PERSONNEL PSYCHOLOGY
2007, 60, 837–874

REVISITING INTERVIEW–COGNITIVE ABILITY RELATIONSHIPS: ATTENDING TO SPECIFIC RANGE RESTRICTION MECHANISMS IN META-ANALYSIS

CHRISTOPHER M. BERRY, PAUL R. SACKETT,
AND RICHARD N. LANDERS
University of Minnesota

This study revisits the relationship between interviews and cognitive ability tests, finding lower magnitudes of correlation than have previous meta-analyses; a finding that has implications for both the construct and incremental validity of the interview. Our lower estimates of this relationship than previous meta-analyses were mainly due to (a) an updated set of studies, (b) exclusion of samples in which interviewers potentially had access to applicants' cognitive test scores, and (c) attention to specific range restriction mechanisms that allowed us to identify a sizable subset of studies for which range restriction could be accurately accounted. Moderator analysis results were similar to previous meta-analyses, but magnitudes of correlation were generally lower than in previous meta-analyses. Findings have implications for the construct and incremental validity of interviews, and meta-analytic methodology in general.

The correlation between applicants' interview and cognitive ability scores has important implications for the construct and incremental validity of the selection interview. First, this correlation helps make clearer what interviews measure. Second, the smaller the correlation, the greater the potential for interview scores to explain variance in job performance over cognitive ability. Three recent meta-analyses (Cortina, Goldstein, Payne, Davison, & Gilliland, 2000; Huffcutt, Roth, & McDaniel, 1996; Salgado & Moscoso, 2002) have estimated the correlation between interview and cognitive test scores. We revisit the interview–cognitive test relationship for three main reasons. First, enough time has passed since the last interview–cognitive test score meta-analysis to allow us to collect an updated set of studies. Second, we wished to explore in greater depth than previous meta-analyses the role that interviewer access to applicants' cognitive test scores plays in the interview–cognitive test relationship. Third,

An earlier version of this manuscript was presented at the 2006 Society for Industrial and Organizational Psychology Conference.

Christopher Berry is now at Wayne State University.
Correspondence and requests for reprints should be addressed to Christopher M. Berry, Wayne State University, Department of Psychology, 5057 Woodward Ave., 7th Floor, Detroit, MI, 48202; berry@wayne.edu.

we wished to demonstrate the importance of attending to specific range restriction (RR) mechanisms in meta-analysis instead of automatically applying direct RR corrections (Hunter, Schmidt, & Le, 2006; Sackett, Lievens, Berry, & Landers, 2007; Schmidt, Oh, & Le, 2006), as was done in previous interview–cognitive test meta-analyses. We outline the previous interview–cognitive test meta-analyses below.

Huffcutt et al. (1996), in their widely cited meta-analysis, concluded that the mean correlation between interview and cognitive test scores was .40 (corrected for direct RR in interviews using a mean RR ratio of .74 drawn from 15 samples, which was reported in Huffcutt and Arthur (1994) and corrected for unreliability in both interviews and cognitive tests). They also concluded that the interview–ability relationship was moderated by interview structure, type of interview questions (situational vs. behavior description), job complexity, and the uncorrected criterion-related validity of the interview. Huffcutt et al. reported correlations between interviews and cognitive tests higher than .5 for many of the moderator levels. Huffcutt et al.'s estimates were later used by Schmidt and Hunter (1998) to estimate the incremental validity of the interview.

Salgado and Moscoso (2002) also performed a meta-analysis of the relationship between interview and cognitive test scores, incorporating a greater number of studies than Huffcutt et al. (1996). Salgado and Moscoso reported two correlations (corrected for direct RR in interviews using an empirical artifact distribution drawn from 38 samples with a mean RR ratio of .61, and corrected for unreliability in both interview and cognitive test scores): .28 between cognitive tests and behavioral interviews (situational interviews [SI] or behavior description interviews [BDI]), and .41 between cognitive tests and conventional interviews (interviews that were not SI or BDI). These general magnitudes were very similar to those reported in Huffcutt et al. (1996).

Finally, Cortina et al. (2000) also performed a meta-analysis of the correlation between interview and cognitive test scores. This was part of a larger study of intercorrelations among various predictors, and it incorporated a smaller number of studies ($k = 21$) than Huffcutt et al. (1996; $k = 49$) or Salgado and Moscoso (2002; $k = 75$). Cortina et al.'s smaller number of studies likely made their meta-analytic estimates less stable than those of Huffcutt et al. and Salgado and Moscoso. Thus, we focus on Huffcutt et al. (1996) and Salgado and Moscoso (2002) as the primary sources of previous information about interview–cognitive test relationships.

Three Main Reasons to Revisit the Interview–Cognitive Test Relationship

This study revisits the correlation between interviews and cognitive test scores. Three things in particular prompted us to revisit the interview–cognitive test relationship. First, since the last interview–cognitive test meta-analysis, empirical work on the relationship between interviews and cognitive tests has continued to be completed. Thus, it was of value to revisit the relationship between interview and cognitive test scores incorporating an updated set of studies.

Second, we were especially concerned about the potentially confounding effects of interviewer access to applicants' cognitive test scores before or during the interview. That is, our focus was determining how much ratings of applicants are affected by impressions of the applicants' cognitive ability formed through interaction with the applicants, not by looking at applicants' cognitive test scores. Huffcutt et al. (1996) demonstrated that the interview–cognitive test relationship was inflated when interviewers were allowed access to cognitive tests. Because of this, Huffcutt et al. (1996) excluded from some analyses, and Salgado and Moscoso (2002) excluded from all analyses those samples in which it was *obvious* that interviewers had access to cognitive test scores.

We felt it prudent to take this notion further and investigate samples in which it was *plausible* that interviewers had access to applicants' cognitive test scores. We posit there are a number of factors that make it more plausible that interviewers had access to applicants' cognitive test scores, even if the primary article did not explicitly state such was the case. For instance, if applicants were first administered cognitive tests and then given interviews, it seems reasonable to suspect that it is more plausible that interviewers had access to test scores. Similarly, if organization members (e.g., supervisors, hiring managers, etc.) administered the interviews and the researchers had little or no control over the interview process, perhaps even only documenting in a post hoc fashion how the interview was carried out, the plausibility of interviewer access probably increases. Finally, if the primary research article makes no explicit statement that interviewers did not have access to applicants' test scores, it seems reasonable to suspect that it is more plausible that interviewers had access to cognitive tests. Though each of these elements in isolation might not be a red flag, when all of these elements converge in one sample, the likelihood that interviewers had access to applicants' test scores seems dramatically increased. As such, we felt it prudent to investigate such samples in which all of these elements converged.

Our third main reason for revisiting the relationship between interviews and cognitive tests was to attend to specific RR mechanisms instead of assuming direct RR. Hunter et al. (2006) and Schmidt et al. (2006) demonstrated the importance of distinguishing between direct and indirect RR, showing that applying direct RR corrections when RR is indirect can misestimate true correlations. Sackett et al. (2007) demonstrated that a failure to distinguish between different RR mechanisms has especially insidious effects on estimates when examining relationships between predictors (such as interviews and cognitive tests). Such information was not widely known when previous interview–cognitive test meta-analyses were completed. Therefore, using techniques that were standard at the time, Huffcutt et al. (1996) and Salgado and Moscoso (2002) attempted to account for RR by applying direct RR corrections to their mean sample-size-weighted correlations. This is implicitly the same as correcting every primary study in the meta-analyses for the same amount of direct RR. Such a practice is commonplace because of the assumption that virtually all validity studies suffer from at least some RR (e.g., Gulliksen, 1950; Hunter & Schmidt, 2004; Thorndike, 1949). Assuming all studies suffer from RR is likely warranted when examining predictor–criterion relationships because the only way a predictor–criterion correlation could be based on an unrestricted sample is in the rare case that all applicants were hired regardless of test scores. The assumption that virtually all correlations drawn from primary studies suffer from RR is not warranted in the special case of examining relationships between predictors. In many instances, correlations between predictors drawn from primary studies will have no RR because both predictors were administered to all applicants, and the primary study reports this correlation based on the entire applicant population.

Imagine, for instance, that 500 applicants apply for a position and are administered an interview and cognitive test, with both predictors used to select among these applicants. In the published article outlining the selection process for these applicants, though, the correlation reported between interview and cognitive test scores was based on the entire 500-applicant pool. In this case there is no RR affecting the correlation between interview and cognitive test scores, and a correction for direct RR would be inappropriate.

We wish to be clear about what we mean by "no RR." First, in the scenario above, it is possible that the 500 applicants were actually screened, based on things such as application blanks, from a larger group of submitted applications. So, the 500 applicants do not represent the pool of every person that submitted an application (typically there are no restrictions on who can at least submit an application), but instead represent the entire pool that the organization deemed worthy of consideration for the job. This latter is what we consider an "unrestricted applicant pool" (the Equal Employment Opportunity Commission and Department of Justice also support a similar view in their recent guidelines on employee selection procedures as they relate to the Internet [Adoption of Additional Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures as They Relate to the Internet and Related Technologies, 2004]), although we recognize that this pool will have less range than the purely hypothetical situation wherein every person who submitted an application is given an interview and cognitive test.

Second, it is true that an applicant pool for, say, a managerial position will probably not have as wide a range of ability scores as will the general population. Whether one wants to consider this a type of RR to be corrected, though, depends on the research question. Correcting for differences in variability between the applicant pool in question and the general population changes the focus away from any applied purpose to a hypothetical scenario in which the entire general population is the employer's applicant pool. Therefore, when we say "no RR," we mean it in the sense that a correlation is based on an applicant pool and not the general population.

This study offers evidence that scenarios in which direct RR corrections are inappropriate (the scenario listed above is only one example) actually represent the bulk of the studies in our interview–cognitive test meta-analytic database. That is, this study found that direct RR was not present in the majority of the studies included in these meta-analyses and that many studies had no RR whatsoever. Therefore, a direct RR correction to studies without direct RR on the interview or cognitive test scores of interest clouds estimates of the true correlation.

In this study, instead of implicitly correcting all studies for direct RR, we used a strategy of grouping studies into subgroups based on the presence of differing RR processes and applying RR corrections appropriate to the RR mechanism in question. This was the strategy proposed by Sackett et al. (2007). This study will first demonstrate that each unique RR process differentially affects magnitudes of relationship between predictors. Further, corrections for some of these RR processes are possible whereas corrections for others are not due to a lack of information in primary studies. By paying close attention to the specific RR processes affecting each primary study, this study identified a sizable subset of samples that either did not have restricted range or that had restricted range, but the specific mechanism causing this RR was known and can be corrected. Another sizable subset of samples was identified in which it was likely that RR existed, but a correction would not be appropriate because of a lack of adequate information in the primary studies. Options for each category of samples are discussed and exercised. Although the process laid out in this study is important for the relationship between interview and cognitive test scores, this process also holds implications for any meta-analysis, especially those examining interrelationships between predictors.

Method

Search for Primary Data

First, attempts were made to locate the articles included in the Cortina et al. (2000), Huffcutt et al. (1996), and Salgado and Moscoso (2002) meta-analyses. The senior authors of these meta-analyses were contacted to request any articles we could not locate. Second, keyword searches of the PsycInfo, ERIC, and MEDLINE databases were conducted. Third, manual searches of *International Journal of Selection and Assessment*, *Journal of Applied Psychology*, *Journal of Occupational and Organizational Psychology*, and *Personnel Psychology* were performed from Winter 1999 (year that Salgado and Moscoso's manual searches of the same journals ended) onward. Finally, the conference programs for the Society for Industrial and Organizational Psychology and the Academy of Management conferences (1998 onward) were manually searched.

Studies were included in the meta-analysis if a correlation between a selection interview (selection interviews include both employment interviews and college admissions interviews; we note that previous interview–cognitive test meta-analyses also included both types of interviews) and a cognitive test could be extracted either from information included in the article or via personal communication with the primary study authors. This resulted in 78 independent samples with a total sample size of 20,014 drawn from 63 articles. Forty samples were drawn from published sources whereas 38 samples were drawn from unpublished sources. The 78 samples included all of the articles used in Huffcutt et al. (1996) and 20 of the 21 articles used in Cortina et al. (2000) (we were not able to locate Friedland [1973], which was included in Cortina et al.). Salgado and Moscoso's (2002) references section did not detail exactly which articles were used in their interview–cognitive test meta-analysis, so we do not know exactly how much overlap there was between our meta-analytic databases, although the overlap should be significant because we used their references section to search for articles. Our meta-analysis also incorporated 15 studies not included in any of the previous meta-analyses.

When drawing correlations between interview and cognitive test scores from primary studies, preference was given to correlations involving overall interview scores and overall cognitive test scores (e.g., overall score on a battery of cognitive tests instead of scores on individual cognitive tests). Therefore, all else equal, the following decision rules were followed. When a correlation with an overall interview score was available, we used that correlation. When correlations were only available for separate interview dimensions, we used an average of those correlations. When a correlation with an overall cognitive test score battery was available, we used that correlation. When correlations were only available for multiple individual cognitive tests, and intercorrelations among the cognitive tests were provided, we estimated what the correlation between the interview score and a composite of those individual cognitive tests would be using formulas provided by Ghiselli, Campbell, and Zedeck (1981, pp. 163–164). When correlations were only available for multiple individual cognitive tests, but the intercorrelations among the cognitive tests were not provided, we used the mean average of the interview–cognitive test correlations.

Coding of Study Characteristics

For each sample, seven study characteristics were coded for use in moderator analyses: type of RR mechanism affecting the interview–cognitive test correlation, four moderators used by Huffcutt et al. (1996), the percentage of interview dimensions in each interview designed to capture cognitively oriented attributes (% cognitive), and the likelihood that interviewers had access to applicants' cognitive test scores. The coding for each of these characteristics is described below.

Type of RR. Each type of RR was expected to differentially affect interview–cognitive test correlations. Specifically, samples were coded into the following five categories:

(1) Samples in which correlations between interview and cognitive test scores were based on the entire applicant pool: Because the entire applicant pool was used in these samples, there was no RR in these cases, and thus any RR correction is inappropriate and overestimates the true correlation.
(2) Samples in which the correlation was drawn from a job incumbent sample: In these samples, because participants were incumbents, there was no direct RR resulting from the interviews or cognitive tests used in the primary studies. It is difficult to imagine, though, that incumbents were not selected for their current positions using some form of an interview. Indeed, Ryan, McFarland, Baron, and Page (1999), in their survey of the prevalence of selection methods used in 959 organizations across 20 countries, found that some form of an interview was almost always used by organizations (especially in the U.S., where the bulk of our samples came from). Because scores assigned to interviewees in separate interviews have been shown to correlate (Conway, Jako, & Goodman, 1995), these incumbent samples likely have interview scores that are indirectly restricted due to selection on an interview other than the one reported in the primary study. Thus, a direct RR correction is inappropriate in such cases. Only a correction for indirect RR caused by selection

on a second interview would be appropriate. The possibility of correcting incumbent samples for indirect RR resulting from selection on a second cognitive test was also considered. Using an empirical estimate of (a) the prevalence of the use of cognitive ability tests to select applicants for hire (20%; Gowing & Slivinski, 1994; Marsden, 1994; Ryan et al., 1999) and (b) the average intercorrelation between cognitive ability tests (.70; Drasgow, 2003), we modeled corrections for indirect RR resulting from selection on a second cognitive test. In no instance did such a correction change any meta-analytic estimate more than .01, so corrections were not made for selection on a second cognitive test.

(3) Samples in which the interview–cognitive test correlation was reduced due to direct RR on only one predictor (e.g., applicants are admitted to an interview based on their scores on a cognitive test): A direct RR correction is appropriate for samples falling in this category.

(4) Samples in which the interview–cognitive test correlation was reduced due to RR on both predictors. When this restriction was due to selection on a composite of cognitive test and interview scores, only an indirect RR correction would be appropriate (as the restriction is not directly due to selection on interview or cognitive test scores, but instead on a composite of the two). A direct RR correction in this case would underestimate the true correlation (Sackett et al., 2007). When the restriction on both predictors was due to multiple hurdles selection, the appropriate correction would be to correct for direct RR on whatever variable was used as the second hurdle and then correct for direct RR again on whatever variable was used as the first hurdle (Sackett & Yang, 2000). Not enough information was provided in any studies to make appropriate corrections to correlations in this fourth category.

(5) Samples in which not enough information was included in the primary study to determine whether the sample was restricted on the interview or cognitive test used in the study: In this case, we do not attempt a correction for RR and do not use these studies in estimating the corrected correlation between interviews and cognitive ability.

Huffcutt et al. (1996) Moderators. Samples were also coded according to four variables that Huffcutt et al. (1996) found moderated the relationship between interview and cognitive test scores. First, interviews were coded according to the level of structure in the interview using a framework presented by Conway et al. (1995). Interviews were coded according to five progressively higher levels of question standardization and three progressively higher levels of standardization of response evaluation. Like Huffcutt et al. (1996), we combined various combinations of these two aspects of structure into three overall levels corresponding to low, medium, and high structure. Second, interviews were coded according to the content of the questions in the interview. Interviews were coded as either BDI (mostly involving questions about past behavior), SI (mostly involving questions about how one would behave in a future situation), other (mostly involving questions that are not situational or behavior description), or composite (interviews including more than one type of the abovementioned questions). Third, the level of job complexity of the job for which applicants were being interviewed was coded. The three-level framework developed by Hunter, Schmidt, and Judiesch (1990) was used to categorize these jobs as low, medium, or high complexity. Fourth, interviews were coded according to whether their uncorrected criterion-related validity for predicting job performance was low ($r \leq .199$), medium (r between .20 and .299), or high ($r \geq .30$).

% cognitive. The degree to which each interview was designed to assess cognitively oriented attributes may moderate the interview–cognitive test score relationship. That is, an interview in which interviewers are asked to rate applicants along dimensions such as intellectual capacity or ability to learn may be more likely to have high cognitive load than an interview in which interviewers are asked to rate applicants along dimensions such as interpersonal skills or dependability. Thus, to capture the degree to which interviews were designed to measure cognitively oriented attributes, we assigned each interview a percentage reflecting the percentage of that interview's dimensions that were cognitive in nature. So, an interview in which interviewers were asked to rate applicants along four dimensions, two of which were cognitive in nature, would be coded as 50%. To determine whether interview dimensions were cognitive in nature, we consulted Huffcutt, Conway, Roth, and Stone (2001) and Arthur, Day, McNelly, and Edens (2003), who listed common labels for interview and assessment center dimensions, respectively. Specifically, if one of the present meta-analysis' interview's dimension labels matched a dimension label listed in either the "mental capability" section of Huffcutt et al.'s (2001) table 1 or the "problem solving" section of Arthur et al.'s (2003) table 2, then that dimension was coded as designed to assess a cognitively oriented attribute for the purposes of the present meta-analysis. Some dimension labels in the present meta-analysis' interview studies did not exactly match dimensions listed by Huffcutt, Conway et al. (2001) or Arthur et al. (2003) but were still cognitive in nature (e.g., strategic skills; range of vocabulary; ability to plan, organize, prioritize; etc.) and were thus also coded as assessing a cognitively oriented attribute.

Interviewer access to cognitive test scores. Interviews were coded according to whether the interviewers were allowed access to applicants' cognitive test scores before or during the interview. We coded samples into three categories representing a continuum of the likelihood that interviewers had access to applicants' test scores. The first category contained samples in which interviewers definitely had access to applicants' cognitive test scores (e.g., the article explicitly stated this was the case). The second category contained samples in which it could not be definitively proven that interviewers had access to test scores, but circumstantial evidence made it very plausible (i.e., in each of these samples applicants were first administered cognitive tests and then given interviews, organization members [e.g., supervisors, hiring managers, etc.] administered interviews and researchers had little to no control over the interview process, researchers only documented [usually post hoc] how the interview had been carried out, and there was no explicit statement that interviewers were not allowed access to applicants' cognitive test scores). The third category contained samples in which it was highly unlikely that interviewers had access to applicants' test scores (e.g., article explicitly stated this was the case, cognitive tests were administered after the interviews, or the researchers themselves conducted the interviews and were thus likely to be aware of the potential to contaminate interview judgments).

The first and third authors independently coded all studies (see Table 1, which outlines the study characteristics of all samples in the present meta-analysis). Initial agreement between raters was 93% for type of RR affecting correlations, 94% for interview structure, 93% for interview content, 85% for job complexity, 92% for uncorrected validity, 92% for % cognitive, and 93% for availability of cognitive test scores. Any disagreements were resolved via discussion as necessary.

Procedure and Analyses

We argue for a thoughtful analysis of each primary study before including it in a meta-analysis. Rather than including in all analyses any study containing a correlation between the variables of interest, our model carefully examines each primary study and only includes in analyses those primary studies with correlations that are not likely to be confounded by uncorrectable methodological or statistical artifacts. Figure 1 outlines the six-step process we used to determine whether primary studies should be included in our final substantive analyses. Our first two steps examined whether availability of applicants' cognitive test scores to interviewers moderated the relationship between interview and cognitive test scores. Thus, in Step 1 we sorted our full 78-coefficient sample into our three

TABLE 1

Coding for All Studies in the Meta-Analysis

Study	*r*	*N*	Type of RR	Test available?	Structure	Content	Complexity	Validity	% Cognitive[a]
Albrecht Glaser, and Marks (1964)	.41	31	Incumbents	Unlikely	Medium	N/A	High	High	33.0%
Bass (1949)	.28	32	App pool	Unlikely	N/A	N/A	Medium	N/A	N/A
Bass (1949)	.27	30	App pool	Unlikely	N/A	N/A	High	N/A	N/A
Berkley (1984)	.12	335	Double RR	Unlikely	High	N/A	Low	Medium	N/A
Bosshardt (1994)	.10	44	Direct RR	Unlikely	High	Composite	Medium	High	16.7%
Bradley, Bernthal, and Thomas (1998)	.13	76	Double RR	Unlikely	Medium	Composite	Medium	Low	N/A
Calkins, Richards, McCanse, Burgess, and Willoughby (1974)	.17	162	Double RR	Plausibly	High	Other	High	N/A	N/A
Calkins et al. (1974)	.19	241	Direct RR	Unlikely	High	Other	High	N/A	15.0%
Calkins et al. (1974)	.20	335	Direct RR	Unlikely	High	Other	High	N/A	15.0%
Campbell, Prien, & Brailey (1960)	.35	84	Direct RR	Yes	Low	N/A	Low	Low	N/A
Campion, Campion, and Hudson (1994)	.60	70	Incumbents	Unlikely	High	Composite	N/A	High	0.0%
Campion, Pursell, and Brown (1988)	.49	140	Direct RR	Unlikely	High	Composite	Low	High	N/A
Conway and Peneno (1999)	−.05	145	Direct RR	Unlikely	High	Composite	High	N/A	0.0%
Davey (1984)	.30	707	App Pool	Unlikely	High	SI	Medium	Medium	25.0%
Delery, Wright, McArthur, and Anderson (1994)	.27	47	Incumbents	Unlikely	High	SI	Medium	High	50.0%
Denning (1993)	.28	322	Double RR	Plausibly	Medium	N/A	Medium	Medium	N/A
Denning (1993)	.29	366	Double RR	Plausibly	Medium	N/A	Medium	Low	N/A
Denning (1993)	.34	329	Double RR	Plausibly	Medium	N/A	Medium	N/A	N/A
Dicken (1969)	.26	53	Double RR	Yes	N/A	N/A	High	Medium	N/A
Dipboye Gaugler, Hayes, and Parker (1992)	.07	420	Double RR	Yes	Low	N/A	Low	Low	N/A
Durivage, St. Martin, and Barrette (1995)	−.06	48	Unclear	Unlikely	N/A	N/A	High	N/A	N/A
Elam and Andrykowski (1991)	.40	356	Direct RR	Yes	Low	Other	High	N/A	N/A
Elliot (1981)	.24	400	Double RR	Plausibly	Low	Other	N/A	N/A	N/A
Fox and Spector (2000)	.18	116	Incumbents	Unlikely	Medium	BDI	High	N/A	0.0%
Friedland (1976)	.19	92	Double RR	Unlikely	Medium	N/A	High	Medium	N/A
Friedland (1980)	.12	243	Double RR	Unlikely	N/A	N/A	Medium	Low	N/A

(Continued)

TABLE 1
(Continued)

Study	r	N	Type of RR	Test available?	Structure	Content	Complexity	Validity	% Cognitive[a]
Handyside and Duncan (1954)	.07	110	Incumbents	Unlikely	Low	Other	High	Low	N/A
Hilliard (2000)	.10	149	App pool	Unlikely	High	BDI	Low	Low	N/A
Huffcutt, Weekly, Wiesner, Degroot, and Jones (2001)	−.07	93	Incumbents	Unlikely	High	Composite	High	Low	30.0%
Huse (1962)	.19	107	Incumbents	Unlikely	N/A	N/A	Medium	Low	25.0%
Johnson (1990)	.13	31	Direct RR	Unlikely	Medium	Other	High	High	16.7%
Johnson (1990)	.24	31	Direct RR	Unlikely	Low	Other	High	Low	16.7%
Johnson (1990)	.08	30	Direct RR	Unlikely	High	SI	High	High	16.7%
Johnson and McDaniel (1981)	.08	158	Double RR	Unlikely	High	SI	Medium	High	N/A
Klehe (2003)	.13	79	Incumbents	Unlikely	High	Composite	High	High	0.0%
Klehe and Latham (2005)	.11	79	Direct RR	Unlikely	High	Composite	High	High	0.0%
Largo (1994)	.04	59	Double RR	Unlikely	High	Other	Low	Medium	N/A
Lievens Harris, Van Keer, and Bisqueret (2003)	.11	166	Unclear	Unlikely	Low	BDI	High	Low	N/A
Lopez (1966)	.01	182	Double RR	Unlikely	High	Other	Medium	Low	N/A
Loven (1998)	.00	31	Double RR	Unlikely	High	SI	High	High	N/A
Melchers, Klehe, Richter, Kleinmann, & Konig (2005)	.19	94	App pool	Unlikely	High	Composite	High	N/A	33.0%
Menkes (2002)	.04	66	Incumbents	Unlikely	High	Composite	High	N/A	25.0%
Meredith, Dunlap, and Baker (1982)	.08	85	Double RR	Unlikely	Low	N/A	High	Medium	N/A
Motowidlo et al. (1992)	.07	107	App pool	Unlikely	High	BDI	Medium	N/A	10.0%
Motowidlo et al. (1992)	−.09	110	Incumbents	Unlikely	High	BDI	High	Medium	10.0%
Motowidlo et al. (1992)	.13	36	Incumbents	Unlikely	High	BDI	Medium	Medium	20.0%
Motowidlo et al. (1992)	.17	875	App pool	Unlikely	High	BDI	Medium	Low	10.0%
Newman, Bobbitt, and Cameron (1946)	.65	498	Unclear	Yes	Low	Other	High	N/A	N/A
Phillips and Dipboye (1989)	.04	129	Direct RR	Yes	Low	Other	Medium	N/A	N/A
Pulakos and Schmitt (1995)	.09	464	Incumbents	Unlikely	High	BDI	Medium	High	28.6%
Pulakos and Schmitt (1996)	.12	461	Incumbents	Unlikely	High	BDI	Medium	High	25.0%
Reeb (1969)	.44	1250	App pool	Unlikely	Medium	Other	Low	High	0.0%
Requate (1972)	.67	26	Unclear	Unlikely	Medium	N/A	Medium	High	N/A

(Continued)

TABLE 1
(Continued)

Study	r	N	Type of RR	Test available?	Structure	Content	Complexity	Validity	% Cognitive[a]
Rhea, Rimland, and Githens (1965)	.11	1882	Unclear	Unlikely	Medium	N/A	High	Low	N/A
Roth and Campion (1992)	.21	172	Double RR	Unlikely	Medium	Other	Medium	High	N/A
Schuler and Funke (1989)	.21	307	Double RR	Unlikely	High	SI	Medium	N/A	N/A
Shahani, Dipboye, and Gehrlein (1991)	.16	2583	App pool	Unlikely	High	Other	High	Low	16.7%
Sparks (1973a)	.09	54	Double RR	Unlikely	N/A	N/A	Medium	High	N/A
Sparks (1973b)	−.01	170	Double RR	Unlikely	N/A	N/A	Medium	Low	N/A
Sparks (1974)	.37	67	Double RR	Unlikely	N/A	N/A	Low	High	N/A
Sparks (1978)	.23	48	Double RR	Unlikely	N/A	N/A	Medium	Medium	N/A
Sue-Chan and Latham (2004)	.20	65	Incumbents	Unlikely	High	SI	High	Medium	0.0%
Tubiana and Ben-Shakhar (1982)	.33	459	App pool	Unlikely	Medium	N/A	Low	High	0.0%
Tziner and Dolan (1982)	.34	193	App pool	Unlikely	N/A	N/A	Medium	Low	N/A
U.S. Office of Personnel Management (1987)	.21	394	Incumbents	Unlikely	High	SI	High	Medium	25.0%
U.S. Office of Personnel Management (1987)	.28	86	Unclear	Unlikely	High	SI	Medium	Low	N/A
U.S. Office of Personnel Management (1987)	.06	55	Unclear	Unlikely	High	SI	Medium	Medium	N/A
U.S. Office of Personnel Management (1987)	.19	67	Unclear	Unlikely	High	SI	Medium	Medium	N/A
U.S. Office of Personnel Management (1987)	.23	102	Unclear	Unlikely	High	SI	High	Low	N/A
Van Iddekinge and Eidson (2005)	.18	412	App pool	Unlikely	High	BDI	Medium	N/A	0.0%
Van Iddekinge and Henry (2006)	.25	203	Direct RR	Plausibly	High	Composite	Medium	N/A	N/A
Van Iddekinge and Henry (2006)	.23	187	Direct RR	Plausibly	High	Composite	Medium	N/A	N/A
Van Iddekinge, Roth, Sager, and Heffner (2005)	.08	440	Direct RR	Unlikely	High	BDI	N/A	Low	0.0%
Vernon and Parry (1949)	.10	411	Double RR	Unlikely	Low	Other	Medium	High	N/A
Vernon and Parry (1949)	.31	460	Double RR	Unlikely	Low	Other	Medium	High	N/A
Villanova, Bernardin, Johnson, and Dahmus (1994)	.17	161	Direct RR	Unlikely	Low	N/A	Low	N/A	N/A
Walters, Miller, and Ree (1993)	.05	223	Direct RR	Unlikely	High	Other	Medium	Medium	25.0%
Zaccaria, Dailey, Tupes, Stafford, Lawrence, and Ailsworth (1956)	.12	83	Incumbents	Unlikely	Medium	Other	Medium	N/A	0.0%

[a]% cognitive was coded post hoc at the request of reviewers. Thus, we only coded % cognitive for the 40 samples that were included in our final analyses. Therefore, some samples that were not included in final analyses may have provided adequate information to assign a percentage, but those samples were simply listed as "N/A" in this table.

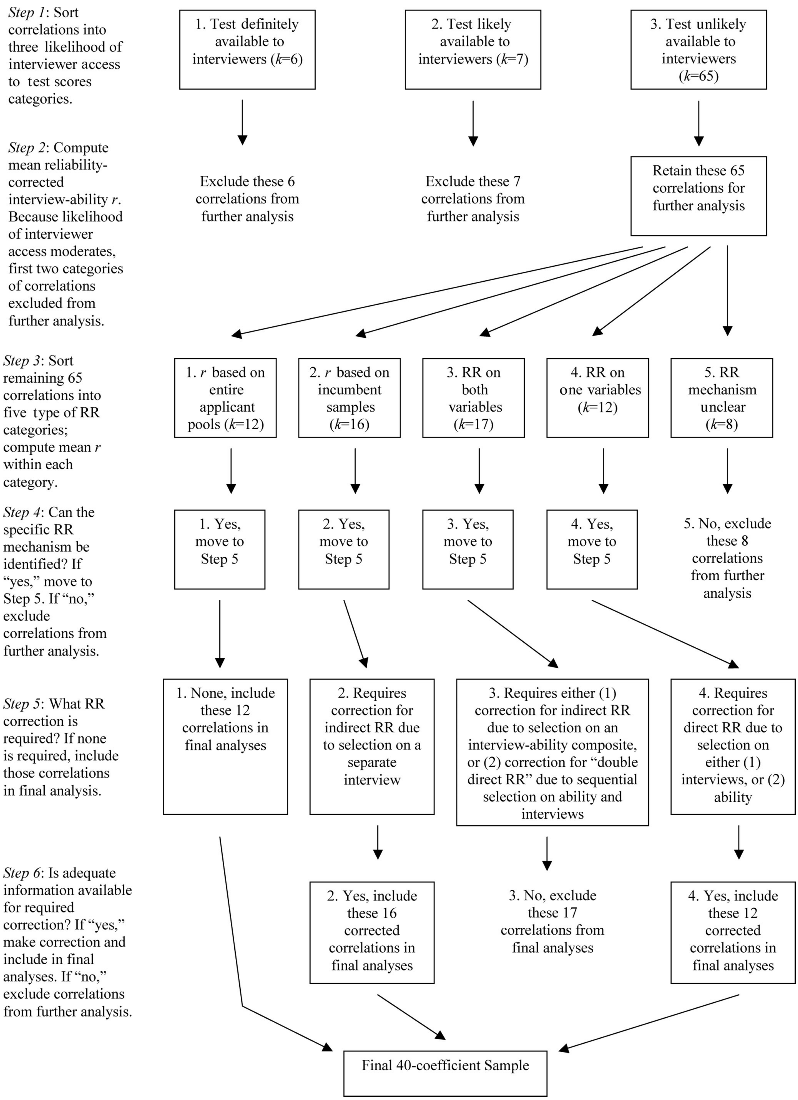

Figure 1: **Six-Step Process for Determining Whether Primary Studies Would be Included in Final Analyses.**

interviewer access to applicants' test scores categories. In Step 2, for each of the three categories of coefficients, we calculated the mean sample-size-weighted correlation between interview and cognitive test scores corrected for unreliability in interview and test scores (we did not correct for RR at this point because we suspected that [a] the type of RR would moderate interview–cognitive test score correlations and [b] some types of RR, such as RR due to selection on both interview and test scores, would not be correctable). Because availability of applicants' test scores moderated relationships, samples in which test scores were at least plausibly available to interviewers (interviewer access categories 1 and 2) were excluded from further analyses as their inclusion may have confounded meta-analytic estimates.

Similarly, if type of RR moderated relationships, then any sample that suffered from a type of RR that was not possible to correct for (or if it was not possible to determine whether the sample suffered from RR) was excluded from further analyses. Thus, in Step 3 the remaining samples were sorted into our five type of RR categories and then Step 4 addressed whether the specific RR mechanism affecting study correlations could be identified in each of the five categories. If the specific RR mechanism could not be identified (this was the case for Category 5), those correlations were excluded from further analyses as the inclusion of such possibly restricted but uncorrectable correlations would have confounded estimates. In Step 5 we addressed what the appropriate RR correction was for each category of the remaining samples (applicant pool samples required no RR correction so they were included in final analyses). In Step 6 we determined whether adequate information was available to apply the RR corrections identified in Step 5. If adequate information for correction was available, RR corrections were applied and RR-corrected correlations were included in final analyses. If adequate information was not available for a type of RR category, coefficients in that category were excluded from further analyses as the inclusion of such restricted but uncorrectable correlations would have confounded estimates. The result of our six-step process was a final database containing only those samples in which interviewers did not have access to applicants' cognitive test scores and in which there was no RR or RR was correctable. All pertinent meta-analyses and moderator analyses described earlier in this paper were then carried out on this final uncontaminated sample.

The Hunter–Schmidt Meta-Analysis Program (Schmidt & Le, 2004) computer software was used to arrive at meta-analytic estimates of the mean correlation and variability of the relationship between interview and cognitive test scores. In all analyses, mean sample-size-weighted correlations were computed. These mean sample-size-weighted correlations were corrected for RR when appropriate. RR corrections were not appropriate for applicant pool samples because these correlations did not have RR. Corrections were not appropriate for samples in which primary studies did not report enough information to determine the specific RR mechanism because there was no basis in such studies for assuming what type of RR (if any) existed. Corrections were not possible when correlations were restricted due to selection on a composite of both interviews and cognitive test scores because no primary study provided enough information to make the appropriate indirect RR correction. At least the following pieces of information would need to be reported in primary studies to make such a correction: (a) unrestricted composite SD, (b) restricted composite SD, (c) correlation between cognitive test score and composite, (d) correlation between interview and composite, and (e) any relative weights assigned to interview or cognitive scores when they were combined into composite form. Finally, corrections were not appropriate when correlations were restricted due to multiple hurdles selection using both interview and cognitive test scores because the appropriate correction would be to apply a direct RR correction on whatever variable was used as the second hurdle and then apply a direct RR correction on whatever variable was used as the first hurdle. In order to do this, one would need pieces of information such as the selection ratios used at each step of the selection process. This information was not provided in any primary studies.

Corrections for RR were only appropriate and possible in two categories of samples: samples with direct RR on only one variable (Type of RR category 4 in Figure 1) and samples comprising incumbents (Type of RR category 2). First, when direct RR resulted from selection on the interview, the RR ratio of .61 drawn from Salgado and Moscoso's (2002) direct RR on interviews artifact distribution was used to correct correlations. When direct RR resulted from selection on the cognitive test, the RR ratio of .67 drawn from Hunter and Hunter's (1984) direct RR on cognitive tests artifact distribution was used to correct correlations.

Second, incumbent samples were corrected for indirect RR. As previously mentioned, incumbent samples were likely affected by indirect RR resulting from selection on an interview correlated with the interview used in the primary study. Although no primary study is likely to report the needed information to correct for such indirect RR, the information can be estimated. Specifically, four pieces of information are needed to make such an indirect RR correction: (a) The mean sample-size-weighted correlation between the interview and cognitive test used in incumbent samples: This correlation was estimated in the present meta-analysis and was .13. (b) The mean correlation between the interviews causing the indirect RR (interview originally used to select incumbents for their current positions) and cognitive tests: In as much as the .13 value estimated for incumbents in the present meta-analysis is a population estimate, this second correlation should also be .13. (c) The mean correlation between the interviews causing indirect RR and the interviews used in primary studies: An estimate of this value can be drawn from Conway et al.'s (1995) meta-analysis of the interrater reliability of interviews. In their table 4, they list the meta-analytic estimates of interrater reliabilities for separate interviews at five different levels of interview question standardization. In as much as these are population estimates of the correlations between scores assigned to the same interviewee in separate interviews, the mean-sample-size-weighted correlation of these five different levels (.53) should estimate the correlation between the separate interviews in the primary studies included in the present meta-analysis. (d) The ratio of restricted standard deviations in incumbent samples to unrestricted standard deviations in the population is needed: Salgado and Moscoso (2002) provided such a value: .61. The above four pieces of information were used to correct incumbent samples for indirect RR.

All samples were corrected for unreliability in interview and cognitive test scores. Because there was not enough information in primary studies to make individual corrections for unreliability, the artifact distribution method was used. For cognitive test scores, we used the same reliability coefficient (.90) as did Huffcutt et al. (1996) and Salgado and Moscoso (2002). For interview scores, we used reliability coefficients drawn from Conway et al.'s (1995) meta-analysis of the interrater reliability of interviews. Specifically, Conway et al. reported interrater reliabilities for each of their five increasing levels of interview question standardization (level 1 = .69, level 2 = .72, level 3 = .75, level 4 = .75, level 5 = .92). For each of our studies for which the level of question standardization could be determined, the appropriate Conway et al. estimate was used. When level of question standardization could not be determined, the mean of the five levels was used (.766).

The variability of the mean observed and corrected correlations were also calculated. Moderator analyses were carried out for all relationships in which each of the following were true: The absolute magnitude of corrected variability ($SD\rho$) was still large enough to suspect moderators, and enough information and enough samples existed for meaningful moderator analyses. For any analyses in which a moderator variable was correlated with another variable (e.g., a continuous moderator such as % cognitive correlated with interview–cognitive test correlations, correlations between moderators, etc.), we ran such correlations both with and without sample-size weighting. Patterns of correlation were similar in both cases, so for ease of interpretation and presentation, reported correlations between moderators and other variables do not use sample-size weighting.

Differences Between the Present and Previous Meta-Analyses

Performing a meta-analysis requires the researchers to make judgment calls (e.g., Wanous, Sullivan, & Mulinak, 1989). Previous meta-analyses of the relationship between interview and cognitive test scores have differed in terms of a number of judgment calls. For instance, whereas Huffcutt et al. (1996) retained for most analyses samples in which interviewers were allowed access to applicants' ability scores, Salgado and Moscoso

(2002) omitted such samples from all analyses. We also made a number of judgment calls in the present meta-analysis, and we wish to be very explicit about the ways in which they differed from the previous meta-analyses, especially given that we arrived at a notably lower estimate of the relationship between interview and cognitive test scores.

Besides differences in the sets of primary studies used, perhaps the most important difference, and the one which we believe is the greatest methodological contribution of the present study, is the way in which RR was handled. Although previous meta-analyses applied a correction for direct RR on interviews to the mean sample-size-weighted correlation, this study only applied RR corrections to samples in which (a) RR was present, and (b) enough information was provided to determine the specific RR mechanism.

Each meta-analysis also handled differently those samples in which interviewers were allowed access to applicants' cognitive test scores. Huffcutt et al. (1996) retained such samples for most analyses but provided some supplementary analyses excluding samples in which it was obvious that interviewers had access to test scores. Salgado and Moscoso (2002) excluded from all analyses any sample in which it was obvious that interviewers had access to test scores. The present meta-analysis, in addition to excluding samples in which it was obvious that interviewers had access to test scores, also excluded samples in which circumstantial evidence made it quite plausible that interviewers had access to applicants' test scores.

Sampling error was also handled differently in each meta-analysis. Although the present meta-analysis and Salgado and Moscoso (2002) weighted each correlation by its sample size, Huffcutt et al. (1996) used a categorical weighting scheme wherein correlations were weighted 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. As a check on whether our use of sample-size weighting affected results, we ran a number of our analyses using Huffcutt et al.'s weighting system. In no case did the sample-size-weighted estimates differ from the three-point weighting system estimates by more than .01.

Corrections for unreliability in interviews were also handled differently in the present meta-analysis. Huffcutt et al. (1996) corrected correlations using reliabilities drawn from Wiesner and Cronshaw's (1988) meta-analysis. Salgado and Moscoso (2002) corrected conventional interviews using a reliability estimate drawn from Conway et al. (1995) and corrected behavior interviews using a reliability estimate drawn from Salgado and Moscoso (1995). The present meta-analysis drew reliability estimates from Conway et al. (1995) for five different levels of interview question standardization and applied whichever estimate was appropriate for the level of question standardization in each interview.

In Salgado and Moscoso (2002) and Huffcutt et al. (1996), when the same interviewees were presented with different types of interview questions and separate correlations were provided for these different types of interview questions, these were treated as separate correlations drawn from independent samples. This study treated such correlations as being drawn from only one sample and called these interviews “composite interviews.” If the Huffcutt et al. and Salgado and Moscoso model had been followed, the present meta-analysis would include 91 coefficients instead of 78.

In some studies correlations between an interview and multiple cognitive tests were reported, but the correlation between the interview and the entire cognitive test battery was not reported. In such cases the present meta-analysis and Salgado and Moscoso (2002) used composite formulas to estimate the correlation between the interview score and a composite of the multiple cognitive test scores when enough information was available. When enough information was not available, the mean was used. Huffcutt et al. (1996), on the other hand, used the highest correlation between the interview and any cognitive test component as an estimate of what the composite correlation would have been.

Results

Availability of Applicants’ Test Scores

The first two steps of our six-step process determined whether the availability of applicants’ cognitive test scores to interviewers moderated the relationship between interview and cognitive test scores. Table 2 lists preliminary meta-analytic estimates based on the full 78-coefficient sample before excluding any confounded samples, correcting mean sample-size-weighted correlations only for unreliability in interviews and cognitive test scores (RR was not corrected at this point because we suspected that the type of RR mechanism would be a moderator and some types of RR would not be correctable). In the six samples in which test scores were definitely available to interviewers, the corrected correlation was .44 whereas it was only .22 in the 65 samples in which it was unlikely that test scores were available. In the seven samples in which circumstantial evidence made it plausible that interviewers had access to test scores, the corrected correlation was .32. Thus, given that as the likelihood that test scores were available to interviewers increased, the correlation between test and interview scores also increased, we decided to eliminate this confound by excluding from any further analyses the 13 samples in which it was at least

TABLE 2

Preliminary Meta-Analytic Estimates of Interview–Ability Correlations Prior to Exclusion of Confounded Samples

Analyses	Total N	k	r_{mean}	SD_r	Corrected r[a]	SD_ρ	10% CV	90% CV
Total sample	20,014	78	.20	.1385	.24	.1536	.04	.44
Test definitely available	1,540	6	.35	.2437	.44	.2988	.06	.83
Test plausibly available	1,969	7	.27	.0454	.32	.0123	.30	.33
Test unlikely available	16,505	65	.18	.1196	.22	.1260	.06	.38

[a]Mean sample-size-weighted correlation (corrected only for unreliability in interview and ability scores).

k = number of correlations; r_{mean} = sample-size-weighted mean observed correlation; SD_r = sample-size-weighted observed standard deviation of correlations; SD_ρ = standard deviation of corrected correlations; 10% and 90% CV = 10% and 90% credibility values, respectively.

plausible that interviewers had access to applicants' cognitive test scores (see also step 2 in Figure 1).

Type of RR Mechanism

The next step was to sort the remaining 65 samples into the five type of RR categories and determine whether the type of RR mechanism moderated the relationship between interview and cognitive test scores (Step 3 in Figure 1). Table 3 lists meta-analytic estimates for each of the five type of RR categories based on these 65 samples, again correcting mean sample-size-weighted correlations only for unreliability in interviews and cognitive test scores. The first noteworthy finding outlined in Table 3 is that sizable numbers of studies fell into each of our five types of RR categories. This is important because it has generally been standard practice in meta-analysis to apply a direct RR correction to the mean sample-size-weighted correlation (implicitly correcting all studies for the same amount of RR) when RR is suspected. Only 12 of the samples in this set of 65 studies were directly restricted on either the interview or ability test used in the primary study, and only three of these samples were directly restricted due to selection on the interview. That so few samples were restricted due to selection on the interview should not be surprising because it is probably rare for organizations to screen applicants using a more expensive interview before sending them on to take a cognitive test. Thus, a blanket direct RR correction on interviews would only be appropriate in 3 out of the 65 samples in the present meta-analysis.

TABLE 3

Meta-Analytic Estimates of Interview–Ability Correlations Excluding Studies in Which Interviewers Were Allowed Access to Applicants' Cognitive Test Scores

Analyses	Total N	k	r_{mean}	SD_r	Corrected r[a]	SD_ρ	10% CV	90% CV
r Based on entire applicant pool	6,891	12	.24	.1139	.29	.1335	.12	.46
r Based on incumbent samples	2,332	16	.13	.1187	.16	.1050	.02	.29
Range restriction on both variables	2,950	17	.15	.1005	.18	.0871	.06	.29
Range restriction on one variable	1,900	12	.14	.1219	.16	.1028	.03	.29
Range restriction mechanism unclear	2,432	8	.12	.0746	.15	.0594	.08	.23

[a]Mean sample-size-weighted correlation corrected only for unreliability in interview and cognitive ability scores.

k = number of correlations; r_{mean} = sample-size-weighted mean observed correlation; SD_r = sample-size-weighted observed standard deviation of correlations; SD_ρ = standard deviation of corrected correlations; 10% and 90% CV = 10% and 90% credibility values, respectively.

The second noteworthy finding outlined in Table 3 is that the type of RR moderated correlations between interview and cognitive test scores. The only samples in Table 3 that are free of RR are the 12 samples in which the reported correlation was based on an entire applicant pool. In these samples, which are free from RR and have been corrected for unreliability in interview and cognitive test scores, the corrected correlation is .29. Some type of RR was likely present in each of the other four categories of samples in Table 3, and corrected correlations in these four categories ranged from .15 to .18.

Correlations Corrected for RR

Steps 4–6 of our six-step process determined whether appropriate RR corrections could be made in each of our Type of RR categories. We only included in our final analyses those samples that we could be reasonably certain were free from the confounding effects of RR. Therefore, we included in our final analyses only those samples that were either based on entire applicant pools or that were affected by a RR mechanism for which we could correct. For reasons explained earlier, we were not able to apply corrections to studies with RR on two variables or studies in which the RR mechanism was unclear, so these two categories of studies were excluded from final analyses (see also Steps 4 and 6 in Figure 1). This resulted in three categories being retained for final analyses (see also Steps 5 and 6 in Figure 1). These three categories were correlations based on the entire applicant pool (because there was no RR), correlations based on job incumbent samples (because we could correct for indirect RR), and correlations in which the sample was directly restricted on only one variable (because we could correct for direct RR on either interviews or cognitive tests).

When these last two categories of samples were corrected for their respective types of RR, both categories had corrected correlations of .24 (Table 4), which converge relatively closely with the unrestricted applicant pool-corrected estimate of .29. As can be seen in Table 4, combining the three categories of samples with known and correctable amounts of RR resulted in a final database of 40 samples that were not confounded by the effects of RR or availability of test scores to interviewers. The mean fully corrected correlation in these 40 samples was .27, which is noticeably lower than overall mean estimates from previous meta-analyses.

Moderator Analyses Using the Final 40-Coefficient Sample

Moderator analyses (job complexity; % cognitive; and interview structure, content, and validity) were carried out on the final database of 40 samples. Unfortunately, some moderator categories contained very few samples (e.g., low structure, low complexity). Therefore, many of the conclusions drawn from moderator analyses on this final sample should only be considered tentative.

Table 4 lists initial moderator results for categorical moderators. The general pattern of results is that interview–cognitive test correlations increase as structure and validity increase, as job complexity decreases, and when interviews are situational (BDI had especially low correlations whereas "composite" and "other" had moderate correlations). % Cognitive was not listed in Table 4 because it was a continuous moderator; so it was correlated with fully corrected interview–cognitive test correlations to test its moderating effect. % Cognitive was correlated .016 with interview–test correlations and thus did not act as a meaningful moderator.

These moderator results are difficult to interpret, though, because many of the moderators were confounded with each other. Table 5 lists correlations between each of the moderators (categorical moderators have been dummy coded), and it is apparent that correlations between many moderators were high. For instance, though Table 4 shows that high structure interviews tended to have lower interview–cognitive test correlations, Table 5 demonstrates that high structure interviews also tended to be BDI and to be used for jobs of medium or high complexity. Thus, it is unclear

TABLE 4

Meta-Analytic Estimates of Interview–Ability Correlations From Studies With Understood and Correctable RR Mechanisms

Analyses	Total N	k	r_{mean}	SD_r	Fully corrected r[a]	SD_ρ	10% CV	90% CV
Total correctable RR sample	11,317	40	.20	.0879	.27	.1328	.10	.44
r Based on entire applicant pool	6,891	12	.24	.1139	.29	.1335	.12	.46
r Based on incumbent samples	2,526	16	.13	.1187	.24	.0676	.16	.33
r Corrected for direct RR	1,900	12	.14	.1219	.24	.1692	.02	.45
Interview structure—high	8,429	27	.16	.0951	.22	.1003	.09	.35
Interview structure—medium	1,970	6	.38	.0948	.48	.0781	.38	.58
Interview structure—low	302	3	.14	.1608	.29	.0000	.29	.29
Interview content—behavior description	3,170	10	.12	.0919	.19	.0000	.19	.19
Interview content—situational	1,243	5	.26	.0517	.34	.0000	.34	.34
Interview content—composite	810	9	.17	.2262	.27	.2657	−.07	.61
Interview content—other	4,887	9	.23	.1290	.29	.1462	.10	.48
Job complexity—high	4,663	19	.15	.0717	.21	.0530	.14	.28
Job complexity—medium	3,791	14	.18	.0848	.24	.0537	.17	.31
Job complexity—low	2,159	5	.38	.1106	.49	.1333	.32	.66
Interview validity—high	3,185	13	.30	.1601	.41	.1593	.21	.61
Interview validity—medium	1,535	6	.19	.1240	.27	.1028	.13	.40
Interview validity—low	4,581	9	.15	.0733	.19	.0339	.15	.24

[a]Mean sample-size-weighted correlations drawn from incumbent samples were corrected for unreliability and indirect RR on interviews. Mean sample-size-weighted correlations drawn from entire applicant pools were only corrected for unreliability because there was no RR in these samples. Mean sample-size-weighted correlations drawn from samples with RR on one variable were corrected for unreliability and direct RR on either interview or cognitive test, as appropriate. k = number of correlations; r_{mean} = sample-size-weighted mean observed correlation; SD_r = sample-size-weighted observed standard deviation of correlations; ρ = mean corrected correlation (corrected for artifacts mentioned above); SD_ρ = standard deviation of corrected correlations; 10% and 90% CV = 10% and 90% credibility intervals, respectively.

TABLE 5

Correlations Between Study Moderators

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Type of RR														
1. *r* Based on entire applicant pool														
2. *r* Based on incumbent samples	−.54													
3. *r* Corrected for direct RR	−.43	−.54												
Interview structure														
4. High structure	.04	−.03	.00											
5. Medium structure	.09	.08	−.16	−.78										
6. Low structure	−.17	−.05	.21	−.52	−.14									
Interview content														
7. Behavior description	.24	.10	−.33	.14	−.04	−.17								
8. Situational	−.04	.15	−.12	.20	−.16	−.11	−.28							
9. Composite	−.19	.03	.14	.29	−.23	−.16	−.40	−.26						
10. Other	−.03	−.25	.29	−.59	.40	.42	−.40	−.26	−.38					
Job complexity														
11. High complexity	−.34	.16	.17	−.03	−.03	.09	−.42	.05	.24	.15				
12. Medium complexity	.19	.05	−.25	.27	−.16	−.22	.42	.04	−.28	−.18	−.76			
13. Low complexity	.24	−.31	.10	−.32	.24	.16	.03	−.14	.06	.03	−.39	−.30		
Uncorrected interview validity														
14. Validity	−.28	.30	−.06	.07	.32	−.54	−.11	.11	.35	−.34	−.09	.09	.01	
% Cognitive														
15. % Cognitive	−.07	.21	−.16	.15	−.17	.03	−.18	.31	−.09	.01	−.13	.22	−.24	−.03

$N = 26$–40; Categorical moderators were dummy coded, whereas "Validity" and "% Cognitive" were treated as continuous variables.

TABLE 6

Fully Corrected Interview–Cognitive Test Correlations in the Final 40-Coefficient Sample Regressed on Moderators

Moderator levels	*B*	SE_B	β	*R*
Intercept	.076	.077		.624
Interview structure				
Medium structure	.143	.095	.328	
Low structure	.075	.127	.128	
Interview content				
Situational	.163	.096	.360	
Composite	.126	.084	.344	
Other	.082	.090	.224	
Job complexity				
Medium complexity	.081	.071	.238	
Low complexity	.222*	.095	.461*	
% Cognitive				
% Cognitive	.001	.002	.099	

$^{*}p < .05$; $N = 30$–38; Except for “% Cognitive,” all variables were dummy coded (e.g., for “Medium structure,” if an interview was medium structure that interview was assigned a 1, whereas interviews that were not medium structure were assigned 0s), and one level for each moderator was excluded from the analysis. Pairwise deletion was used, as the question of interest was whether each of the primary study correlations falling into each moderator category were confounded with each of the other moderator categories, regardless of whether any specific study provided complete information for all moderators.

in this instance whether the lower interview–cognitive test correlation is due to interview structure, content, or job complexity.

Therefore, to facilitate interpretation of the independent contributions of each moderator, we simultaneously regressed the fully corrected interview–cognitive test correlations on each moderator in the final 40-coefficient sample (see Table 6). The regression was run using both weighted least squares (e.g., Steel & Kammeyer-Mueller, 2002) and ordinary least squares, with virtually identical patterns of results. Thus, for the sake of interpretability of statistical significance, only the ordinary least squares results are presented. Although this use of multiple regression facilitates interpretation of independent effects, we stress that these regression results are likely very unstable due to the high ratio of predictors (8 moderator levels) to data points (ranging between 26–38), and due to the very small number of samples in many moderator categories. Further, although we are unable to calculate the reliability of our dependent variable (fully corrected correlations), we note that the reliability should be reduced due to artifact corrections. Only one moderator level was statistically significant: Interviews for low complexity jobs were positively related to interview–cognitive test correlations. Still, the pattern of regression results generally substantiated the moderator results from Tables 4 and 5. Holding other moderators constant, interview–cognitive test correlations were lower when interview structure was high, when interviews were BDI, and when job complexity was high. Interview-cognitive test correlations were highest when interviews were SI, when interview structure was medium, and when job complexity was low (though only six and five correlations, respectively, contributed to these last two moderator levels). The percentage of cognitive interview dimensions had virtually no effect on interview–cognitive test correlations.

Because the regression results demonstrated that, when all moderators were held constant, the low complexity moderator level was the only statistically significant level, it was possible that each of the results thus far reported in this study were unduly influenced by low complexity samples. For instance, perhaps it was not really interviewer access to applicants' cognitive test scores that inflated interview–test relationships; instead interviewers were simply more likely to have access to applicants' scores when the job was of low complexity. To test for such possibilities, we reran all analyses reported in Tables 2, 3, and 4 with low complexity samples omitted. No patterns of relationships changed enough to overturn study conclusions, except in the case of interview validity. When low complexity samples were omitted, mean corrected interview–test correlations were .25, .28, and .20 in high, medium, and low validity samples, respectively. Thus, the pattern reported in Table 4, wherein interview–test correlations increase as validity increases, disappears when low complexity samples are omitted from analysis.

Discussion

Summary of Findings

This study argued for a careful analysis of each primary study in meta-analyses. Such a careful analysis identified a number of samples in the present meta-analysis in which interviewers were allowed access to applicants' cognitive test scores. Availability of test scores was viewed as a likely confound when assessing the relationship between interview and cognitive test scores, as our focus was determining how much ratings of applicants are affected by impressions of the applicants' cognitive ability formed through interaction with the applicants, not by looking at applicants' cognitive test scores. Indeed, as the likelihood that interviewers had access to test scores increased, so did the correlation between interview and test scores. Therefore, these samples were excluded from final substantive analyses.

An anonymous reviewer brought up the possibility that our "test unlikely available" category ($k = 65$) of samples is not an adequately precise category. The reviewer suggested splitting samples in this category into two subcategories: (a) samples in which it can definitely be determined that interviewers did not have access to applicants' test scores (e.g., article explicitly states this was the case or interviews were administered before cognitive tests), and (b) samples in which it cannot be definitively determined but circumstantial evidence made it unlikely. If our retention of samples in our "test unlikely available" category was justified, correlations in the two subcategories should be relatively comparable. Of our 65 "test unlikely available" samples, 21 fell into the first subcategory and had a mean correlation corrected for unreliability of .21. The remaining 44 samples had a mean corrected correlation of .22, providing evidence that our "test unlikely available" category was likely adequately meaningful and precise.

The present study's careful analysis also identified five different categories of samples, with each category representing a different type of RR (or lack thereof). This study excluded from final substantive analyses any studies that were either affected by RR that was uncorrectable or were possibly affected by RR but the specific mechanism was unknown. This approach resulted in a final database of 40 samples whose correlations were not confounded with interviewer access to cognitive test scores or RR. In these 40 samples, correlations drawn from entire applicant pools, from incumbent samples corrected for indirect RR, and from samples corrected for direct RR on only the one restricted variable converged on corrected correlations of .24–.29 between interviews and cognitive tests, noticeably lower than corrected correlations reported in previous meta-analyses. This range of correlations represents the best estimate to date of the mean unrestricted and unattenuated correlation between interviews and cognitive tests because these estimates were based on a sizable number of samples in which statistical and methodological artifacts should be of almost no consequence.

Our finding of a lower correlation than previous meta-analyses reflects the multiple differences between this study and prior ones. Prior meta-analyses used a different, but overlapping, set of studies, as well as a number of different decision rules. Our meta-analysis points toward a number of methodological points that future meta-analysts of predictor interrelationships should take into account. Of most importance to meta-analyses in general is our treatment of RR and the effect this had on corrected estimates. We note that our uncorrected meta-analytic estimates do not differ from previous meta-analyses' uncorrected estimates as much as our corrected estimates differ from previous meta-analyses' corrected estimates. This is as it should be. Because the technique we offer for dealing with RR was not widely known when previous meta-analyses were carried out, previous meta-analyses' corrected correlations were higher than they would have been if this study's techniques had been used (because they corrected all studies for direct RR even though many samples were unrestricted or not directly restricted), but their uncorrected correlations were not higher than they would have been if this study's techniques had been used (because no corrections were required for previous meta-analyses to arrive at mean uncorrected correlations). Thus, between the present and previous meta-analyses, uncorrected correlations should only differ by a relatively small amount due to sampling error and some differential representation of moderator categories (assuming that the proportions of samples affected by each type of RR are relatively similar in the present and previous meta-analyses), whereas correlations corrected for RR should differ relatively widely.

We believe it is instructive to demonstrate how our own conclusions would have differed had we used the RR correction methods of previous interview–cognitive test meta-analyses. Table 7 provides side-by-side comparisons of fully corrected meta-analytic estimates for full sample and categorical moderator analyses using the present meta-analysis' method versus previous meta-analyses' method of correcting for RR. That is, when the present meta-analysis' method was used, coefficients were only corrected for RR and unreliability and included in final analyses when those samples had known and correctable types of RR (or had no RR). When previous meta-analyses' method was used, the sample-size-weighted mean correlation was corrected for direct RR (using Salgado and Moscoso's direct RR artifact distribution value of .61) and unreliability without first accounting for specific RR mechanisms and excluding confounded samples. Note that in order to not confound comparisons between the two methods in Table 7, samples in which interviewers had access to applicants' test scores were excluded. Previous meta-analyses' RR correction method always resulted in higher estimates, mostly ranging from about 20% to 40% higher than estimates using our method (see far right column of Table 7). For instance, the fully corrected interview–cognitive test correlation across all samples using the present meta-analysis' method was .27 based on the 40-coefficient final sample. The same estimate using previous meta-analyses' RR correction method was 29.6% larger at .35 ($k = 65$ for the previous meta-analyses' method because previous meta-analyses would not have excluded the 25 samples with unknown or uncorrectable RR mechanisms). Thus, Table 7 makes clear how, even holding all else constant, applying a blanket direct RR correction instead of accounting for specific RR mechanisms can affect meta-analytic estimates.

The mean corrected correlation was also found to be moderated by a number of variables. Many of the moderator analyses were based on small

TABLE 7

Comparison of Fully Corrected Meta-Analytic Estimates Using RR-Correction Methods of the Present Versus Previous Meta-Analyses

Analyses	Fully corrected estimates using two different correction methods: Present meta-analysis' RR correction methods	Fully corrected estimates using two different correction methods: Previous meta-analyses' RR correction methods	Percentage over estimated using previous meta-analyses' methods[a]
Full sample	.27 (40)[b]	.35 (65)	29.6%
Interview structure—high	.22 (27)	.30 (37)	36.4%
Interview structure—medium	.48 (6)	.48 (11)	0.0%
Interview structure—low	.29 (3)	.35 (7)	20.7%
Interview content—behavior description	.19 (10)	.24 (11)	26.3%
Interview content—situational	.34 (5)	.44 (12)	29.4%
Interview content—composite	.27 (9)	.29 (10)	7.4%
Interview content—other	.29 (9)	.42 (14)	44.8%
Job complexity—high	.21 (19)	.29 (24)	38.1%
Job complexity—medium	.24 (14)	.31 (26)	29.2%
Job complexity—low	.49 (5)	.53 (13)	8.2%
Interview validity—high	.41 (13)	.52 (21)	26.8%
Interview validity—medium	.27 (6)	.35 (13)	29.6%
Interview validity—low	.19 (9)	.27 (17)	42.1%

[a]This reflects the degree to which previous meta-analyses' RR correction methods overestimate relative to this meta-analysis' methods; for example, .35 is 29.6% larger than .27 ([.35 − .27]/.27 = .296).

[b]Numbers in parentheses are the numbers of independent correlations (k) used in the meta-analytic estimate.

numbers of samples, so conclusions drawn can only be considered tentative. Future research should study the interview–cognitive test correlations in the moderator levels for which there was not enough information in this study.

Still, some tentative conclusions can be reached about moderators. First, high interview structure tends to result in lower interview–cognitive test correlations. Second, the greater the complexity of the job for which applicants are interviewing, the smaller the interview–cognitive test correlation. Third, the greater the uncorrected criterion-related validity of interviews, the greater tends to be the correlation between interview and cognitive test scores. Fourth, the content of the interview moderates its correlation with cognitive tests, with BDI being least correlated and SI being most correlated with cognitive tests. Each of these moderator results echo findings from previous reviews of the literature (e.g., Huffcutt et al., 1996; Salgado & Moscoso, 2002), but the magnitudes of correlation are generally considerably lower. Finally, the percentage of interview dimensions designed to capture cognitively oriented attributes did not have an effect on the interview–cognitive test correlation.

This last finding at first appears counterintuitive because it seems that interviews designed to measure more cognitively oriented constructs should correlate more highly with cognitive test scores. However, even if interviewers are not explicitly instructed to take account of applicants' cognitive ability, applicants with higher cognitive ability may speak more fluently or be better able to answer interview questions satisfactorily. For instance, Klehe, Koenig, Melchers, Kleinmann, Richter, and Vaccines (2006) demonstrated there were individual differences in applicants' ability to ascertain the constructs an interview was intended to measure and that this ability to ascertain was related to both verbal ability and interview performance. Thus, applicants with higher cognitive ability may simply be doing better in interviews, and thus, interviewers tend to rate them higher regardless of whether interviewers are explicitly trying to account for applicants' cognitive ability.

One possible issue with our % cognitive results could be that we were too liberal in what we coded as "cognitively oriented" interview dimensions. For instance, although scores in interview dimensions such as "ability to plan, organize, and prioritize" may to some degree be a function of cognitive ability, the case could be made that such an interview dimension is not as purely cognitive as dimensions such as "intellectual capacity." Thus, we reran our moderator regression including as cognitively oriented dimensions only those dimensions that were very clearly cognitive (e.g., intellectual capacity, ability to learn, problem solving). Regression results (both in terms of total variance accounted for and beta weights) were virtually identical, so for the purposes of this study, liberal versus conservative decision rules regarding which dimensions were cognitively oriented made no substantive difference in results.

We again stress, though, that many of these moderator results are likely to be unstable due to the small number of coefficients included in some moderator categories such as low and medium structure, and low complexity. The results for medium structure and low complexity are illustrative of the problem. Table 4 lists the fully corrected correlations for medium structure and low complexity as .48 and .49, respectively, correlations that are considerably higher than most other moderator categories. However, because the medium structure and low complexity categories only contained six and five correlations, respectively, both results were heavily influenced by one large sample study (Reeb, 1969) with an especially high interview–cognitive test correlation (uncorrected $r = .44$). We chose to include Reeb (1969) because the sample comprised an entire applicant pool and because the study explicitly stated interviewers were not allowed access to applicants' cognitive test scores. However, if Reeb (1969) were excluded, the fully corrected medium structure interview–cognitive test correlation would be reduced from .48 to .37, and the fully corrected low complexity correlation would be reduced from .49 to .40. Thus, the instability of such small *k* estimates should be clear, and we caution against overinterpreting moderator results based on such relatively small samples.

Implications for Incremental Validity of Interviews

As mentioned before, the correlation between interview and cognitive test scores has important implications for the incremental validity of interviews. It is therefore interesting to substitute our estimates of the interview–cognitive test correlation with well-known incremental validity analyses, such as those of Schmidt and Hunter (1998). Our reanalysis of these incremental validity analyses is by no means an indictment against Schmidt and Hunter's (1998) excellent work. Schmidt and Hunter were using Huffcutt et al.'s (1996) estimate of the relationship between interview and cognitive test scores, which was the best estimate up until that point in time. Schmidt and Hunter (1998) used correlations between cognitive tests and structured and unstructured interviews of .30 and .38 (mean sample-size-weighted correlations corrected for direct RR but not unreliability), respectively, drawn from Huffcutt et al. (1996) to estimate the incremental validity of the interview over cognitive ability. This study identified 40 samples in which the effects of RR could be controlled. The correlation in these samples between high structured interviews and cognitive tests ($k = 27, N = 8{,}429$) comparable to the value used by Schmidt and Hunter is .19 (this mean sample-size-weighted correlation is corrected for indirect RR in only the incumbent samples, for direct RR in only the direct RR samples, and is not corrected for unreliability). Unfortunately, only three of the interviews in the final 40-coefficient database were low structure interviews, so a comparable analysis of the incremental validity of low structure interviews was not possible. Using Schmidt and Hunter's methods and numbers, but using .19 as the correlation between high structure interviews and cognitive tests, to estimate the multiple correlation for job performance regressed on cognitive test and interview scores, this multiple correlation is .66, higher than any other combination of predictors reported by Schmidt and Hunter (1998), including integrity and conscientiousness tests. Thus, one tentative conclusion that might be made from the evidence in this paragraph is that given the high criterion-related validity of the interview (especially of structured interviews; i.e., McDaniel, Whetzel, Schmidt, & Mauer, 1994), and its low correlation with cognitive tests, the interview may be a useful supplement to cognitive tests for many employers.

Additional Issues, Limitations, and Implications for Future Research

One caveat, though, is that interviews are a method of obtaining information, not a measure of a specific construct. Interviews can be constructed to be highly ability saturated or to be relatively uncorrelated with cognitive ability. Even specific types of interviews, such as BDI that have generally been shown to have quite small correlations with cognitive tests, could be constructed with questions specifically tapping cognitive ability. So, the results of this meta-analysis do not guarantee that all interviews will have unrestricted and unattenuated correlations with cognitive tests, similar to the estimates reported in this meta-analysis. The present meta-analysis simply reports what the state of the literature is up until this point in time.

One possible criticism is that in the final analysis we excluded so many studies. Our counter is that because primary studies generally do not report enough information to understand the exact RR mechanisms at work, the meta-analyst in this domain is faced with only two options. One is to use RR artifact distributions to correct the mean correlation, which is the equivalent of applying the same correction to each individual correlation. The other is to only use samples in which it can be ascertained that RR is of little or no consequence or in which the meta-analyst can be quite certain that an appropriate RR correction is possible. As we had a relatively sizable set of studies ($k = 40$, $N = 11{,}317$) in which RR processes were well understood, we chose the latter and encourage future meta-analysts to do the same. We note that the exclusion of primary studies in meta-analyses due to concern about confounds in those primary studies is not a new strategy. For instance, Ones, Viswesvaran, and Schmidt (1993), in their meta-analysis of the relationship between integrity tests and job performance, collected 222 independent samples containing a correlation between integrity tests and job performance. Ones et al. based substantive conclusions on only 23 of these samples incorporating job applicants in predictive validity designs. We agree with this decision by Ones et al. to sacrifice total sample size in the name of only including interpretable coefficients in a meta-analysis.

One limitation of our study that came about due to our exclusion of studies in the final analysis is that high interview structure and high job complexity samples were overrepresented in our final sample. All else equal, these two study characteristics minimize the correlation between interview and cognitive test scores. Still, our estimates of the correlation between interview and cognitive test scores in high structure and high complexity samples are much lower than in previous meta-analyses. For instance, Huffcutt et al. (1996) reported a mean correlation in high structure interviews of .35 whereas the comparable estimate in our meta-analysis was .22; Huffcutt et al. reported a mean correlation in high complexity samples of .30 whereas the comparable estimate in our meta-analysis was .21.

This overrepresentation of high structure and complexity samples is an unfortunate consequence of a lack of reporting of information important to meta-analysts in primary studies, a problem that has been lamented by meta-analysts for quite some time (e.g., Hunter et al., 2006; Schmidt et al., 2006). For some reason, high structure and high complexity studies were more likely to report needed information. We therefore (a) strongly encourage researchers of primary studies to include as much information as possible regarding RR, and (b) strongly encourage further research in the correlation between interview and cognitive test scores in low structure and low job complexity samples.

Conclusions

The results of this meta-analysis demonstrate that cognitive ability does not commonly saturate the selection interview as much as our field has previously concluded. This provides scientists and practitioners alike with a clearer picture of both how much interview scores are typically reflective of applicants' cognitive ability and what is the selection interview's incremental validity over cognitive ability. This meta-analysis also serves as an example of the importance of carefully examining each primary study before including them in meta-analyses. It is hoped that future meta-analysts will follow our model.

REFERENCES

References marked with an asterisk (*) provided one or more independent primary samples for the meta-analysis.

Adoption of Additional Questions and Answers To Clarify and Provide a Common Interpretation of the Uniform Guidelines on Employee Selection Procedures as They Relate to the Internet and Related Technologies, 69 Fed Reg. 10152-10158 (Mar. 4, 2004).

*Albrecht PA, Glaser EM, Marks J. (1964). Validation of a multiple-assessment procedure for managerial personnel. *Journal of Applied Psychology, 48*, 351–360.

Arthur W, Day EA, McNelly TL, Edens PS. (2003). A meta-analysis of the criterion-related validity of assessment center dimensions. PERSONNEL PSYCHOLOGY, *56*, 125–154.

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

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

*Bosshardt MJ. (1994). *Situational interviews versus behavior description interviews: A comparative validity study*. Unpublished dissertation, University of Minnesota.

*Bradley KM, Bernthal PR, Thomas JN. (1998, May). *Factor structure and predictive validity of structured employment interviews: An application of industrial psychology in a Kuwaiti petrochemical corporation*. Poster presented at the 13th Annual Conference of the Society for Industrial and Organizational Psychology Dallas, TX.

*Calkins EV, Richards JM, McCanse A, Burgess MM, Willoughby TL. (1974). Impact on admission to a school of medicine of an innovation in selection procedures. *Psychological Reports, 35*, 1135–1142.

Campbell JT, Prien EP, Brailey LG. (1960). Predicting performance evaluations. PERSONNEL PSYCHOLOGY, 13*, 435–440.

*Campion MA, Campion JE, Hudson JP, Jr. (1994). Structured interviewing: A note on incremental validity and alternative question types. *Journal of Applied Psychology, 79*, 998–1102.

Campion MA, Pursell ED, Brown BK. (1988). Structured interviewing: Raising the psychometric properties of the employment interview. PERSONNEL PSYCHOLOGY, 41*, 25–42.

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

Conway JM, Jako RA, Goodman DF. (1995). A meta-analysis of interrater and internal consistency reliability of selection interviews. *Journal of Applied Psychology, 80*, 565–579.

Cortina JM, Goldstein NB, Payne SC, Davison HK, Gilliland SW. (2000). The incremental validity of interview scores over and above *g* and conscientiousness scores. PERSONNEL PSYCHOLOGY*, 53*, 325–351.

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

*Delery JE, Wright PM, McArthur K, Anderson DC. (1994). *g* tests and the situational interview: A test of incremental validity. *International Journal of Selection and Assessment, 2*, 53–58.

*Denning DL. (1993). *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.

*Dicken C. (1969). Predicting the success of Peace Corps community development workers. *Journal of Consulting and Clinical Psychology, 33*, 597–606.

*Dipboye RL, Gaugler BB, Hayes TL, Parker DS. (1992). *Individual differences in the incremental validity of interviewers' judgments*. Unpublished manuscript.

Drasgow F. (2003). Intelligence and the workplace. In Borman WC, Ilgen DR, Klimoski RJ (Eds.), *Handbook of psychology: Volume 12, industrial and organizational psychology* (pp. 107–130), Hoboken, NJ: Wiley.

*Durivage A, St. Martin J, Barrette J. (1995). Intelligence pratique ou traditionnelle: Que mesure l'entrevue structure situationnelle? *European Review of Applied Psychology, 45*, 171–178.

*Elam CL, Andrykowski MA. (1991). Admission interview ratings: Relationship to applicant academic and demographic variables and interviewer characteristics. *Academic Medicine, 66*, 13–15.

*Elliot AGP. (1981). Sex and decision making in the selection interview: A real-life study. *Journal of Occupational Psychology, 54*, 365–273.

*Fox S, Spector PE. (2000). Relations of emotional intelligence, practical intelligence, general intelligence, and trait affectivity with interview outcomes: It's not all just "G". *Journal of Organizational Behavior, 21*, 203–220.

Friedland D. (1973). *Selection of police officers*. Los Angeles, CA: 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.
Ghiselli EE, Campbell JP, Zedeck S. (1981). *Measurement theory for the behavioral sciences*. San Francisco. W. H. Freeman & Co.
Gowing MK, Slivinski LW. (1994). A review of North American selection procedures: Canada and the United States of America. *International Journal of Selection and Assessment, 2*, 103–114.
Gulliksen H. (1950). *Theory of mental tests*. New York: Wiley.
*Handyside JD, Duncan DC. (1954). Four years later: A follow-up of an experiment in selecting supervisors. *Occupational Psychology, 28*, 9–23.
*Hilliard PA. (2000). *Comparison of the predictive validity of a written test, an integrity test, a conscientiousness questionnaire, a structured behavioral interview and personality inventory in the assessment of job applicants' background investigations, and subsequent task and contextual job performance*. Unpublished dissertation.
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 AI, Conway JM, Roth PL, Stone NJ. (2001). Identification and meta-analytic assessment of psychological constructs measured in employment interviews. *Journal of Applied Psychology, 86*, 897–913.
Huffcutt AI, Roth PL, McDaniel MA. (1996). A meta-analytic investigation of *g* in employment interview evaluations: Moderating characteristics and implications for incremental validity. *Journal of Applied Psychology, 81*, 459–473.
Huffcutt AI, Weekly JA, Wiesner WH, Degroot TG, Jones C. (2001). Comparison of situational and behavior description interview questions for higher-level positions. PERSONNEL PSYCHOLOGY, 54*, 619–644.
Hunter JE, Hunter RF. (1984). Validity and utility of alternative predictors of job performance. *Psychological Bulletin, 96*, 72–98.
Hunter JE, Schmidt FL. (2004). *Methods of meta-analysis* (2nd ed.). Thousands Oaks, CA: Sage.
Hunter JE, Schmidt FL, Judiesch MK. (1990). Individual differences in output variability as a function of job complexity. *Journal of Applied Psychology, 75*, 28–42.
Hunter JE, Schmidt FL, Le H. (2006). Implications of direct and indirect range restriction for meta-analysis methods and findings. *Journal of Applied Psychology, 91*, 594–612.
Huse EF. (1962). Assessments of higher-level personnel: IV. The validity of assessment techniques based on systematically varied information. PERSONNEL PSYCHOLOGY, 15*, 195–205.
*Johnson EK. (1990). *The structured interview: Manipulating structuring criteria and the effects on validity, reliability, and practicality*. Unpublished doctoral dissertation, Tulane University.
*Johnson EM, McDaniel MA. (1981, June). *Training correlates of a police selection system*. Paper presented at the International Personnel Management Association conference, Denver, CO.
*Klehe UC. (2003). *Towards an understanding of the constructs underlying the situational interview and the patterned behavior description interview in predicting typical versus maximum performance*. Unpublished dissertation.

Klehe UC, Koenig C, Melchers KG, Kleinmann M, Richter G, Vaccines C. (2006, May). Candidates ability to identify criteria in structured interviews and its predictive power of performance in an assessment center. In Ferstle KL, Klehe UC (Chairs). *New insights into constructs underlying structured interview performance and validity*. Symposium presented at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

*Klehe UC, Latham GP. (2005). The predictive and incremental validity of the situational and patterned behavior description interviews for teamplaying behavior. *International Journal of Selection and Assessment, 13*, 108–115.

*Largo E. (1994). *An assessment of the incremental validity of a structured employment interview*. Unpublished dissertation.

*Lievens F, Harris MM, Van Keer E, Bisqueret C. (2003). Predicting cross-cultural training performance: The validity of personality, *g*, and dimensions measured by an assessment center and a behavior description interview. *Journal of Applied Psychology, 88*, 476–489.

Lopez FM. (1966). Current problems in test performance of job applicants: I. PERSONNEL PSYCHOLOGY, 19*, 10–18.

*Loven MJ. (1998). *Selection factors related to satisfaction, persistence, and job performance for technical writers at the Iowa State University Child Welfare Project*. Unpublished dissertation.

Marsden PV. (1994). Selection methods in U.S. establishments. *Acta Sociologica, 37*, 287–301.

McDaniel MA, Whetzel DL, Schmidt FL, Mauer SD. (1994). The validity of employment interviews: A comprehensive review and meta-analysis. *Journal of Applied Psychology, 79*, 599–616.

*Melchers KG, Klehe UC, Richter GM, Kleinmann M, Konig CJ. (2005). *The impact of interviewees' ability to identify criteria in structured interviews*. Unpublished manuscript.

*Menkes J. (2002). *What do structured interviews actually measure? A construct validity study*. Unpublished dissertation.

*Meredith KE, Dunlap MR, Baker HH. (1982). Subjective and objective admissions factors as predictors of clinical clerkship performance. *Journal of Medical Education, 57*, 743–751.

*Motowidlo SJ, Carter GW, Dunnette MD, Tippins N, Werner S, Burnett JR, et al. (1992). Studies of the structured behavioral interview. *Journal of Applied Psychology, 77*, 571–587.

*Newman SH, Bobbitt JM, Cameron DC. (1946). The reliability of the interview method in an officer candidate evaluation program. *American Psychologist, 1*, 103–109.

Ones DS, Viswesvaran C, Schmidt FL. (1993). Comprehensive meta-analysis of integrity test validities: Findings and implications for personnel selection and theories of job performance. *Journal of Applied Psychology, 78*, 679–703.

*Philips AP, Dipboye RL. (1989). Correlational tests of predictions from a process model of the interview. *Journal of Applied Psychology, 74*, 41–52.

Pulakos ED, Schmitt N. (1995). Experience-based and situational interview questions: Studies of validity. PERSONNEL PSYCHOLOGY, 48*, 289–308.

*Pulakos ED, Schmitt N. (1996). An evaluation of two strategies for reducing adverse impact and their effects on criterion-related validity. *Human Performance, 9*, 241–258.

*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, FL: City of Miami Personnel Department.

*Rhea BD, Rimland B, Githens WH. (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 PL, Campion JE. (1992). An analysis of the predictive power of the panel interview and pre-employment tests. *Journal of Occupational and Organizational Psychology, 65*, 51–60.

Ryan AM, McFarland L, Baron H, Page R. (1999). An international look at selection practices: Nation and culture as explanations for variability in practice. PERSONNEL PSYCHOLOGY*, 52*, 359–391.

Sackett PR, Yang H. (2000). Correction for range restriction: An expanded typology. *Journal of Applied Psychology, 85*, 112–118.

Sackett PR, Lievens F, Berry CM, Landers R. (2007). A cautionary note on the effects of range restriction on interpredictor correlations. *Journal of Applied Psychology*, 92, 538–544.

Salgado JF, Moscoso S. (1995). Validity of the behavior structured interviews. *Psicologia del Trabajo y las Organizaciones, 11*, 9–24.

Salgado JF, Moscoso S. (2002). Comprehensive meta-analysis of the construct validity of the employment interview. *European Journal of Work and Organizational Psychology, 11*, 299–324.

Schmidt FL, Hunter JE. (1998). The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin, 124*, 262–274.

Schmidt FL, Le HA. (2004). Software for the Hunter-Schmidt Meta-Analysis methods [computer software]. Iowa City: University of Iowa. Department of Management and Organizations.

Schmidt FL, Oh IS, Le H. (2006). Increasing the accuracy of corrections for range restriction: Implications for selection procedure validities and other research results. PERSONNEL PSYCHOLOGY*, 59*, 281–305.

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

*Shahani C, Dipboye RL, Gehrlein TM. (1991). The incremental contribution of an interview to college admissions. *Educational and Psychological Measurement, 51*, 1049–1061.

*Sparks CP. (1973a). *Validity of a selection program for process trainees* (Personnel Research No. 73–1). Unpublished manuscript.

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

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

*Sparks CP. (1978). *Preemployment tests for pipefitter trainee* (Personnel Research No. 78–19). Unpublished manuscript.

Steel PD, Kammeyer-Mueller JD. (2002). Comparing meta-analytic moderator estimation techniques under realistic conditions. *Journal of Applied Psychology, 87*, 96–111.

*Sue-Chan C, Latham GP. (2004). The situational interview as a predictor of academic and team performance: A study of the mediating effects of *g* and emotional intelligence. *International Journal of Selection and Assessment, 12*, 312–320.

Thorndike RL. (1949). *Personnel selection*. New York: Wiley.

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

*Van Iddekinge CH, Eidson CE, Jr. (2005). *Using personality tests and structured interviews to select customer service managers*. Unpublished manuscript, Clemson University.

*Van Iddekinge CH, Henry MS. (2006, May). Are facets of cognitive ability differentially predictive of interview performance? In Ferstl K, Klehe UC (Chairs), *New insights into constructs underlying structured interview performance and validity*. Symposium to be conducted at the 21st Annual Conference of the Society for Industrial and Organizational Psychology, Dallas, TX.

*Van Iddekinge CH, Roth PL, Sager CE, Heffner TS. (2005). *A construct-oriented investigation of a structured employment interview*. Paper presented at the 20th Annual Conference of the Society for Industrial and Organizational Psychology, Los Angeles, CA.

*Vernon PE, Parry JB. (1949). *Personnel selection in the British forces*. London: University of London Press.

Villanova P, Bernardin HJ, Johnson DL, Dahmus SA. (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 LC, Miller MR, Ree MJ. (1993). Structured interviews for pilot selection: No incremental validity. *International Journal of Aviation Psychology, 3*, 25–38.

Wanous JP, Sullivan SE, Malinak J. (1989). The role of judgment calls in meta-analysis. *Journal of Applied Psychology, 74*, 259–264.

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.

*Zaccaria MA, Dailey JT, Tupes EC, Stafford AR, Lawrence HG, Ailsworth KA. (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.