A Meta-Analytic Investigation of Cognitive Ability in Employment Interview Evaluations: Moderating Characteristics and Implications for Incremental Validity

Allen I. Huffcutt
Bradley University

Philip L. Roth
Clemson University

Michael A. McDaniel
University of Akron

The purpose of this investigation was to explore the extent to which employment interview evaluations reflect cognitive ability. A meta-analysis of 49 studies found a corrected mean correlation of .40 between interview ratings and ability test scores, suggesting that on average about 16% of the variance in interview constructs represents cognitive ability. Analysis of several design characteristics that could moderate the relationship between interview scores and ability suggested that (a) the correlation with ability tends to decrease as the level of structure increases; (b) the type of questions asked can have considerable influence on the magnitude of the correlation with ability; (c) the reflection of ability in the ratings tends to increase when ability test scores are made available to interviewers; and (d) the correlation with ability generally is higher for low-complexity jobs. Moreover, results suggest that interview ratings that correlate higher with cognitive ability tend to be better predictors of job performance. Implications for incremental validity are discussed, and recommendations for selection strategies are outlined.

Understanding of the validity of the employment interview has increased considerably in recent years. In particular, a series of meta-analyses has affirmed that the interview is generally a much better predictor of performance than previously thought and is comparable with many other selection techniques (Huffcutt & Arthur, 1994; Marchese & Muchinsky, 1993; McDaniel, Whetzel, Schmidt, & Maurer, 1994; Wiesner & Cronshaw, 1988; Wright, Lichtenfels, & Pursell, 1989). Moreover, these studies have identified several key design characteristics that can improve substantially the validity of the interview (e.g., structure).

However, much less is understood about the constructs assessed in interviews (Harris, 1989; Schuler & Funke, 1989). Scattered primary studies have suggested that

general factors such as motivation, cognitive ability, and social skills may be commonly captured. For example, Landy (1976) factor analyzed ratings on nine separate dimensions from a structured interview and found three general factors: manifest motivation, communication, and personal stability. Campion, Pursell, and Brown (1988) found a significant correlation between interview evaluations and a cognitive test battery. Schuler and Funke (1989) found that a multimodal interview that included vocational, biographical, and situational questions correlated highly with a social skills criterion. To date, there has been no summary level research in the literature to assess the extent to which these factors are evaluated and their consistency across interviews or their change with interview design (e.g., panel format or level of structure).

Understanding the constructs involved is potentially important. For one thing, there may be overlap between interviews and other selection approaches. The more similar the constructs, the greater the possibility that interviews may duplicate what could be accomplished with less costly paper-and-pencil tests (Dipboye, 1989; Dipboye & Gaugler, 1993; Harris, 1989). Furthermore, as Hakel (1989) noted, the incremental validity provided by interviews is a key issue in selection. In addition, understanding the constructs involved could lead to general

improvements in interview design, including better recognition of which constructs are most effective for particular jobs. Such improvements could ultimately raise the level of validity attainable with interviews.

The purpose of this investigation was to explore empirically the extent to which employment interview evaluations reflect cognitive ability. We felt that cognitive ability was a particularly important construct to study in relation to the interview for two reasons. First, no other construct measures have been shown to predict job performance as accurately or as universally. In addition, intelligence has become a very prominent social issue, as evidenced by publication of *The Bell Curve* (Herrnstein & Murray, 1994). We begin with a conceptual discussion of why one would expect interview evaluations to be saturated with cognitive ability. Then, we present and discuss five potential factors that may influence the strength of this relationship.

Cognitive Ability in Interviewer Evaluations

There are at least four reasons why interviewer evaluations are expected to reflect cognitive ability in a typical interview. First, intelligence may be one of a small number of issues that are highly salient in many interview situations. Research suggests that interviewers tend to base their judgments on a limited number of factors (Roth & Campion, 1992; Valenzi & Andrews, 1973). This is not surprising given general limitations in human information processing (see Solso, 1991). Moreover, it appears that the judgments behavioral observers make are typically guided by overall impressions (Kinicki & Lockwood, 1985; Srull & Wyer, 1989). Thus, many interviewers may be focusing on a limited number of general themes, such as whether the applicant has the appropriate background, can fit in with other employees, and is bright enough to learn the job requirements quickly. In turn, their general impressions along these themes are likely to have considerable influence on the final evaluations.

Second, applicants with greater cognitive ability may be able to present themselves in a better light than applicants with lower cognitive ability. Applicants clearly engage in impression management behaviors (Gilmore & Ferris, 1989), using techniques such as ingratiation, intimidation, self-promotion, exemplification, and supplication (Jones & Pittman, 1982; Tedeschi & Melburg, 1984). Those with greater cognitive ability may be better at knowing which strategies are most likely to succeed in that situation and when to back off from such strategies (see Barron, 1989). At a more general level, the link between impression management and cognitive ability is highlighted in a relatively new theory of intelligence (see Gardner & Hatch, 1989). In his theory, Gardner main-

tains that interpersonal skills, namely the capacity to discern and respond appropriately to other people, are one form of intelligence.

Third, at least some of the questions commonly asked in employment interviews could elicit ability-loaded responses. For example, questions of a more technical nature are likely to be answered more effectively by applicants with higher cognitive ability. These applicants probably are able to think in more complex ways and have a greater base of retained knowledge from which to work. Abstract questions may also be answered more effectively by applicants with higher cognitive ability. For example, two of the most frequently asked questions are, "What do you consider your greatest strengths and weaknesses?" and "What [college or high school] subjects did you like best and least?" (see Bolles, 1995). More intelligent applicants may be better at thinking through such questions and giving more desirable responses.

Fourth, cognitive ability may be indirectly captured through background characteristics. Intelligence, more so than any other measurable human trait, is strongly related to many important educational, occupational, economic, and social outcomes ("Mainstream science," 1995). Thus, on average, more intelligent people are likely to have more and better education, greater social and economic status, and better previous employment. Such information, whether it is reviewed before the interview or emerges during the interview, could influence interviewers' ratings in a favorable manner. For example, researchers (Dipboye, 1989; Phillips & Dipboye, 1989) have found that preinterview information can strongly influence both the interview process and subsequent ratings. In general, the more influence background information has on the ratings, the greater the chance that these ratings will reflect cognitive ability.

In summary, it appears that a number of mechanisms by which interview ratings can become saturated with cognitive ability. One of these mechanisms, interviewer evaluation of applicants' ability to learn job requirements quickly, represents a relatively direct measurement of the ability construct. Each of the other three mechanisms represents more of an indirect influence from ability. In particular, cognitive ability influences applicant behavior during the interview, generation of responses, and background characteristics, all of which in turn influence the final ratings. It should also be noted that these four mechanisms are not mutually exclusive in that more than one may be operating in a given situation.

Potential Moderators of the Interview–Ability Correlation

The level of structure is a widely researched characteristic of interview design. The two most prominent aspects

of interview structure are standardization of the questions and standardization of the evaluative criteria (Huffcutt & Arthur, 1994). Standardizing the questions effectually establishes a common content across all of the interviews, because interviewers are no longer free to ask whatever questions they wish. Typically, this content is based on some type of job analysis (Campion et al., 1988). Standardizing the evaluative criteria focuses the evaluations more directly on the actual content, thereby reducing the influence of global impressions and resulting in more relevant and differentiated ratings.

Whereas these aspects of structure increase reliability and validity (Conway, Jako, & Goodman, 1995; Huffcutt & Arthur, 1994), their effect on the constructs assessed is less clear. On one hand, they may actually decrease the correlation with ability because they limit the influence of background information, general impressions of intelligence, and possibly impression management behaviors as well. On the other hand, because the questions are used consistently across all applicants, there is the potential for a fairly high correlation if that type of question does load on ability. In short, a structured interview does not automatically imply that certain constructs such as ability will be reflected in the ratings. Rather, the above line of reasoning suggests that the correlation with ability could be either higher or lower than that typically found in unstructured interviews, depending on the type of question used. Moreover, the above reasoning also suggests that even if the correlation is somewhat similar, structured interviews are likely to reflect ability for different reasons than unstructured interviews. Namely, the nature of the questions is likely to be the dominant factor with structured interviews, whereas other factors such as impression management behaviors and background characteristics probably assume a more prominent role with unstructured interviews.

In fact, several different types of questions have been used in structured interviews. For example, a *situational interview question* (Latham, Saari, Pursell, & Campion, 1980) involves presenting applicants with a hypothetical job-related situation, to which they must indicate how they would respond. A *behavior description interview question* (Janz, 1982) involves asking applicants to describe some real situation from their past relevant to the job for which they are applying. Other types have been used as well, including job knowledge, job simulation, and worker requirement questions (Campion et al., 1988).

In comparison, there may be important differences among the various types of questions in terms of how much they assess cognitive ability. For instance, situational interview questions are thought to load more heavily on verbal and inductive reasoning abilities than behavior description questions (Janz, 1989) because ap-

plicants most likely must think through and analyze each situation. Job knowledge questions should also load somewhat on ability because causal analyses of performance ratings have suggested that the most direct effect of intelligence is on acquisition of job knowledge (Borman, White, Pulakos, & Oppler, 1991; Schmidt, Hunter, & Outerbridge, 1986).

In summary, the first potential moderator of the interview–ability correlation is the level of structure. Structured interviews tend to be more reliable than unstructured interviews and could correlate more highly with an ability measure because of this psychometric advantage. However, after accounting for differences in reliability, we predicted that there would be little or no difference in ability saturation across levels of structure. The basis for our prediction was that looking at structure by itself tends to collapse across content, which should make the extent of ability in the evaluations at least somewhat similar.

The second potential moderator is the content of the questions. We predicted that differences in the magnitude of the interview–ability correlation would begin to emerge when interviews at various levels of structure were further broken down by content. For high-structure interviews, we predicted that interviews containing situational and/or job knowledge questions would be more saturated with ability than interviews that were based on other types of questions such as past behavior. For low-structure interviews, we similarly predicted that asking hypothetical and/or technical questions would increase the extent to which evaluations reflect ability. (Although, as noted below, we were not able to test low-structure interviews because of a lack of information regarding their content.)

Third, allowing interviewers access to cognitive ability test scores may influence the extent to which their ratings reflect ability. As noted above, preinterview information can have considerable influence on both the interview process and postinterview evaluations (Dipboye, 1989; Phillips & Dipboye, 1989). Seeing ability test scores may cause interviewers to form an early impression of applicants' intellectual skills. In turn, this could affect their final ratings either directly or through impression-confirming behavior during the interview. Alternately, access to ability scores may simply make the general issue of intellectual capability even more salient, thus increasing the focus on it during the interview. We predicted that making ability test scores available would increase the degree to which interviewer evaluations reflect cognitive ability, at least for low-structure interviews. With high-structure interviews, the constraints on the questions and the evaluation process could minimize any influences that arise from seeing ability test scores.

Fourth, the interview–ability correlation may vary ac-

cording to the complexity of the job for which the applicants are applying. Jobs differ widely in complexity, and jobs of greater complexity generally require a higher level of mental skills ("Mainstream science," 1995). Such a tenet is supported empirically by the finding that the validity of ability tests tends to increase with the level of complexity (Gandy, 1986; Hunter & Hunter, 1984; McDaniel, 1986). Therefore, it is posited in an interview that the interviewers recognize the increased necessity for such skills with more complex jobs, and they place more emphasis on their assessments. Such a tenet assumes that interviewers not only recognize a job as being of high complexity but also successfully incorporate ability into their assessments. In general, we predicted that the correlation with ability would be greater with interviews for more complex jobs.

Finally, there may be an association between the magnitude of the interview–ability correlation and the magnitude of the validity coefficient (i.e., the correlation between interview ratings and job performance). Research suggests that cognitive ability is a strong and consistent predictor of job performance (Hunter & Hunter, 1984). Accordingly, it can be argued that the more saturated the ratings are with ability, the higher the resulting validity of those ratings is likely to be. Alternately, it can be said that interviews become more valid when they capture cognitive ability. Thus, we predicted that in general, ratings that correlate more highly with cognitive ability should be more valid predictors of job performance. Such a prediction is relevant to all jobs regardless of complexity because cognitive ability is still a valid predictor, even for low-complexity jobs.

Method

Search for Primary Data

We conducted an extensive search for interview studies that reported a correlation between interview ratings and some type of cognitive ability test. Datasets from previous meta-analyses were prime sources for locating studies (Huffcutt & Arthur, 1994; McDaniel et al., 1994; Wiesner & Cronshaw, 1988). Supplemental inquiries were also made of prominent researchers in the interview area in order to obtain any additional studies not included in the above datasets.

Two main criteria were used in deciding which of the studies reporting an interview–ability correlation would be retained. First, the interview had to represent a typical employment interview. Eight studies did not meet this criteria and were excluded. Three of these involved a procedure known as an *extended interview*, where the interview is combined with several assessment center exercises (Handyside & Duncan, 1954; Trankell, 1959; Vernon, 1950). Two studies used objective biographical checklists rather than true interviews (Distefano & Pryer, 1987; Lopez, 1966). One interview was designed deliberately to induce stress (Freeman, Manson, Katzoff, & Pathman, 1942).

Finally, in two studies interviews were used as an alternate method to assess job proficiency (Hedge & Teachout, 1992; Ree, Earles, & Teachout, 1994). Second, a study had to provide sufficient information to allow coding on a majority of the five moderator characteristics. Three studies did not report sufficient information and were dropped (Conrad & Satter, 1945; Darany, 1971; Friedland, 1973). In total, we were able to locate 49 usable studies after application of the above decision rules, with a total sample size of 12,037.

Such a dataset is notable given the general difficulty in finding studies that report correlations among predictors (Hunter & Hunter, 1984). These studies represented a wide range of job types, organizations, subjects, and interview designs. Sources for the studies were similarly diverse and included including journals, unpublished studies, technical reports, and dissertations. Thus, we were reasonably confident that these studies represented a broad sampling of employment interviews.

As expected, there was considerable diversity among the ability measures to which interview ratings were correlated. To better understand the ability measures used, we compiled some summary statistics. Of the 49 studies in our dataset, 11 (22.4%) used a composite test such as the Wonderlic Personnel Test (Wonderlic, 1983) where various types of ability-loading questions (e.g., math, verbal, and spatial) are combined into one test. In 31 of the studies (63.3%), separate subtests of individual factors were administered and these scores were then combined to form a composite. Lastly, in 7 of the studies (14.3%), separate subtests of individual factors were administered, but no ability composite was formed.

In general, we felt that the first two categories of tests listed above were all reasonable (albeit not perfect) measures of general cognitive ability. In the first category, the test itself was a composite of individual factors, and in the second category, a composite was formed from individual subtests. We had some concern about the third category because no composite was formed. However, eight studies from the second category reported ability correlations with individual factors, as well as with the composite ability measure. Our analysis of these eight studies suggested that the highest individual correlation was a fairly accurate estimate of the composite correlation. In particular, the highest individual correlation from these studies correlated .98 ($p < .0001$) with the composite correlation. Thus, we took the highest individual correlation in these seven studies.

Coding of Study Characteristics

Level of interview structure was coded with a variation of the framework developed by Huffcutt and Arthur (1994). They identified four progressively higher levels of question standardization, which Conway et al. (1995) later expanded to five. They also identified three progressively higher levels of response evaluation. We combined various combinations of these two aspects of structure into three overall levels corresponding to low, medium, and high. Studies were classified as low structure if there were no constraints or very limited constraints on the questions and the evaluation criteria. Studies were classified as medium structure if there were a higher level of constraints on the questions and responses were evaluated along a set of clearly defined dimensions. Finally, studies were classified as high structure if

there were precise specifications in the wording and in the number of questions without variation or with very limited flexibility to choose questions and probe, and the responses were evaluated individually by question or along multiple dimensions.

We also attempted to code both medium- and high-structure interviews by type of question or content. Studies were classified as *situational* if most or all of the questions involved presentation of job-related scenarios to which the applicants indicated how they would respond. Similarly, studies were coded as *behavior description*[1] if most or all of the questions involved asking applicants to describe actual situations from their past that would be relevant to the job for which they were applying. Although a number of studies clearly fell into one of these two categories, there were two studies that used more than one type of question. Campion, Campion, and Hudson (1994) had both a situational section and a past behavior section. Furthermore, Campion et al. (1988) used a combination of situational, job knowledge, job simulation, and worker requirement questions. In the former case, we coded the two sections as separate studies because the two types of questions were not mixed (i.e., one section was given and then the other), and separate information on the correlation with ability was provided. In the latter case, the question types were mixed, and separate information was not provided. Harris (1989) denoted this mixture of four question types as a comprehensive structured interview. We coded this study as comprehensive but did not analyze it as a separate content category because there was only one of its type.

Although it is obviously possible for studies that use a particular type (or a combination) of questions to be of medium structure, most tend to be of high structure. The most likely exception is with behavior description studies because the design allows for considerable interviewer discretion (Janz, 1982). In our dataset, all of the studies for which we could make a content classification were of high structure, including the behavior description ones. Consequently, for the medium-structure studies, we attempted to make a simple dichotomous classification. Specifically, studies were coded as to whether at least some technical, problem-solving, or abstract reasoning (e.g., situational) questions were systematically included.

In effect, content was a nested variable operating under structure in this investigation, in that it had different categories at different levels of structure (see Keppel, 1991, for a discussion of nested variables.) Such nesting did not present a problem because our initial hypothesis was that differences in ability saturation would emerge when interviews at various levels of structure are further broken down by content. No distinction of content was made with low-structure interviews because information regarding the types of questions asked by the interviewers was generally not provided.

Availability of ability test information was coded as a dichotomy. Specifically, studies were coded as to whether interviewers had access to cognitive ability test scores at any time during the interview process.

Job complexity was coded with a three-level framework developed by Hunter, Schmidt, and Judiesch (1990). This framework is based on ratings of "Data and Things" from the *Dictionary of Occupational Titles* (U.S. Department of Labor, 1977) and is a modified version of Hunter's (1980) original job complexity system. Specifically, unskilled or semiskilled jobs such as truck driver, assembler, and file clerk were coded as *low complexity*. Skilled crafts, technician jobs, first-line supervisors, lower level administrators, and other similar jobs were coded as *medium complexity*. Finally, jobs such as managerial, professional, and those involving complex technical set-up were coded as *high complexity*. As Gandy (1986) noted, job complexity classifications essentially reflect the information-processing requirements of a position, and they do not capture the complexities relating to interactions with people.

The validity coefficient of the interview was recorded as reported in the studies and were uncorrected for artifacts. We included only coefficients involving job performance criteria because mixing performance and training criteria did not seem appropriate, and there were too few training coefficients to do a separate analysis. Also, coefficients representing overall ratings on both the interview and performance criteria were preferred. If not presented, the coefficients for the individual dimensions were averaged. In cases where multiple performance evaluations were made, typically in situations where more than one appraisal instrument was used, the resulting validity coefficients were averaged.

Whereas the above five factors constituted the independent (i.e., moderator) variables, the dependent variable in this investigation was the degree to which ability was reflected in the interview ratings. We recorded the observed (uncorrected) correlation between interview ratings and ability test scores. Preference was given to correlations involving overall interview ratings rather than individual dimensions and to correlations involving a composite ability score rather than individual ability factors. As noted above, when correlations were reported only for individual ability factors, we took the highest individual correlation as an estimate of the correlation with the ability composite.

As expected, some of the studies did not report enough information to make a complete coding on all of the above characteristics. In these cases, we made a concerted attempt to contact the authors directly to obtain further information. Although this was not always possible, we did manage to reach a number of them and were able to make additional codings. In total, we were able to code all 49 (100%) of the studies for level of struc-

---

[1] We intended the behavior description label to be somewhat general. Janz (1982) actually called his original format the *patterned behavioral description interview* because interviewers could choose selectively from patterns of questions established for each dimension and probe applicant responses freely. Several later studies used the same type of question but with higher levels of constraint, and renaming the format in the process. For example, Pulakos and Schmitt (1995) and Motowidlo et al. (1992) completely standardized the questions, with the former being called an *experienced-based interview* and the latter being called a *structured behavioral interview*. The key factor for classification in our behavior description category was not the extent of constraints (i.e., medium versus high structure), but rather that the questions requested information about past real situations that would be relevant to the job. However, in this meta-analysis, all of the studies with situational behavior description and comprehensive content were of high structure.

ture, 46 (94%) for availability of ability scores, 49 (100%) for job complexity, 41 (84%) for the validity coefficient, and 49 (100%) for the interview–ability correlation. Regarding content, we were able to code 18 high-structure studies as having situational, behavior description, or comprehensive content, and 19 medium-structure studies as containing some or no deliberate cognitive content. Thus, in total we were able to code 37 studies for content (76%). Overall, these results indicate that with the additional information obtained, we were able to code most studies on most variables.

To ensure accuracy of coding, we independently coded every study on the above characteristics. Differences were then investigated and resolved by consensus. In select cases, the authors of the study were contacted for verification. The correlations between our initial set of ratings (before consensus resolution) were .90 for structure, .71 for availability of ability scores, 1.00 for interview content, .89 for job complexity, .98 for the validity coefficient, and .95 for the interview–ability correlation ($p <$ .001). The somewhat lower correlation for availability of ability scores was due to the same coding difference on a set of three studies from one researcher, who was subsequently contacted for verification. Excluding these three studies, the correlation was .90. In summary, these data suggest that the coding process was reliable across raters. Moreover, the above numbers are underestimates of the true reliability of the final dataset because all differences were investigated and resolved.

Preliminary Assessment of the Variables

Before conducting the actual analyses, we did some preliminary evaluation of the study variables. The first purpose of these evaluations was to ensure that all variables had sufficient variability to allow meaningful analyses because low variability would restrict assessment of a variable and reduce the power to find a true effect. In the case of the four variables comprised of distinct levels or categories (i.e., structure, content, ability score availability, and job complexity), we compiled the number and percent of data points at each level or category. Distributions for these variables are shown in Table 1. As indicated in Table 1, the variables in general appeared to have adequate variability. Availability of ability test scores had the most nonuniform distribution, where scores were withheld from interviewers much more often than they were made available. Although, as noted below, the total sample size for the six studies where scores were made available was 2,455.

For the two variables that were continuous, namely the interview–ability correlations and the performance validity coefficients, we computed simple statistics. The interview–ability correlations ranged from −.09 to .74 ($M = .25, SD = .19$). The validity coefficients ranged from .00 to .51, ($M = .27, SD = .12$). Both of the continuous variables appeared to have acceptable variability.

The second purpose of the preliminary analyses was to determine whether the five moderator variables were reasonably independent of each other. Intercorrelation (i.e., multicollinearity) among these variables would make it difficult to isolate the individual effects of the variables involved and could necessitate modifications to the approach used for the analyses

Table 1
Preliminary Assessment of the Moderator Variables: Analysis of Variability

Variable and level	Number	Percentage
Structure		
Low	8	16.3
Medium	19	38.8
High	22	44.9
Content, high structure		
Situational	10	45.5
Behavior description	7	31.8
Comprehensive	1	4.5
Other	4	18.2
Content, medium structure		
No cognitive	14	73.7
Some cognitive	5	26.3
Ability score availability		
No	40	81.6
Yes	6	12.2
Unknown	3	6.1
Job complexity		
Low	13	26.5
Medium	24	49.0
High	12	24.5

Note. Percentages for content categories are based on 22 studies for high structure and 19 studies for medium structure, respectively. Percentages for all other variables are based on 49 studies.

of the interview–ability correlations. Multicollinearity was assessed through both formation of a matrix of simple correlations and by computation of the variance inflation factor (VIF) for each variable (Freund & Littell, 1991). The intercorrelation matrix gives the simple correlations among the variables, whereas the VIF's indicate the overall overlap between one independent variable and all of the other independent variables. As Myers (1990) noted, VIF values exceeding 10 indicate serious collinearity. Results are presented in Table 2, which shows no correlation above .5, and only one correlation was in the .4 range, with structure and ability score availability correlating −.46. Such a correlation is not surprising because many of the high-structure techniques routinely withhold test information. Availability of ability scores also correlated −.29 with job complexity, indicating a tendency for scores to be made available more often for lower complexity jobs. The VIFs in Table 2 were all of a relatively low magnitude, with none reaching the critical level suggested by Myers (1990).

In summary, it appears that collinearity was not a serious problem in this investigation. Neverthess, we took a closer look at availability of ability test scores because it was involved in the two highest simple correlations and had the highest VIF. Of the six studies where ability scores were made available, three were of low structure and three were of medium structure; four were of low complexity, and one each of medium complexity and high complexity. Although such collinearity is not of sufficient concern to warrant modifying the planned analyses, we decided to conduct supplementary analyses for structure and job complexity by removing the studies where test scores were made available.

Lastly, it should be noted that content was not included as a

Table 2
Preliminary Assessment of the Moderator Variables: Analysis of Multicollinearity

Variable	VIF	n	1	2	3	4
1. Level of structure	1.35	48	—	−.46	.05	.11
2. Availability of ability scores	1.46	45		—	−.29	.05
3. Job complexity	1.11	48			—	−.10
4. Validity coefficient	1.01	40				—

Note. VIF = variance inflation factor, a measure of multicollinearity among the predictor variables.

variable in the multicollinearity assessment for two reasons. First, it involved categories rather than levels. Therefore, regression analysis would not have been as appropriate. In addition, as noted above, content represented a nested variable rather than a separate moderator variable in and of itself, having different categories at different levels of structure.

Meta-Analysis for the Average Level of Ability Saturation

We first attempted to estimate the average level to which interview ratings reflect cognitive ability, collapsing across all study types, designs, and characteristics. We started by computing the sample-weighted mean of the observed (i.e., uncorrected) correlations between interview ratings and ability test scores. Computations were performed with a statistical analysis software (SAS) program developed by Huffcutt, Arthur, and Bennett (1993).

We used a modified version of sample weighting in this investigation because of our concern about a handful of studies dominating the analysis. For example, two relatively old military studies (Reeb, 1969; Rhea, Rimland, & Githens, 1965) combined would have contributed over 26% to the overall sample size. Such reliance on a few studies goes counter to the basic logic of meta-analysis, which is to avoid placing too much emphasis on any individual study (Hunter & Schmidt, 1990). Moreover, because of the sparse information on artifacts reported in interview studies, an unusually large but unreported level of an artifact such as range restriction in one of the larger sample studies could have skewed the results considerably.

Accordingly, we used a 3-point weighting system for our analyses. Specifically, studies were weighted as 1 if the sample size was 75 or less, 2 if the sample size was between 75 and 200, and 3 if the sample size was 200 or more. The number of studies at each of the three weights was 14, 18, and 17, respectively. Such a weighting scheme retained the general notion that larger samples are more credible than smaller samples, but it also ensured that no one study contributed any more than three times any other study to the results.[2]

After computing the sample-weighted mean of the observed correlations, we corrected it for range restriction in the interview by using the artifact distribution approach (Hunter & Schmidt, 1990). Because there was not enough information provided in the studies to form an artifact distribution for range restriction, we used data from a larger interview meta-analysis. Specifically, we used the mean range restriction ratio of .74

found by Huffcutt and Arthur (1994). This value is very similar to the mean range restriction ratio of .68 reported by McDaniel et al. (1994) in their interview meta-analysis, albeit slightly more conservative. Squaring the resulting corrected mean correlation provided an estimate of the average proportion of variance in observed (and unrestricted) interview ratings that was common with ability test scores.

Then, we further corrected the mean correlation for measurement error in the ability tests by using an average reliability of .90, a value that seemed reasonably representative of these tests in general (see Wechsler, 1981; Wonderlic, 1983). Squaring the mean correlation then yielded an estimate of the average proportion of variance in observed, unrestricted interview ratings that represented the construct cognitive ability. Finally, we corrected for measurement error in interview ratings. Wiesner and Cronshaw (1988) found average reliabilities of .61 and .82 for studies with low- and high-structure respectively. Therefore, we used .715 as the average interview reliability (i.e., the mean of the two values). Squaring the resulting mean correlation provided an estimate of the average proportion of variance in true interview ratings (i.e., unrestricted and without measurement error) that represented the construct cognitive ability (see Kaplan & Saccuzzo, 1993, who discussed true versus observed test scores).

Additionally, we attempted to determine the stability of the interview–ability relationship across different interviews. We started by computing the sample-weighted variance in the observed interview–ability correlations. Then, we computed the variance attributable to sampling error and added to this the variance attributable to study-to-study differences in the level of range restriction, again with the artifact data from Huffcutt and Arthur (1994). The variances from these two artifact sources were then totaled, divided by the observed variance, and multiplied by 100. The result was the estimated percentage of variance in the interview–ability correlations that was attributable to artifacts. As Hunter and Schmidt (1990) noted, the likelihood of other variables moderating a relationship is fairly low if at least 75% of the variance is accounted for by artifacts. However, our estimate of the percentage of variance accounted for was probably conservative because at least some variance may have resulted from use of different ability tests and different ability factors.

Meta-Analyses for the Five Moderator Variables

For each moderator variable, we sorted the studies into the various categories of that variable and conducted a separate meta-analysis for each with the procedures and values described above. The following caveats should be noted. First, for interview structure and content, we corrected the low-, medium-, and high-structure studies for measurement error in the interview separately by using the more precise estimates from Wiesner and Cronshaw (1988). Specifically, we used .61 for low structure, .715 (the average) for medium structure, and .82 for high structure. These corrections were more accurate than uniformly using the average value, and with regard to structure, this allowed us to analyze whether there really were differences in ability saturation across levels of structure after accounting for differences in reliability.

Second, it was also necessary to categorize the uncorrected validity coefficients because they were continuous. Muchinsky (1993) noted that a validity coefficient of .30–.40 is desirable in validation studies. Accordingly, we categorized studies as *high validity* if their coefficient was .30 or higher. The remainder of the studies were classified either as *low* or *medium validity*. In particular, we classified studies as low if their validity coefficient was less than .20 and medium if their validity coefficient was between .20 and less than .30. In total, 13 studies were classified as low, 9 as medium, and 19 studies as high. The remaining 8 studies did not report a validity coefficient.

In regard to the validity coefficients, it is important to note that at least a few of the studies in the two lower validity categories may have been inadvertently classified because of artifacts. Specifically, unusually high levels of an artifact such as range restriction or criterion unreliability could have made a study with a higher level of validity appear to have a lower level. For example, a study with high validity in reality could have ended up in the medium- or even the low-validity category. Inclusion of such studies in the lower validity categories would tend to dilute any underlying differences in the interview–ability correlation. Thus, our results may have slightly underestimated true differences in ability saturation among levels of structure.

Lastly, we decided to do supplementary analyses for structure and job complexity. To remove any possible confound from availability of ability test scores, we reran the analyses as described above after removing the six studies where ability scores were made available to the interviewers.

Results

The mean sample-weighted correlation between interview ratings and cognitive ability test scores across all 49 studies was .25. Correcting for range restriction in the interview scores increased the mean correlation to .32, indicating that approximately 10.2% of the variance in the observed, unrestricted interview ratings was common with ability test scores. Correcting for measurement error in the ability tests further increased the mean correlation to .34, indicating that approximately 11.6% of the variance in observed, unrestricted interview ratings reflected the construct cognitive ability. Lastly, correcting for mea-

surement error in the interview resulted in a mean correlation of .40, indicating that approximately 16.0% of the variance in interviews (at a true score level) represented the construct cognitive ability.

The sample-weighted variance across the observed interview–ability correlations was .033676. Variance from sampling error was estimated to be .003610, whereas variance from study-to-study differences in range restriction was estimated to be .002072. Combined, these two artifacts accounted for only 16.9% of the variance in the observed correlations. Thus, it appeared that other variables moderated the extent to which ability was reflected in the interview ratings.

Results for the five moderator variables are summarized in Table 3, where Contrary to our prediction of little or no difference, the interview–ability correlation is shown to decrease as the level of structure increases. After final correction for interview reliability, the interview–ability correlations were .52, .40, and .35, respectively, for low, medium, and high structure. Therefore, the percentage of variance in true interview ratings representing the construct cognitive ability was 27.0, 16.0, and 12.3, respectively. The same inverse relationship was found with observed interview ratings (i.e., without correction for range restriction, test reliability, or interview reliability), although, as expected, the magnitude of the differences were smaller.

Consistent with our prediction, differences in ability saturation emerged when interviews at various levels of structure were further broken down by content. For high-structure interviews, situational interviews correlated more highly with ability than behavior description interviews. The fully corrected correlations were .32 and .18, respectively; the corresponding percentages of variance associated with cognitive ability were 10.2 and 3.2, respectively. For medium-structure interviews, deliberate inclusion of at least some cognitive content did appear to increase reflection of ability in the ratings, although the magnitude of the difference was not nearly as pronounced as for high-structure interviews.

As we predicted, making ability test scores available to interviewers appeared to increase the extent to which ability was reflected in their evaluations. After all corrections, the interview–ability correlation was .38 when scores were not made available and .59 when they were made available. Therefore, the corresponding percentages of variance in true interview ratings representing the construct cognitive ability were 14.4 and 34.8, respectively. These findings may relate primarily to low- and medium-structure interviews and to low-complexity jobs.

Similar to the results for structure and contrary to our prediction, we discovered that the extent of ability saturation was inversely related to job complexity. Interviews

Table 3
Assessment of Cognitive Ability in the Interview

Analysis	No. of study coefficients	Total sample size	Estimated mean correlation				Observed variance (%)[a]
			obs	rr	rr-t	true	
Overall	49	12,037	.25	.32	.34	.40	16.9
Structure							
Low	8	3,147	.30	.38	.40	.52	9.7
Medium	19	3,974	.25	.32	.34	.40	44.3
High	22	4,916	.23	.30	.31	.35	13.9
Content, high structure							
Situational	10	1,463	.21	.27	.29	.32	67.9
Behavior description	7	1,881	.12	.15	.16	.18	16.3
Content, medium structure							
No cognitive	14	2,490	.23	.30	.32	.38	40.1
Some cognitive	5	1,484	.28	.36	.38	.45	69.6
Ability score availability							
No	40	7,185	.23	.30	.32	.38	21.2
Yes	6	2,455	.37	.47	.50	.59	15.3
Job complexity							
Low	13	3,195	.36	.46	.49	.58	22.9
Medium	24	7,375	.20	.27	.29	.34	27.8
High	12	1,467	.19	.25	.27	.32	16.1
Validity							
Low	13	4,786	.19	.25	.26	.31	27.7
Medium	9	2,154	.17	.23	.24	.29	44.7
High	19	3,679	.35	.45	.47	.56	12.6

Note. obs = correlation between interview ratings and test scores; rr = correlation corrected for range restriction in the interview; rr-t = correlation corrected for range restriction in the interview and measurement error in the ability tests; true = correlation further corrected for measurement error in the interview.
[a] Accounted for by artifacts of sampling error and study-to-study differences in level of range restriction.

for low-complexity jobs had the highest interview–ability mean correlation, whereas the mean correlation for medium-complexity jobs was slightly higher than that for high-complexity jobs. In particular, the final mean correlations were .58, .34, and .32, respectively, indicating that the percentages of variance in true interview ratings representing the construct cognitive ability were 33.6, 11.6, and 10.2, respectively.

As predicted, interviews with high criterion-related validity tended to have more cognitive ability reflected in their ratings. The final mean correlation for the high-validity studies was .56, indicating that on average about 31.4% of the variance reflected ability. The mean interview–ability correlations for low- and medium-validity studies were .31 and .29, respectively and the corresponding percentages of variance were 9.6 and 8.4, respectively. It is surprising that the mean correlation for the medium-validity studies was slightly lower than that for low-validity studies, as it was expected to fall between the low and high means. Follow-up analysis suggested one possible explanation. The six studies where ability test scores were made available were split between the low- and high-validity categories, with none falling in the medium-validity category. Because making these scores available does appear to increase saturation of ability, the lack of such studies in the medium category may have

reduced the interview–ability correlation relative to the other two categories. The reason our preliminary analyses for multicollinearity among the moderator variables did not detect this situation is that the relationship between validity and test score availability was nonlinear (i.e., high, low, high).

In terms of variability, the percentage of variance accounted for by artifacts tended to increase somewhat when studies were broken down by the five moderator variables. Across all 15 moderator categories (see Table 3), the average percentage of variance accounted for was 30.0, considerably higher than the initial value of 16.9 for the overall analysis. The percentage of variance accounted for was the highest for content, confirming its important role in moderating the extent to which interview evaluations reflect cognitive ability.

However, few of the moderator categories reached the commonly cited level at which the presence of other moderator variables can largely be ruled out (i.e., the 75% rule). In general, we did not expect the percentage of variance accounted by artifacts to be particularly high for any one moderator variable, because looking at one variable collapses across all of the other variables. Of course, the ideal solution would be to group studies by combinations of moderator characteristics (e.g., high structure or low complexity) and then reanalyze the per-

Table 4

Supplemental Analyses of Cognitive Ability Assessment With Studies Removed Where Ability Scores Were Made Available

Analysis	No. of study coefficients	Total sample size	Estimated mean correlation				Observed variance (%)[a]
			obs	rr	rr-t	true	
Structure							
Low	5	2,232	.25	.32	.34	.44	11.9
Medium	16	2,434	.22	.28	.30	.36	65.3
High	22	4,916	.23	.30	.31	.35	13.9
Job complexity							
Low	9	1,191	31	.40	.42	.50	32.4
Medium	23	6,955	21	.28	.30	.35	29.0
High	11	1,436	18	.24	.26	.30	14.8

Note. obs = correlation between interview ratings and test scores; rr = correlation corrected for range restriction in the interview; rr-t = correlation corrected for range restriction in the interview and measurement error in the ability tests; true = correlation further corrected for measurement error in the interview.
[a] Accounted for by artifacts of sampling error and study-to-study differences in level of range restriction.

centage of variance accounted for by artifacts. Given the relatively large number of moderator variables in this investigation, we could not have done this easily because there would be a small number of studies in many of the combinations.

However, we did find a few combinations with enough studies to do at least somewhat meaningful analyses. The percentage of variance accounted for in these combinations was generally much higher. For example, there were 11 studies that had medium structure, medium complexity, and no availability of test scores. The percentage of variance accounted for was 50.3 in this grouping. With further breakdown by content, the percentage of variance accounted for across the studies with no deliberate cognitive content was 85.4 and 69.6 across the studies with some cognitive content. Similarly, there were 10 studies that had high structure, medium complexity, and no availability of test scores. The percentage of variance accounted for was 58.0 in this grouping. With further breakdown by content, the percentage of variance accounted for was 100.0 in the situational studies and 93.2 in the behavior description studies.

Results of the two supplemental analyses are presented in Table 4. Comparison with results in Table 3 suggests that the overall pattern of ability saturation did not change. Both structure and job complexity still had inverse relationships with ability, even after removal of the studies where ability scores were made available. However, for both variables, there was a noticeable decrease in the magnitude of the differences among levels. For example, the difference in the final interview–ability correlation between low and high structure was originally .17 (.52–.35), which dropped to .09 (.44–.35) after removal of the studies where ability scores were made available. Similarly, the difference between low and high complexity dropped from .26 (.58–.32) to .20 (.50–.30). Thus,

overall conclusions did not change, but the collinearity between ability score availability and structure and job complexity nevertheless appeared to have at least some influence on the results. In general, these findings underscore the importance of assessing intercorrelations among the study variables when conducting a meta-analysis, especially because such intercorrelations are not uncommon and could be much higher than those found here.

Moreover, the supplementary analysis for structure provided an additional insight into the interview–ability relationship. Although we had predicted that making ability test scores available would increase the interview–ability correlation, we also had noted that the effect of seeing these scores could diminish at higher levels of structure from the constraints on the questions and the evaluation process. Eliminating the studies where scores were available changed the final mean interview–ability correlation for low structure from .52 to .44, a difference of .08. The average interview–ability correlation for medium structure changed from .40 to .36, a difference of only .04. In short, removing the studies where ability scores were available had less effect on medium-structure studies than on low-structure studies. Overall, this finding may suggest that influences from seeing test scores decrease as the level of structure increases.

Discussion

Results of this investigation confirm that employment interview evaluations tend to reflect cognitive ability. The mean corrected interview–ability correlation was .40, suggesting that on average about 16% of the variance in interview constructs represents cognitive ability. Such a level is high enough to suggest that correlation with ability should be a consideration in interview design. It also

highlights the importance of continued research into other constructs such as interpersonal skills and motivation. Construct validation could very well be the next major thrust of interview research, as empirical validity has been for the last decade or so.

The findings here are important because it appears that interviews which correlate higher with ability also tend to be better predictors of job performance. And, this relationship appears to generalize across level of structure. That is, low-structure interviews with more ability saturation tend to have higher validity than low-structure interviews with less saturation, medium-structure interviews with more ability saturation tend to have higher validity than those with less saturation, and high-structure interviews with more ability saturation tend to have higher validity than those with less saturation. Such a finding is not surprising given the large body of research indicating that mental ability tests are one of the best and most consistent predictors of job performance (Hunter & Hunter, 1984).

We identified several characteristics that appear to influence the extent to which interview ratings reflect cognitive ability. For example, making ability test scores available to interviewers tends to increase the extent to which ability is reflected in their evaluations, particularly for low-structure interviews. However, we found less of an effect for medium-structure interviews and were not able to test the effect on high-structure interviews.

Another characteristic that appears to affect the extent to which ability is captured in interview evaluations is job complexity. Somewhat surprisingly, on average we found that ratings for low-complexity jobs were considerably more saturated with ability than those for medium- or high-complexity jobs. Such a finding might appear to present a paradox because prior research (Hunter & Hunter, 1984) suggested that cognitive skills become increasingly important as the level of complexity increases. However, it is important to note that these results concerned degree of association rather than mean levels. In particular, they suggested that interview ratings for low-complexity jobs were much more closely linked to cognitive ability (i.e., higher ratings and higher scores; lower ratings and lower scores), not that applicants for low-complexity jobs as a whole scored better on the ability tests.

Thus, although cognitive skills most likely do become increasingly important with higher complexity jobs, intellectual differences among applicants are not as strongly reflected in the interview ratings. Interviewers could simply be assuming that most applicants for complex jobs already possess a respectable level of intelligence; therefore, they focus their evaluation on other areas such as motivation and interpersonal skills. It is also possible that differences in intelligence are much less apparent among applicants for more complex jobs, either because of a narrower range or to better impression management skills. For example, many jobs of higher complexity require a fairly high level of education. This tends to screen out applicants with lower cognitive ability. In any respect, preliminary analyses of the study variables suggest that our findings for job complexity cannot be attributed to differences in the level of structure among the levels of job complexity. Additionally, our supplementary analysis suggests that the inverse relationship between complexity and ability holds up even after the tendency for ability scores to be made available to interviewers more often for low-complexity jobs is taken into account.

Also surprising were our results for interview structure. Some researchers have suggested that structure increases ability saturation and even that structured interviews represent little more than verbally administered intelligence tests (e.g., Hunter & Hirsh, 1987). Our findings suggest the opposite. Specifically we found that the extent of cognitive ability reflected in interview evaluations tends to decrease as the level of structure increases.

However, our finding of an inverse relationship between structure and ability must be qualified by taking content into account. We noted above that the content of a structured interview could affect the extent of ability in the ratings, making it either high or low. That appeared to be the case, and in particular, we found fairly low-ability saturation for most of the situational interviews (about 10% at a construct level) and even less for the behavior description interviews (about 3% at a construct level). Yet, there were a handful of high-structure studies that were considerably more saturated with ability (generally over 60% at a construct level).

The direct implication of the above is that it may be possible to develop high-structure interviews that correlate either very high or low with cognitive ability. Exactly why these handful of studies correlated so much more highly with ability is not known at the present time, and it should definitely be pursued in future research. It could possibly relate to the method of job analysis (e.g., worker-oriented vs. job-oriented; see Campion, Palmer, & Campion, 1995) or the manner in which the questions were written. Latham and Saari (1984) noted that the process of writing questions from a job analysis amounts to "literary license." In turn, this might suggest that question development processes may be as influential as the intent of the questions (i.e., which dimension each question was designed to assess).

Knowing how to develop high-structure formats, such as situational and past behavior interviews, that can be either high or low on ability would then allow researchers to customize the interview to the situation. If an ability test is also to be used in selection, then the researcher

could attempt to minimize ability saturation in the interview in order to increase the probability that it will have incremental validity. If an organization chooses not to use an ability test for legal or other reasons, then the researcher might attempt to maximize ability saturation in order to increase the probability that the interview will be a valid predictor of job performance.

One point about interview structure needs to be clarified. Given the findings that ability saturation tends to increase validity and ability saturation is inversely related to structure, it might be concluded that low-structure interviews are better than high-structure interviews because they tend to be more saturated with ability. However, that is not the case, as the structured interviews in our dataset did had higher overall validity than the unstructured interviews. Why then do structured interviews have higher validity when they are typically less saturated with ability? Possibly, structured interviews are better at assessing other constructs that are also related to validity. For example, Bosshardt (1992) developed a behavior description interview that correlated .00 with an ability test but had an uncorrected validity coefficient of .36.

Isolating what other constructs structured interviews are commonly capturing could very well be the next major breakthrough in construct research. Empirically, factor analyzing the questions in a structured interview may isolate the constructs. This in turn could be identified by content and then individually correlated with performance criteria. For example, certain questions that assess a generalized motivation factor may emerge.

A number of directions for future research emerged from this investigation. Some have already been mentioned, particularly in regard to continued construct research. Another area is to identify what drives capturing of ability with low-structure interviews, whether it is direct assessment of applicants' ability to learn job requirements quickly, applicant impression management skills, the type of questions asked, or more indirect influences from background characteristics. It would be interesting to see how much effect individual differences among interviewers can have on the relative influence from these sources.

A related area for future research involves the issue of adverse impact. Certain minority groups consistently score lower on cognitive ability tests, generally about one standard deviation (Hunter & Hunter, 1984). Because structured interviews appear to provide comparable validity (Huffcutt & Arthur, 1994; McDaniel et al., 1994; Wiesner & Cronshaw, 1988), it would be interesting to see whether they also result in the same overall level of adverse impact and whether that level varies depending on the level of cognitive ability reflected in the interview ratings. Eventually, it may be possible to identify or design interview formats that maintain high validity while reducing the level of adverse impact.

Limitations of this investigation should be noted. First, the size of our dataset was relatively modest, even after an extensive search, and a larger dataset would have been desirable. However, it is unlikely that many more studies could have been found given the difficulty noted by Hunter and Hunter (1984) in finding studies that reported correlations among predictors. Second, different cognitive ability tests were used in the interview studies, and this may have added "noise" to the data. Such noise might have increased the residual variances and could help explain why the percentages of variances accounted for by artifacts was relatively low. Ideally, the same general cognitive ability test would be used in a study of this nature. Third, we found only one comprehensive structured interview study. This prohibited us from analyzing it as a separate content category. Fourth, meta-analysis as a technique is by nature correlational, and drawing causal inferences from findings should always be done with caution. Findings represent average trends across a diversity of interview conditions, and the outcome for any one individual situation can vary from these trends.

Nonetheless, these results here do make a significant contribution to the interview literature. At a conceptual level, they provide understanding of the relationship between interviews and cognitive ability and the characteristics that moderate this relationship. At a more pragmatic level, they suggest when the interview is likely to have both validity and incremental validity, thereby allowing design of optimal selection strategies. In short, results of this investigation have provided an important first step in establishing construct validity of the interview.

References

The asterisk (*) indicates studies that were included in the meta-analysis.

*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.

Barron, R. A. (1989). Impression management by applicants during employment interviews: The "too much of a good thing" effect. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 204–215). Newbury Park, CA: Sage.

*Bass, B. M. (1949). *Comparison of the leaderless group discussion and the individual interview in the selection of sales and management trainees.* Unpublished doctoral dissertation, Ohio State University.

*Berkley, S. (1984). *VII. Validation Report Corrections Officer Trainee.* Harrisburg: Commonwealth of Pennsylvania Civil Service Commission.

Bolles, R. N. (1995). *What color is your parachute?* Berkeley, CA: Ten Speed Press.

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.

*Campbell, J. T., Prien, E. P., & Brailey, L. G. (1960). Predicting performance evaluations. *Personnel Psychology, 13,* 435–440.

*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,* 998–1102.

Campion, M. A., Palmer, D. K., & Campion, J. E. (1995). *A review of structure in the employment interview.* Manuscript submitted for publication.

*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.

Conrad, H. S., & Satter, G. A. (1945). *The use of test scores and quality-classification ratings in predicting success in electrician's mates school* (Report No. 23). Princeton, NJ: Naval Office of Scientific Research and Development.

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.

Darany, T. (1971). *Summary of State Police Trooper 07 Validity Study.* Lansing, MI: Department of Civil Service.

*Davey, B. (1984, May). *Are all oral panels created equal?* Paper presented at the Annual Conference of the International Personnel Management Association Assessment Council.

*Delery, J. E., Wright, P. M., McArthur, K., & Anderson, D. C. (1994). Cognitive ability tests and the situational interview: A test of incremental validity. *International Journal of Selection and Assessment, 2,* 53–58.

*Denning, D. L. (1993, April). *An initial investigation of constructs assessed in employment interviews.* Paper presented at the 8th Annual Conference of the Society for Industrial and Organizational Psychology, San Francisco.

Dipboye, R. L. (1989). Threats to the incremental validity of interviewer judgments. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 45–60). Newbury Park, CA: Sage.

Dipboye, R. L., & Gaugler, B. B. (1993). Cognitive and behavioral processes in the selection interview. In N. Schmitt & W. Borman (Eds.), *Personnel selection in organizations* (pp. 135–170). San Francisco: Jossey-Bass.

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

Distefano, M. K., Jr., & Pryer, M. W. (1987). Evaluation of selected interview data in improving the predictive validity of a verbal ability test with psychiatric aide trainees. *Educational and Psychological Measurement, 47,* 189–192.

Freeman, G. L., Manson, G. E., Katzoff, E. T. & Pathman, J. H. (1942). The stress interview. *Journal of Abnormal and Social Psychology, 37,* 427–447.

Freund, R. J., & Littell, R. C. (1991). *SAS system for regression* (2nd ed.). Cary, NC: SAS Institute.

Friedland, D. (1973). *Selection of police officers.* Los Angeles: City of Los Angeles Personnel Department.

*Friedland, D. (1976). *Section II—City of Los Angeles Personnel Department: Junior Administrative Assistant Validation Study.* Los Angeles: City of Los Angeles Personnel Department.

*Friedland, D. (1980). *A predictive validity study of selected tests for police officer selection.* Los Angeles: City of Los Angeles Personnel Department.

Gandy, J. A. (1986). *Job complexity, aggregated subsamples, and aptitude test validity: Meta-analysis of the GATB database.* Washington, DC: U.S. Office of Personnel Management.

Gardner, H., & Hatch, T. (1989). Multiple intelligences go to school: Educational implications of the theory of multiple intelligences. *Educational Researcher, 18*(8), 4–10.

Gilmore, D. C., & Ferris, G. R. (1989). The politics of the employment interview. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 195–203). Newbury Park, CA: Sage.

Hakel, M. D. (1989). The state of employment interview theory and research. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 285–293). Newbury Park, CA: Sage.

Handyside, J. D., & Duncan, D. C. (1954). Four years later: A follow-up of an experiment in selecting supervisors. *Occupational Psychology, 28,* 9–23.

Harris, M. M. (1989). Reconsidering the employment interview: A review of recent literature and suggestions for future research. *Personnel Psychology, 42,* 691–726.

Hedge, J. W., & Teachout, M. S. (1992). An interview approach to work sample criterion measurement. *Journal of Applied Psychology, 77,* 453–461.

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

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., Arthur, W., Jr., & Bennett, W. (1993). Conducting meta-analysis using the PROC MEANS procedure in SAS. *Educational and Psychological Measurement, 53,* 119–131.

Hunter, J. E., & Hirsh, H. R. (1987). Applications of meta-analysis. In C. L. Cooper & I. T. Robertson (Eds.), *International review of industrial and organizational psychology* (pp. 321–357). Chichester, England: Wiley.

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.

Hunter, J. E., Schmidt, F. L., & Judiesch, M. K. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology, 75,* 28–42.

*Huse, E. F. (1962). Assessments of higher-level personnel: IV. The validity of the assessment techniques based on systemat-

ically 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.

Janz, T. (1989). The patterned behavior description interview: The best prophet of the future is the past. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 158–168). Newbury Park, CA: Sage.

*Johnson, E. K. (1990). *The structured interview: Manipulating structuring criteria and the effects on validity, reliability, and practicality.* Unpublished doctoral dissertation, Tulane University.

*Johnson, E. M., & McDaniel, M. A. (1981, May). *Training correlates of a police selection system.* Paper presented at the International Personnel Management Association Conference, Denver.

Jones, E. E., & Pittman, T. S. (1982). Toward a general theory of strategic self-presentation. In J. Suls (Ed.), *Psychological perspectives on the self.* Hillsdale, NJ: Erlbaum.

Kaplan, R. M., & Saccuzzo, D. P. (1993). *Psychological testing: Principles, applications, and issues* (3rd ed.). Pacific Grove, CA: Brooks/Cole.

Keppel, G. (1991). *Design and analysis: A researcher's handbook* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Kinicki, A. J., & Lockwood, C. A. (1985). The interview process: An examination of factors recruiters use in evaluating job applicants. *Journal of Vocational Behavior, 26,* 117–125.

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. (1984). Do people do what they say? Further studies on the situational interview. *Journal of Applied Psychology, 69,* 569–573.

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

Lopez, F. M., Jr. (1966). Current problems in test performance of job applicants: I. *Personnel Psychology, 19,* 10–18.

Mainstream science on intelligence: A statement signed by 52 experts in intelligence and allied fields. (1995). *Industrial-Organizational Psychologist, 32*(4), 67–72.

Marchese, M. C., & Muchinsky, P. M. (1993). The validity of the employment interview: A meta-analysis. *International Journal of Selection and Assessment, 1,* 18–26.

McDaniel, M. A. (1986). The evaluation of a causal model of job performance: The interrelationships of general mental ability, job experience, and job performance (Doctoral dissertation, George Washington University, 1990). *Dissertation Abstracts International, 46,* AAD86-08356.

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.

*Meredith, K. E., Dunlap, M. R., & Baker, H. H. (1982). Subjective and objective admissions factors as predictors of clinical clerkship performance. *Journal of Medical Education, 57,* 743–751.

*Motowidlo, S. J., Carter, G. W., Dunnette, M. D., Tippins, N.,

Werner, S., Burnett, J. R., & Vaughan, M. J. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology, 77,* 571–587.

Muchinsky, P. M. (1993). *Psychology applied to work* (4th ed.). Pacific Grove, CA: Brooks/Cole.

Myers, R. H. (1990). *Classical and modern regression with applications* (2nd ed.). Boston: PWS and Kent.

Phillips, A. P., & Dipboye, R. L. (1989). Correlational tests of predictions from a process model of the interview. *Journal of Applied Psychology, 74,* 41–52.

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

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than g. *Journal of Applied Psychology, 79,* 518–524.

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

*Requate, F. (1972). *Metropolitan Dade County Fire Fighter Validation Study.* Miami: City of Miami Personnel Department.

*Rhea, B. D., Rimland, B., & Githens, W. H. (1965). *The development and evaluation of a forced-choice letter of reference form for selecting officer candidates* (Technical Bulletin STB 66-10). San Diego, CA: U.S. Naval Personnel Research.

*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.

Schmidt, F. L., Hunter, J. E., & Outerbridge, A. N. (1986). Impact of job experience and ability on job knowledge, work sample performance, and supervisory ratings of job performance. *Journal of Applied Psychology, 71,* 432–439.

*Schuler, H., & Funke, U. (1989). The interview as a multimodal procedure. In R. W. Eder & G. R. Ferris (Eds.), *The employment interview: Theory, research, and practice* (pp. 183–192). Newbury Park, CA: Sage.

Solso, R. L. (1991). *Cognitive psychology* (3rd ed.). Boston: Allyn & Bacon.

*Sparks, C. P. (1973). *Validity of a selection program for process trainees* (Personnel Research No. 73-1). Unpublished manuscript.

*Sparks, C. P. (1973). *Validity of a selection program for refinery technician trainees* (Personnel Research No. 73-4). Unpublished manuscript no longer available for distribution.

*Sparks, C. P. (1974). *Validity of a selection program for commercial drivers* (Personnel Research No. 74-9). Unpublished manuscript no longer available for distribution.

*Sparks, C. P. (1978). *Selecting laboratory technicians* (Personnel Research No. 78-5). Unpublished manuscript no longer available for distribution.

Srull, T. K., & Wyer, R. S. (1989). Person memory and judgment. *Psychological Review, 96,* 58–83.

Tedeschi, J. T., & Melburg, V. (1984). Impression management and influence in the organization. In S. Bacharach & E. J. Lawler (Eds.), *Research in the sociology of organizations,* (Vol. 3, pp. 31–58). Greenwich, CT: JAI Press.

Trankell, A. (1959). The psychologist as an instrument of prediction. *Journal of Applied Psychology, 43,* 170–175.

*Tubiana, J. H., & Ben-Shakhar, G. (1982). An objective group questionnaire as a substitute for a personal interview in the prediction of success in military training in Israel. *Personnel Psychology, 35*, 349–357.

*Tziner, A., & Dolan, S. (1982). Validity of an assessment center for identifying future female officers in the military. *Journal of Applied Psychology, 67*, 728–736.

U.S. Department of Labor. (1977). *Dictionary of occupational titles* (4th ed.). Washington, DC: U.S. Government Printing Office.

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

Valenzi, E., & Andrews, I. R. (1973). Individual differences in the decision processes of employment interviews. *Journal of Applied Psychology, 58*, 49–53.

Vernon, P. E. (1950). The validation of civil service selection board procedures. *Occupational Psychology, 24*, 75–95.

*Vernon, P. E., & Parry, J. B. (1949). *Personnel selection in the British forces.* London: University of London Press.

*Villanova, P., Bernardin, H. J., Johnson, D. L., & Dahmus, S. A. (1994). The validity of a measure of job compatibility in the prediction of job performance and turnover of motion picture theater personnel. *Personnel Psychology, 47*, 73–90.

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

Wechsler, D. (1981). *Wechsler Adult Intelligence Scale—Revised.* San Antonio, TX: Psychological Corp.

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.

Wonderlic, E. F. (1983). *Wonderlic Personnel Test manual.* Northfield, IL: Wonderlic & Associates.

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.

*Zaccaria, M. A., Dailey, J. T., Tupes, E. C., Stafford, A. R., Lawrence, H. G., & Ailsworth, K. A. (1956). *Development of an interview procedure for USAF officer applicants* (AFPTRC-TN-56-43, Project No. 7701). Lackland Air Force Base, TX: Air Research and Development Command.