlations do not estimate the reliability of job performance ratings. *Personnel Psychology, 53,* 873–900.

Oliver, P. J. (1989). Towards a taxonomy of personality descriptors. In D. Buss & N. Cantor (Eds.), *Personality psychology: Recent trends and emerging directions* (pp. 261–271). New York: Springer-Verlag.

Ones, D., & Viswesvaran, C. (1998). Gender, age, and race differences on overt integrity tests: Results across four large-scale job applicant data sets. *Journal of Applied Psychology, 83,* 35–42.

Orpen, C. (1985). Patterned behavior description interviews versus unstructured interviews: A comparative validity study. *Journal of Applied Psychology, 70,* 774–776.

Osburn, H. G., & Callender, J. (1992). A note on the sampling variance of the mean uncorrected correlation in meta-analysis and validity generalization. *Journal of Applied Psychology, 77,* 115–122.

*Prewett-Livingston, A. J., Feild, H. S., Veres, J. G., III, & Lewis, P. M. (1994). Effects of face on interview ratings in a situational panel interview. *Journal of Applied Psychology, 81,* 178–186.

*Prien, E. P. (1962). Assessments of higher-level personnel: V. An analysis of interviewers' predictions of job performance. *Personnel Psychology, 15,* 319–334.

*Pulakos, E. D., & Schmitt, N. (1995). Experience-based and situational interview questions: Studies of validity. *Personnel Psychology, 48,* 289–308.

*Raza, S. M., & Carpenter, B. N. (1987). A model of hiring decisions in real employment interviews. *Journal of Applied Psychology, 72,* 596–603.

*Reeb, M. (1969). A structured interview for predicting military adjustment. *Occupational Psychology, 43,* 193–199.

Riggio, R. (1996). *Introduction to industrial/organizational psychology* (2nd ed.). New York: Harper Collins.

*Robertson, I. T., Gratton, L., & Rout, U. (1990). The validity of situational interviews for administrative jobs. *Journal of Organizational Behavior, 11,* 69–76.

*Roth, P. L., & Campion, J. E. (1992). An analysis of the predictive power of the panel interview and pre-employment tests. *Journal of Occupational and Organizational Psychology, 65,* 51–60.

Rotter, J. B. (1966). Generalized expectancies for internal vs. external reinforcement. *Psychological Monographs, 80*(1, Whole No. 609).

Rynes, S. L., & Gerhart, B. (1990). Interviewer assessments of applicant "fit": An exploratory investigation. *Personnel Psychology, 43,* 13–35.

Sackett, P. R., & Dreher, G. F. (1982). Constructs and assessment center dimensions: Some troubling empirical findings. *Journal of Applied Psychology, 67,* 401–410.

Schmidt, F. L., & Hunter, J. E. (1992). Development of a causal model of processes determining job performance. *Current Directions in Psychological Science, 1,* 89–92.

Schmidt, F. L., & Hunter, J. E. (1998). The validity of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124,* 262–274.

Schmidt, F. L., & Rader, M. (1999). Exploring the boundary conditions for interview validity: Meta-analytic validity findings for a new interview type. *Personnel Psychology, 52,* 445–464.

*Sparks, C. P. (1974). *Validity of a selection program for commercial drivers* (PR 74-9). Unpublished manuscript.

*Sparks, C. P. (1977). *Selecting process operator and laboratory technician trainees* (PR 77-15). Unpublished manuscript.

*Sparks, C. P., & Manese, W. R. (1970). Interview ratings with and without knowledge of pre-employment test scores. *The Experimental Publication System, 4,* 142.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

*Spychalski, A. (1994). *A test of a model of employment interview information gathering.* Unpublished master's thesis, Rice University.

Sternberg, R. J., Conway, B. E., Ketron, J. L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology, 41,* 37–55.

Sternberg, R. J., & Kaufman, J. C. (1998). Human abilities. *Annual Review of Psychology, 49,* 479–502.

Stevens, M. J., & Campion, M. A. (1999). Staffing work teams: Development and validation of a selection test for teamwork settings. *Journal of Management, 2,* 207–228.

Taylor, H. R. (1949). [Review of the Cardall Test of Practical Judgment]. In O. K. Buros (Ed.), *The third mental measurements yearbook* (pp. 735–736). New Brunswick, NJ: Rutgers University Press.

Terman, L. M. (1918). The use of intelligence tests in the army. *Psychological Bulletin, 15,* 177–187.

Tett, R. P., Jackson, D. N., & Mitchell, M. (1991). Personality measures as predictors of job performance: A meta-analytic review. *Personnel Psychology, 44,* 703–742.

Thorndike, E. L. (1920). Intelligence and its uses. *Harper's Magazine, CXL,* 227–235.

Thorndike, E. L. (1921). *Educational psychology: The original nature of man* (Vol. 1). New York: Columbia Teacher's College.

Thurstone, L. L. (1938). Primary mental abilities. In *Psychometric monographs* (No. 1). Chicago: University of Chicago Press.

Ulrich, L., & Trumbo, D. (1965). The selection interview since 1949. *Psychological Bulletin, 63,* 100–116.

*U.S. Office of Personnel Management. (1987). *The structured interview.* Washington, DC: Office of Examination Development, Alternative Examining Procedures Division.

*Van Iddekinge, C. H., Roth, P. L., Huffcutt, A. I., & Eidson, C. E., Jr. (2000). *Structured interview ethnic group differences: Greater than we thought?* Manuscript submitted for publication.

Viswesvaran, C., Ones, D. S., & Schmidt, F. L. (1996). Comparative analysis of the reliability of job performance ratings. *Journal of Applied Psychology, 81,* 557–574.

Walsh, U. R. (1975). *A test of the construct and predictive validity of a structured interview.* Unpublished doctoral dissertation, University of Nebraska—Lincoln.

Walters, L. C., Miller, M. R., & Ree, M. J. (1993). Structured interviews for pilot selection: No incremental validity. *International Journal of Aviation Psychology, 3*(1), 25–38.

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

Wiesner, W., & Cronshaw, S. (1988). A meta-analytic investigation of the impact of interview format and degree of structure on the validity of the employment interview. *Journal of Occupational Psychology, 61,* 275–290.

*Wiesner, W. H., Latham, G. P., Bradley, P. J., & Okros, A. C. (1992, June). *A comparison of the situational and behavior description interviews in the selection of naval officers: Preliminary results.* Paper presented at the annual convention of the Canadian Psychological Association, Quebec City, Quebec, Canada.

Winograd, T. (1975). Frame representations and the declarative–procedural controversy. In D. Bobrow & A. Collins (Eds.), *Representation and understanding: Studies in cognitive science* (pp. 185–210). New York: Academic Press.

*Woodworth, D. G., Barron, F., & MacKinnon, D. W. (1957). *An analysis of life history interviewer's ratings for 100 Air Force captains* (Research Rep. No. AFPTRC-TN-57-129). Lackland Air Force Base, TX: Air Force Personnel and Training Research Center.

Wright, P. M., Lichtenfels, P. A., & Pursell, E. D. (1989). The structured interview: Additional studies and a meta-analysis. *Journal of Occupational Psychology, 62,* 191–199.

*Zedeck, S., Tziner, A., & Middlestadt, S. E. (1983). Interviewer validity and reliability: An individual analysis approach. *Personnel Psychology, 36,* 355–370.

Received July 7, 1999
Revision received October 26, 2000
Accepted October 26, 2000 ■