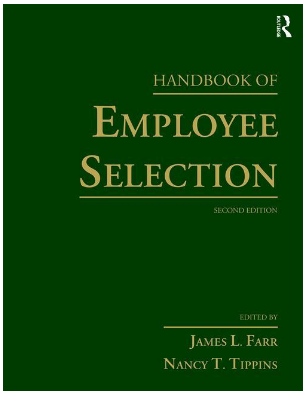

Handbook of Employee Selection

James L. Farr, Nancy T. Tippins, Walter C. Borman, David Chan, Michael D. Coovert, Rick Jacobs, P. Richard Jeanneret, Jerard F. Kehoe, Filip Lievens, S. Morton McPhail, Kevin R. Murphy, Robert E. Ployhart, Elaine D. Pulakos, Douglas H. Reynolds, Ann Marie Ryan, Neal Schmitt, Benjamin Schneider

Cognitive Ability

Publication details
https://test.routledgehandbooks.com/doi/10.4324/9781315690193-11
Deniz S. Ones, Stephan Dilchert, Chockalingam Viswesvaran, Jesús F. Salgado
Published online on: 22 Mar 2017

11

COGNITIVE ABILITY

Measurement and Validity for Employee Selection

DENIZ S. ONES, STEPHAN DILCHERT, CHOCKALINGAM VISWESVARAN, AND JESÚS F. SALGADO

Cognitive ability (or "intelligence") affects individuals' lives in countless ways, and it influences work lives of employees perhaps to a greater extent than any other individual differences trait. Cognitive ability determines whether an employee will be able to acquire the required job knowledge and perform assigned tasks. It is the strongest predictor of learning and acquisition of job knowledge as well as overall job performance. It is remarkably relevant regardless of the occupation one holds (Ones, Dilchert, & Viswesvaran, 2012). It even predicts extrinsic career success (i.e., earnings and promotions). As such, it is an exceedingly important trait to include in employee selection systems.

In this chapter, we provide an overview of cognitive ability's key role in staffing organizations and provide evidence-based practice recommendations. We first present a brief synopsis of the history, current usage, and acceptance of cognitive ability tests in employee selection. Second, we highlight the theoretical underpinnings and structure of cognitive ability as a construct. Third, we discuss developments in its measurement. Fourth, we present an overview of the criterion-related validity of cognitive ability tests in predicting valued work behaviors and outcomes, including non-task-performance criteria that have been increasingly investigated in recent years. Fifth, we discuss the issue of group differences in cognitive ability test scores both within the United States and internationally. We conclude by discussing future research and challenges facing organizations that intend to use cognitive ability tests in making employee selection decisions.

HISTORY, CURRENT USAGE, AND ACCEPTABILITY OF COGNITIVE ABILITY MEASURES IN EMPLOYEE SELECTION

It has been more than 110 years since the publication of Spearman's influential (1904) article "'General Intelligence,' Objectively Determined and Measured."[1] Early in the 20th century, researchers began to study the usefulness of cognitive ability measures for predicting learning and performance in educational settings. For personnel decision making, standardized, objective cognitive ability tests first saw large-scale use in military settings. Group tests of intelligence were developed prior to World War I and used extensively during both World Wars. European and U.S. armed forces continued to utilize cognitive ability tests for selection and placement, and many business organizations followed suit. However, during the 1940s, '50s, and '60s, research revealed much variability, particularly with regard to the supposed usefulness of such measures to predict job performance. It seemed that the specific jobs under investigation, specific organizational settings, the particular ability measures used, and many unidentified (and unidentifiable) factors all contributed to the variability of observed results (e.g., Hull, 1928). Moreover, validation results differed even when jobs, organizations, and measures were held constant. Industrial psychologists came to believe that subtle, undetectable differences in situations were responsible for differences observed in the predictive value of the test studied. By the 1960s, this belief in situational specificity dominated the scientific literature and was well entrenched among practitioners (see Chapter 4, this volume). The breakthrough came in the 1970s. Frank Schmidt and Jack Hunter demonstrated (Schmidt & Hunter, 1977) that most differences observed across studies of cognitive ability were due to sampling error (sample sizes for validation studies in the 1960s displayed a median of 68; see Lent, Aurbach, & Levin, 1971), differences in level of restriction of range in samples (typically employees in concurrent studies who had already been selected into an organization), and differences across studies in the unreliability of criterion measurement (typically supervisory ratings of job performance). These statistical and measurement artifacts were responsible for the differences in results observed across most previous validation studies. The invention of meta-analysis (known as "validity generalization" in the employee selection literature), and the consistent findings from meta-analytic studies discredited the theory of situational specificity and paved the way to systematic investigations of predictor validity, also reaching beyond the domain of cognitive abilities.[2]

Today, cognitive ability measures are used in educational admissions and civilian personnel staffing, but how widespread is the use of cognitive ability measures in organizational settings in general, as well as vis-à-vis other tools available for personnel decision making? At the turn of the 21st century, the most comprehensive survey was conducted by Ryan, McFarland, Baron, and Page (1999). Ryan and colleagues surveyed 959 organizations from 20 countries by randomly sampling 300 large organizations (with more than 1,000 employees) in each country. The focus of their study was the examination of national and cultural influences on many selection system features. The pervasiveness of cognitive ability tests was also surveyed. Across the 18 countries for which data were reported by Ryan and colleagues, on average, cognitive ability tests were used between 21% and 50% of the time in employee selection. Organizations in the Netherlands, Belgium, Portugal, Spain, South Africa, New Zealand, and the United Kingdom reported above average use, whereas organizations in Germany, Hong Kong, and Italy reported especially low levels of cognitive test use. Within each country, of 14 selection methods presented to respondents (cognitive ability tests, physical ability tests, foreign language tests, work samples, personality tests, integrity tests, interest inventories, simulation exercises, situational judgment tests [SJTs], video-based tests, projective techniques, drug tests, medical screens, and graphology), cognitive ability tests were ranked in the top three most frequently utilized methods in 15 of 18 countries. It is of value to note that some of the methods listed, such as SJTs or simulations, can be used to measure a variety of constructs, and thus data on their use are not necessarily directly comparable to that of construct-specific ones such as standardized tests of cognitive ability and personality. However, it appears that if objective tests are utilized at all in personnel staffing decisions, cognitive ability measures are included with frequency.

Several other, often region-specific, surveys have documented similar prevalence rates as well as other interesting trends. Salgado and Anderson's (2002) summary of such studies revealed that cognitive ability test use seems more common for selection into graduate and managerial-level positions compared to low-complexity jobs. More recent surveys (e.g., Chartered Institute of Personnel Development, 2007; Taylor, Keelty, & McDonnell, 2002) also seem to suggest an increase in cognitive ability test use, at least in some countries. In addition, cognitive ability tests appear to be used more frequently by larger organizations compared with smaller ones (Salgado, in press). Unfortunately, even when considering these more recent surveys, the available data on the extensiveness of cognitive ability test use come from countries that are not entirely representative of the world's cultural regions. Data from Eastern Europe (e.g., Ukraine, Russia) and southern Asia (e.g., India, Pakistan) are meager; systematic, large-scale surveys from Latin

America, the Middle East, and Africa are also lacking. Research studies on other issues relating to cognitive ability tests (e.g., validity, group differences) are increasingly being published by authors in these regions, which might be interpreted as an indicator that their use in practice is also increasing (see, for example, Barros, Kausel, Cuadra, & Diaz, 2014; Kriek & Dowdeswell, 2009; Thadeu & Ferreira, 2013). Countries from these world regions offer a unique opportunity for industrial-organizational (I-O) psychologists to assess cultural variability in the extensiveness of use of as well as reactions to cognitive ability tests.

Prevalence data provide an index of *organizational acceptance* of selection tools. Another perspective on this issue can be gained by examining *applicants' acceptance*. Applicant reactions to selection tests vary by test type (Kluger & Rothstein, 1993). There are now many international (including some comparative) studies of applicant reactions to selection tests (see Bertolino & Steiner, 2007; Moscoso, 2006; Nikolaou & Judge, 2007, Ryan et al., 2009). An early meta-analysis by Hausknecht, Day, and Thomas (2004) of 10 studies on selection tool perceptions showed that the mean favorability ratings for cognitive ability tests were lower than those for interviews, work samples, resumes, and references, but higher than those for (in descending order) personality tests, biodata, personal contacts, honesty tests, and graphology. However, this meta-analysis included both respondents in laboratory research and field settings. Caution is warranted in drawing conclusions about reactions of actual job applicants to cognitive ability tests on the basis of these results: participants in the studies contributing to the Hausknecht et al. meta-analysis were not necessarily applying for jobs, were not in selection settings, and did not experience each of the tools they were rating. A more recent meta-analysis by Anderson, Salgado, and Hülsheger (2010), which included studies from all countries listed above (and summarized data for job applicants as well as some "student surrogate" samples), found that cognitive ability tests are among the selection procedures rated most favorably by job applicants. Those authors conclude that the high favorability of such tests, despite the drawback of being perceived as relatively impersonal, was due to perceptions of standardized tests being scientifically valid, respectful of applicants' privacy, and providing them with an opportunity to perform.

Recent research indicates there is cross-national similarity in organizational acceptance and use of cognitive ability test use for employee selection (Ryan et al., under review). Furthermore, applicants view cognitive ability tests relatively favorably, and again similarly so across several countries where data are available (e.g., see Ryan et al., 2009 for a 21-country investigation). Figure 11.1 presents meta-analytic data on applicant reactions to cognitive ability tests in comparison to one of the most favorably rated selection methods (interviews), the least favorably ranked method for each justice dimension, as well as mean applicant reactions across all methods investigated. In sum, cognitive ability tests are perceived more favorable than average on all relevant dimensions except "interpersonal warmth," and even exceed favorability ratings of interviews in terms of scientific evidence and respect for applicants' privacy.

Scholars have rightfully pointed out that applicant reactions to personnel selection procedures are largely determined by their perceived fairness and their perceived predictive validity (Chan & Schmitt, 2004). However, it has also been shown that cognitive ability is a common antecedent not only of performance on standardized tests but also of perceived test fairness and test-taking motivation (Chan, Schmitt, Jennings, Clause, & Delbridge, 1998; Reeve & Lam, 2007). This is also true for self-assessed performance. Applicants' "guesses" of how well they performed on cognitive tests have been demonstrated to relate to perceptions of predictive validity and job relatedness in several cultures (Ryan et al., 2009). Applicant perceptions of fairness are likely to present a challenge for cognitive ability tests as long as any traditionally disadvantaged group (broadly defined) scores systematically lower on a given predictor battery or applicants perceive a systematic bias hindering their test performance on such tests. Of course, this issue has been much discussed in relation to race and ethnic group mean score differences, particularly in the U.S. context. However, with changes in test technology and the increase of online testing, as well as new and innovative item formats to measure various cognitive abilities (especially inductive reasoning), similar concerns might occur with regard to other groups, such as older job applicants (see below for a discussion of oft-neglected age differences and potential for adverse impact). Tackling this issue will be a major task for our profession in the years to come if organizations and society as a whole is to benefit from the use of the most reliable

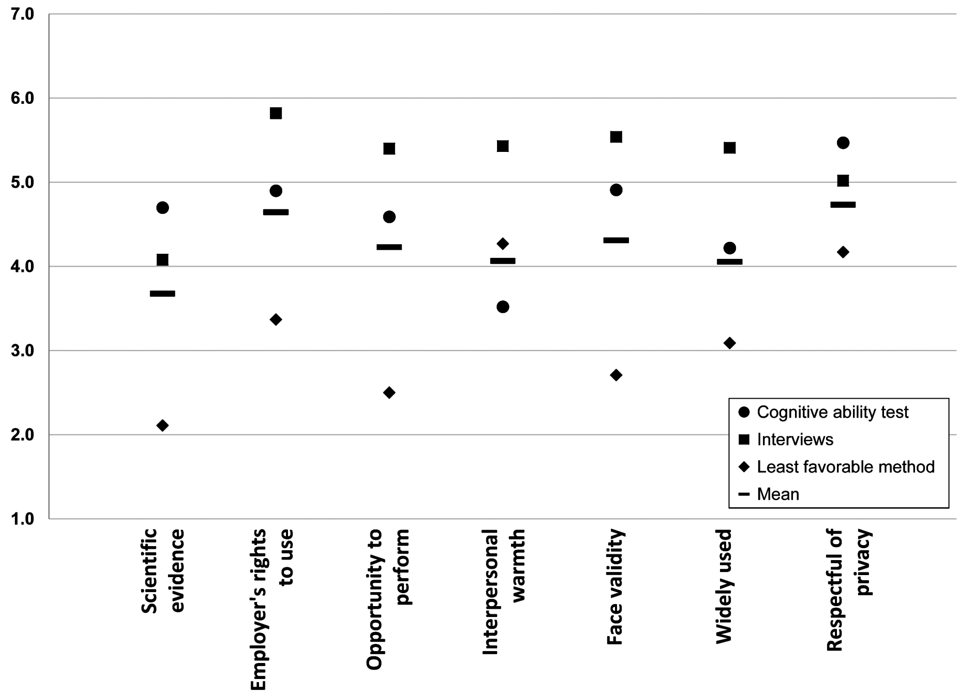

FIGURE 11.1 Applicant Reactions to Cognitive Ability Tests and Other Selection Methods by Dimension

Based on meta-analytic results from Anderson et al. (2010). Values represent sample-size weighted means for each method after all ratings were transformed to a 7-point scale; higher values indicate more favorable reactions. Graphology was excluded from consideration as the "least favorable" method due to lack of validity evidence and relative infrequency of use compared to all other methods compared. For the five dimensions, "personal contacts" was consistently the least favorably ranked method; for the dimensions "widely used" and "respectful of privacy," the least favorable method was honesty tests.

and valid assessments available for hiring and placements. Vocal opponents of high-stakes testing have promulgated myths about the ability of intelligence tests to predict valued outcomes, as well as their fairness (see Schmidt et al., 2007, for examples). These myths, although often entirely unsupported by empirical evidence or even common logic, are difficult to dispel and, if allowed to inform organizational decisions on selection tool use, pose a threat to organizations' and ultimately societies' economic welfare (Sackett, Borneman, & Connelly, 2008; Schmidt & Hunter, 1981).

DEFINITIONS AND THEORETICAL UNDERPINNINGS

The core of intelligence as a psychological construct has long been conceptualized as reasoning ability and a form of mental adaptability (Stern, 1911). Despite the central place this construct takes in determining individual behavior, it took almost a century for a broad scientific consensus to emerge on its definition. A group of 52 experts that included luminaries of psychological science defined intelligence as "a very general mental capacity that, among other things, involves the ability to reason, plan, solve problems, think abstractly, comprehend complex ideas, learn quickly and learn from experience" (Gottfredson, 1997, p. 13). This group of scholars, drawn from various psychological disciplines (including I-O psychology), goes on to state that intelligence "is not merely book learning, a narrow academic skill, or test-taking smarts. Rather, it reflects a broader and deeper capability for comprehending our surroundings—'catching on,' 'making sense' of things, or 'figuring out' what to do" (p. 13). In the words of William Stern, one of the forefathers of modern-day research on cognitive ability, intelligence is the "general mental adaptability to new problems and conditions of life" (2011, p. 3).

The importance of such a broad definition (in contrast to folk concepts such as "book smarts") cannot be overstated. Conceptually, intelligence, in humans and other species, indicates the complexity and efficiency of cognitive functioning. Here, complexity refers to the "sophistication of the intellectual repertoire" (Lubinski, 2004, p. 98), whereas the efficiency aspect refers to the effectiveness of information-processing skills. Both aspects are critical to performance in all domains of life (in interpersonal interactions, at home, school, or work), and their impact on individual differences in problem-solving ability can be observed in individuals of all ages. The realization that such information-processing skills "can be applied to virtually any kind of content in any context" (Gottfredson, 2004b, p. 23) is of relevance to scientists and practitioners alike. The application of this principle to organizational contexts forms the conceptual basis of Campbell's (1990) fundamental statement that "general mental ability is a substantively significant determinant of individual differences in job performance for any job that includes information-processing tasks" (p. 56). It is difficult to imagine any job that does not include information processing of some form.

Cognitive ability is an integral part in models of job performance because of its relation to knowledge and skill acquisition. General mental ability predicts job performance because it is a causal determinant of acquisition of job knowledge (McCloy, Campbell, & Cudeck, 1994; Schmidt, Hunter, & Outerbridge, 1986). The more cognitively demanding the knowledge to be acquired and the more complex the task to be performed, the greater is the relationship between cognitive ability and performance (Hunter & Hunter, 1984).

STRUCTURE OF COGNITIVE ABILITY

Although there are numerous specific cognitive abilities, they all share a common construct core: general mental ability, popularly called *intelligence*, or *g* (for *general* intelligence factor) in many scientific writings. As Gottfredson (2002) so aptly noted, the multitude of ways to measure *g* attest to its generality. Although measures of intelligence may look different, employ different item types (e.g., verbal, figural, numerical, cognitive/neuropsychological tasks), and use different formats (e.g., individually administered tasks, paper-and-pencil tests, computerized batteries, and even game-like mobile applications), this does not mean they assess entirely distinct constructs.

The structure of cognitive abilities has been examined extensively since Spearman's (1904) distinction of *g* and *s* (*specific* abilities). A century of research has yielded hundreds of data sets in which individuals took multiple cognitive ability measures. Carroll (1993) compiled, analyzed, and summarized the correlation matrices resulting from more than 460 such data sets. The result was his popular three-stratum model (see McGrew, 2009; Schneider & McGrew, 2012; for fuller descriptions of the Cattell-Horn-Carroll [CHC] model of intelligence). Cognitive abilities are hierarchically organized. At the apex is *g*, or an ability that is general. At the second stratum are group factors or broad abilities, including fluid reasoning, memory (short-term as well as long-term storage and retrieval), visual processing, processing speed (including perceptual speed), but also previously acquired knowledge (including quantitative ability, comprehension, reading and writing, and domain-specific knowledge). At the lowest level of the hierarchy are specific factors or narrow abilities such as induction for fluid abilities, ideational fluency for long-term memory, or lexical knowledge for comprehension. Individuals of similar intelligence (i.e., at the same trait level of general mental ability) differ in their standing on specific abilities because of differential "investment" of their cognitive capacity (guided by other personal characteristics as well as idiosyncratic developmental and educational experiences) in these narrow cognitive domains.

The distinction between fluid and crystallized intelligence (termed g_f and g_c, respectively; see Cattell, 1971) provides a still-popular conceptual model but has also been shown to distinguish between *g* and lower-level abilities, rather than ability factors at the same level of the taxonomical hierarchy. Fluid and crystallized intelligence tend to correlate around .70, and some scholars argue that g_f is indistinguishable from *g* (Gustafsson, 2002; other scholars go as far as classifying certain fluid ability tests, such as the *Raven's* matrices, as direct measures of *g*). The most significant *content* domains that surface in most ability models are verbal/linguistic, quantitative/numerical, and spatial/mechanical.

There are several other popular models of cognitive ability structure, both competing as well as converging (e.g., Vernon's, Cattell and Horn's, Holzinger's, Johnson & Bouchard's, as well as the Berlin model). Most of them are hierarchical in nature but differ in terms of both number of strata as well as nature of primary and secondary factors they postulate. An overview and illustration of these models is provided in Salgado (in press). Even though researchers continue to clarify and refine the structure of intelligence in individual studies (e.g., see Johnson & Bouchard, 2005; Johnson, te Nijenhuis, & Bouchard, 2007, 2008), the Cattell-Horn-Carroll (CHC) model (McGrew, 2009) still dominates the thinking about the structure of intelligence today.

What is undebated is that when various cognitive ability tests reflecting the entire range of intelligence from the general population are administered to test takers, a large proportion of variance can be attributed to a general factor. Lubinski (2004) found that in such cases, about 50% of the common variance is due to *g*, whereas 8–10% of the remaining common variance is attributable to verbal, quantitative, and spatial abilities (Lubinski, 2004).

It has been found that relationships among cognitive ability scales are weaker at higher levels of the ability spectrum (e.g., see Detterman & Daniel, 1989; Kane, Oakland, & Brand, 2006), implying a smaller amount of common variance due to *g*. Theoretically, one implication could be that there may be more room among high-ability individuals for specific abilities to yield incremental validities over tests of general mental ability. However, investigations of incremental validity in highly complex jobs have so far yielded mixed results. For example, Olea and Ree (1994) reported that specific abilities contributed little beyond *g* to the prediction of job performance among pilots and navigators, whereas Ree and Carretta (1996) concluded that some specific abilities had the potential to add incremental value at least for prediction of military pilot performance. If it were consistently found that the common variance among individual ability tests accounted for by *g* was smaller than in samples of broad talent, then it is plausible that specific abilities could add incremental value over general mental ability for such groups. For the prediction of training performance, there is some initial evidence to support this hypothesis in primary samples of apprentices in low- compared with medium-complexity jobs (Ziegler, Dietl, Danay, Vogel, & Bühner, 2011). However, both the Germanic context of these data (e.g., relatively high educational standards) and the reliance on suboptimal regression analyses in this research points to the need for replication in other countries, using more appropriate statistical approaches (Wiernik, Wilmost, & Kostal, 2015). Moreover, direct tests among *job applicants*, especially high-ability samples, and for the prediction of job performance are still called for. However, such investigations would have to sort out potentially complex range restriction effects in these samples.

When broad job categories and applicants of a wide range of talent are studied, analyses directed at incremental validities of specific ability measures over *g* have yielded disappointing results: Specific abilities do not provide substantial incremental validity over *g*. Nonetheless, in some meta-analyses, specific abilities have been shown to be similarly valid for the prediction of some criteria (see below). In addition, there may also be nonvalidity-related considerations for practitioners to include specific ability measures in their selection systems, such as the consequences of anticipated group differences or applicant reactions. Some survey results on the use of specific versus general mental ability test use seem to reflect such considerations, as specific ability tests see significant use in pre-hire assessments, albeit not at the same level as general mental ability tests (Krantowitz, 2014).

MEASUREMENT

The list of cognitive ability measures available to scientists, individual practitioners, and organizations runs in the hundreds and includes everything from simple, homegrown measures to tests of wide circulation supported by many decades of empirical research evidence. A discussion of the merits of individual measures cannot be provided in this chapter. However, a brief discussion of commonly used methods, as well as current trends in cognitive ability assessment, is warranted.

Traditional, standardized tests are the most widespread method for measuring all types of cognitive abilities. Their popularity is not because of a lack of alternative methods, but primarily because of their excellent reliability, ease of administration, and scoring. Although validity (including predictive validity) is the property of the inferences made about a psychological construct (e.g., the abilities measured by a test, not the test itself), the reliability of the assessment methods provides a ceiling for validities that can be obtained in applied settings. From this point of view, standardized tests provide the best solution for organizations looking to assess cognitive ability in a reliable, standardized, and objective manner.

The use of standardized tests in employee selection and academic settings is not without controversy. Unfortunately, criticism levied against cognitive ability tests, like that directed at other standardized testing, often falls victim to "content-format confusion" (Chan & Schmitt, 1997, 2004; Ryan & Greguras, 1998) and failure to distinguish the nature of the test response (Lievens, De Corte, & Westerveld, 2015). In addition to standardized tests, many other methods can be used to assess cognitive ability constructs, and a careful investigation of these methods and how they are typically used can inform decisions on whether they are suitable for a given purpose and setting. Interviews, assessment centers (ACs), and SJTs are all methods that assess cognitive ability to varying degrees—sometimes by design, sometimes by accident. Early meta-analyses estimated the overlap between interviews and cognitive ability at approximately $\rho = .40$ (Huffcut, Roth, & McDaniel, 1996). A more recent meta-analysis reported a mean, range-restriction corrected correlation of .27 ($N = 11{,}317$, $k = 40$; Berry, Sackett, & Landers, 2007). A re-analysis by Roth & Huffcutt (2013) showed that interviews conducted specifically in employment settings (versus for academic admissions) are more saturated with cognitive ability variance, in line with earlier findings ($\rho = .41$, $N = 840$, $k = 5$). However, the analysis by Berry and colleagues provides some intriguing moderator results, including higher interview-ability test correlations when interview validity is high and job complexity is low. Interviews with greater cognitive content can be expected to yield higher criterion-related validities. Also, for low-complexity jobs, interviews may function as more of a cognitive screen than for higher-complexity jobs.

Relationships between cognitive ability and overall AC ratings have also been examined. A meta-analysis by Collins et al. (2003) reported that cognitive ability test scores correlated .43 with overall AC ratings ($N = 5{,}419$, $k = 34$). AC dimensions may have a differential cognitive load. In a large-scale study, Dilchert and Ones (2009) reported that the highest correlations were found for the AC dimension problem solving ($r = .32$, $N = 4{,}856$), providing further evidence for the fact that cognitive ability measures capture real-world problem-solving abilities, including those displayed in business simulations (cf. Gottfredson, 1997). In an integrative meta-analysis of the AC literature, Meriac, Hoffman, and Woehr (2014) reported similar findings, with the AC dimensions problem solving, communication, and organizing/planning all displaying mean unreliability-corrected correlations of .29 with GMA. Meriac and colleagues estimated that the general factor that spans AC dimensions (see Kuncel & Sackett, 2013) is correlated .26 with GMA (the maximum correlation with any of the Big Five personality dimensions was .14 with Extraversion). Assessment center exercises are similarly related to GMA. Hoffman, Monahan, Lance, and Sutton's (2015) meta-analysis reported unreliability corrected correlations of .30 for in-baskets, but relations in the range of .13 to .22 for leaderless group discussions, role plays, case analyses, and oral presentations.

Increasingly popular SJTs are also correlated with cognitive ability; however, the magnitude of the correlation depends on the instructions given to participants. Knowledge instructions (e.g., "what should one do," "rate the best/worst option") in completing SJTs produce an observed correlation of .32 ($N = 24{,}656$, $k = 69$), whereas SJTs with behavioral tendency instructions (e.g., "what would you do") correlate .17 ($N = 6{,}203$, $k = .26$) with cognitive ability (McDaniel, Hartman, Whetzel, & Grubb, 2007). Thus, if job applicants complete SJTs, especially under knowledge instructions, assessments produce a ranking of job applicants on cognitive ability to a certain degree. However, Christian, Edwards, and Bradley's (2010) meta-analysis of SJT validity by construct domain indicates that job knowledge and skills-focused SJTs predict job performance at lower levels than those established for traditional cognitive ability tests. More construct-focused research on SJTs is warranted, however, as total sample sizes for these analyses were very small, and SJTs specifically designed to assess general mental ability were not included (likely because few such measures exist).

Many assessment methods increasingly rely on formats other than the traditional paper-and-pencil form, a trend that is also reflected in ability measurement. Earlier research used meta-analysis to establish the equivalence of computerized and paper-and-pencil versions of cognitive ability tests (Mead & Drasgow, 1993). Recent trends in web-based assessment and test content delivery build on the fact that tests of many individual difference predictors (not only cognitive ability) have been shown to be equivalent between paper-and-pencil and computerized versions. However, the real challenge arises not from a change in test format but from a change in administration mode.

Web-based, unproctored cognitive ability assessment is aimed at streamlining the application process for applicants and organizations. Critics argue that this approach requires strong confidence in the honesty of test takers (who, at least in selection contexts, presumably have a strong incentive to cheat). Some organizations, confronted by the real-world challenges of having to assess hundreds of thousands of applicants every year, are already using computerized adaptive testing and constantly updated test materials to conduct unproctored web-based testing (Gibby, 2008). Although there is considerable range in estimates of the magnitude of applicant cheating on such assessments (Arthur, Glaze, Villado, & Taylor, 2010; Hense, Golden, & Burnett, 2009; Lievens & Burke, 2011), they are lower than initially assumed, and not high enough to justify foregoing the significant efficiencies realized by unproctored testing (Tippins, 2015). Commercial test publishers and assessment providers have also developed several strategies to address issues of cheating and test security, ranging from regular monitoring for item piracy and systematic, proctored retesting of test takers (Burke, 2008) to remote, video-based proctoring and biometric test taker identification (Foster, 2008), as well as algorithmic identity monitoring via means such as keystroke analysis, facial-, voice-, and even palm/knuckle recognition. We are certain that for large-scale assessments, such trends will soon become the everyday reality, dictated by demands for more streamlined assessment procedures from applicants and organizations alike (see also Chapter 39, this volume). The challenges posed by remote, unproctored cognitive ability assessment will need to be addressed by a more intense collaboration of scientists and practitioners, as well as by drawing on expertise from outside of the psychological domain. The challenges are worth tackling, because the utility gains of expanding testing programs to larger numbers of test takers earlier in the hiring process, as well as improvements in fairness gained from reaching additional and unique applicant populations, are likely to outweigh the costs (see, for example, the simulations provided by Landers & Sackett, 2012). Organizations and providers that invest in the appropriate know-how and technology are increasingly at the forefront of big development in cognitive ability measurement.

CRITERION-RELATED VALIDITY EVIDENCE

The job relatedness and usefulness of standardized cognitive ability tests in employee selection have been documented in dozens of quantitative reviews in the form of publications and technical reports, incorporating more than 1,300 meta-analyses summarizing results from more than 22,000 primary studies. The total sample size of job applicants and employees providing data for these validation studies is well in excess of 5 million individuals (Ones, 2004; Ones & Dilchert, 2004). The question of whether cognitive ability tests are useful predictors of performance in occupational settings has been definitely answered: yes, they are excellent predictors of training performance and job performance. In fact, no other predictor construct in employee selection produces as high validities, as consistently, as does cognitive ability. In addition, no assessment method has so far achieved as reliable assessment of cognitive ability as standardized tests, making such tests the ideal choice for predicting performance in organizational settings.

Meta-analyses of cognitive ability test validities have been tabulated and summarized by Ones, Viswesvaran, and Dilchert (2005), Dilchert (in press), and Salgado (in press). In this section we provide an overview of conclusions from these quantitative reviews. Readers interested in the specific meta-analyses supporting each conclusion are encouraged to review these chapters.

Cognitive Ability Tests Predict Learning, Acquisition of Job Knowledge, and Job Training Performance with Outstanding Validity (Operational Validities in the .50 to .70 Range)

Validities for training criteria generalize across jobs, organizations, and settings. Meta-analyses provide voluminous evidence of high validity for training success in military and civilian organizations. Operational validities (correlations corrected for attenuation due to unreliability in criterion measures and range restriction, where applicable) are highest for general mental ability and specific quantitative and verbal abilities, and somewhat lower for memory (although still highly useful with a sample-size-weighted operational validity of .46). Validities are moderated by job complexity. The greater the complexity of jobs being studied, the higher the validity of cognitive ability tests in predicting training performance (Hunter & Hunter, 1984; Salgado, Anderson, Moscoso, Bertua, de Fruyt, & Rolland, 2003; Ziegler et al., 2011). Superior validities of cognitive ability tests for learning are in line with findings that cognitive ability is the strongest determinant of knowledge acquisition, in this case acquisition of job knowledge (Schmidt et al., 1986). The more complex jobs are, the more complex and vast the knowledge to be acquired. Brighter individuals learn more quickly, learn more, and can acquire more complex knowledge with ease.

Cognitive Ability Tests Predict Overall Job Performance with High Validity (Operational Validities in the .35 to .55 Range)

Table 11.1 summarizes the potential moderators of cognitive ability test validity in employment settings, indicating those supported and those rejected on the basis of meta-analyses, as well

TABLE 11.1

Hypothesized Moderators of Cognitive Ability Test Validities

Yes: Confirmed Moderators	*No: Rejected Moderators*	*?: Moderating Effect Unknown*
Job complexity	Situational variables	Time of study (historical age)
Criterion predicted	Organizational setting	Age
Training performance	Race	Race
Job performance	African Americans[a]	Asian Americans
Leadership	Hispanics[a]	Native Americans
Turnover, etc.	Sex[b]	National setting and culture (except for some countries)
Cognitive ability construct assessed	Military/civilian setting	
GMA	Validation design (concurrent/predictive)	
Verbal ability	Length of time on the job (up to 5 years)	
Memory, etc.	Method of criterion-measurement (e.g., ratings, production quantity, work samples)	

[a] Meta-analytic evidence for race and comparisons between Caucasians and African Americans as well as Hispanic/Latino Americans in civilian employment suggests operational validities for Whites may be .02 to .04 correlational points higher; differences in military settings were found to be somewhat higher.
[b] Meta-analytic evidence suggests operational validities for men may be negligibly higher.

as those awaiting investigation. Validities for overall job performance criteria generalize across jobs, organizations, and settings. Support for these key conclusions comes from meta-analyses of studies using narrow job groupings (e.g., mechanical repair workers, first-line supervisors, health technicians, computer programmers, lawyers, retail sales personnel, firefighters), broad job groupings (e.g., clerical jobs, law enforcement, maintenance trades), and heterogeneous job groupings (e.g., by job complexity). Individual large sample studies (e.g., Project A) also point to the same conclusions. Operational validities are highest for general mental ability and quantitative abilities and somewhat lower for memory (although still useful with a sample-size-weighted operational validity of .39 across 12 different meta-analyses; Ones & Dilchert, 2004). The method of performance measurement employed (objective vs. subjective) does not lead to different conclusions about the usefulness of cognitive ability tests, and different indices of performance (rankings, ratings, etc.) produce similar operational validities. Job complexity also moderates the validities of cognitive ability tests for predicting job performance. Higher validities are found for jobs of higher complexity. Although content validation has gained popularity in recent years, validity generalization studies have clearly demonstrated that matching specific cognitive abilities to aspects of task performance deemed important in a given job is not necessary. Basing selection systems on one or two specific abilities based on content validity evidence can be expected to result in lower levels of learning, less new job knowledge acquisition, and poorer adaptation to changing work environments (Ones, 2016). In sum, it is remarkable that even when moderators have been reported for cognitive ability test validity, they do not result in validities reversing direction or shrinking to negligible levels in magnitude. Useful levels of validity are found even for the more specific cognitive abilities and for lowest levels of job complexity.

A multitude of additional variables have been tested as potential moderators of validity in the meta-analyses reviewed in Ones et al. (2005), and most can be dismissed based on empirical evidence. These include organizational setting, method of criterion measurement (ratings, rankings, etc.; Nathan & Alexander, 1988), sex (see below), validation study design (concurrent vs. predictive; Barrett, Phillips, & Alexander, 1981), and length of time on the job (experience up to five years; Schmidt, Hunter, Outerbridge, & Goff, 1988). In sum, cognitive ability test validity does not vary substantially and systematically across organizational settings or for most subgroups that have been examined. Concurrent validities approximate predictive validities, and cognitive ability tests show no declines in validity as workers gain experience. There are, nonetheless, still some potential moderators waiting to be tested in large-scale, representative studies, or more systematically or thoroughly investigated using meta-analytic approaches. Studies of Asians and Native Americans as well as older adults are notably absent from the I-O psychology literature (see below for more details).

Although we know a great deal about the validity of cognitive ability tests for predicting training, task, and overall job performance criteria, knowledge of how cognitive ability relates to other aspects of work behavior (e.g., organizational citizenship) has been limited. Initial intriguing findings were reported by Alonso, Viswesvaran, and Sanchez (2008), who found that cognitive ability correlated more highly with contextual performance than personality factors. In a recent meta-analysis of 43 studies, Gonzalez-Mule, Mount, and Oh (2014) reported mean true-score correlation of .24 between cognitive ability and supervisor-rated organizational citizenship behaviors (the validity was corrected for indirect range restriction on the predictor measure and unreliability both the predictor and criterion measures).

Investigations of cognitive ability validities for counterproductive work behaviors (CWB) have also been reported. A large predictive validity study relating a cognitive ability measure to counterproductive behaviors indicated that intelligent individuals avoid engaging in organizational and interpersonal deviance on the job (Dilchert, Ones, Davis, & Rostow, 2007). A meta-analysis of 16 studies with non-self-report CWB criteria pointed out the need to better understand the CWB criterion domain (Gonzalez-Mule et al., 2014). More research into correlates of cognitive ability outside of the traditional job performance domain (e.g., adaptive performance, employee green behaviors) would be welcome.

So far, there have been no large-scale investigations of cognitive ability test validity across time (i.e., has validity for cognitive ability tests in general changed over time?). Labor force changes paired with progressive changes in the way work is done in many fields (more complex processes, greater use of technology) call for renewed inquiries. Of course, given the current reward system in academia and the overemphasis on theory that favors small-scale studies of new effects over large-scale replications (cf. Campbell & Wilmot, in press), new academically based research on cognitive ability test validities is unlikely (Salgado, in press). One hypothesis that we would like to offer is that cognitive ability tests today have greater validity than half a century ago. As job roles and tasks for jobs in most sectors change over time to include more complex tools (e.g., computers), processes (e.g., virtual teamwork), and requirements (e.g., multiple languages), the power of general mental ability as the basic learning skill in predicting performance may increase substantially, especially in mid- to high-complexity jobs.

Another change that has put increasing demand on individuals and organizations over the last two decades or so is internationalization. We already know that organizations, small and large, compete for customers in a global economy. However, in a time when mobility—real and virtual—is greater than ever in humanity's history, organizations now also compete internationally for their labor force or are faced with mobile labor forces. Validity of cognitive ability tests has been studied in international contexts, most extensively in Europe (Hülsheger, Maier, & Stumpp, 2007; Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003; Salgado, Anderson, Moscoso, Bertua, de Fruyt, & Roland, 2003), but also in Asia (Takahasi & Nimura, 1994; Nimura, Imashiro, & Naito, 2000; Oh, 2010; Lee, 2005). Table 11.2 summarizes the results of these international meta-analyses for job performance. Findings are mostly parallel to those from the United States: Cognitive ability tests show substantial validity, and higher validities are found for higher-complexity jobs. Moreover, highest validities are found for general mental ability rather than specific abilities (Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003). The only international exception to the pattern of very strong validities seems to exist in Japan. One explanation that has been put forth for this finding is that job performance evaluations in this context often emphasize non-task-performance aspects, such as citizenship behaviors, more strongly than in other cultures (Nimura et al., 2000; Salgado, in press). However, another

TABLE 11.2

Validity of Cognitive Ability Tests for Predicting Job Performance in International Contexts

Across European countries	*k*	*N*	*r*	ρ
High-complexity jobs	14	1,604	.23	.64
Medium-complexity jobs	43	4,744	.27	.53
Low-complexity jobs	12	864	.25	.51
Analyses by country				
Belgium and the Netherlands	15	1,075	.24	.63
France	26	1,445	.48	.64
Germany	8	746	.33	.53
Japan	126	26,095		.20
South Korea	8	1,098		.57
Spain	11	1,182	.35	.64
United Kingdom	68	7,725	.26	.56

N = total number of subjects; *k* = number of studies summarized in meta-analysis; *r* = sample size weighted mean observed correlation; ρ = operational validity, corrected only for sampling error and attenuation due to unreliability in the criterion.
Sources: Data across European countries summarized from Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P., *Journal of Applied Psychology, 88*, 1068–1081, 2003. Data for individual European countries except Germany summarized from Salgado, J. F., & Anderson, N., *European Journal of Work and Organizational Psychology, 12*, 1–17, 2003. Data for Germany summarized from Hülsheger, U. R., Maier, G. W., & Stumpp, T., *International Journal of Selection and Assessment, 15*, 3–18, 2008. Data for Japan based on three different meta-analyses, synthesized by Salgado (in press); data for South Korea based on two meta-analyses also synthesized by Salgado.

explanation might be that nearly half of the data summarized in the two available Japanese meta-analyses were collected with one specific ability test. Although the criterion-related validity of cognitive ability tests for predicting job performance is among the most established findings in applied psychology, there is room for additional research in specific cultural regions.

Several practical issues are noteworthy. First, it is often argued that educational requirements serve as proxies for cognitive ability. This argument suggests that using a cognitive ability test would not be necessary if a screening based on educational credentials were in place. There are two flaws in this line of reasoning. Educational qualifications of applicants to the same jobs are very similar, or at least more homogenous than those of the population at large. Conversely, even among those who hold advanced degrees (e.g., doctoral, medical, and law degrees), there is still substantial variability in cognitive ability (Sackett & Ostgaard, 1994; Wonderlic Inc., 2002), indicating room for utility to be derived from cognitive testing. Findings of highest predictive validities for the most complex jobs (e.g., lawyers, physicians) underscore the usefulness of cognitive tests even when there is homogeneity in the educational level and credentials of job applicants. If, for some reason, individuals of mixed educational level were to apply for the same job, a cognitive ability test is a more precise, valid, and efficient selection tool to use (Berry, Gruys, & Sackett, 2006).

It is also often suggested that beyond a certain required level of ability, cognitive capacity does not contribute to performance. Such arguments essentially suggest a nonlinear relationship between cognitive ability and performance: Cognitive ability predicts up to a certain point on the trait continuum, but validity drops off beyond that. The data that have been brought to bear on this question seem to tell the opposite story (Arneson, 2007; Coward & Sackett, 1990; Ghiselli & Kahneman, 1962; Hawk, 1970; Tiffin & Vincent, 1960). An issue commonly encountered in investigations of nonlinearity, however, is the lack of sensitivity of such investigations at the parts of the trait continuum that actually matter (in this case, high ability levels). This is due to both the sensitivity of the ability measures employed as well as the number of (extremely) high-ability individuals in the respective data sets. However, large-scale investigations that address these issues now exist. Arneson, Sackett, and Beatty (2011) provided a particularly strong illustration for the case of ability tests used in academic admissions decisions. Not only did these authors find no evidence for the "good-enough" hypothesis on ability–performance relationships, but they also showed that at high ends of the ability spectrum, the relationship with performance was typically *stronger*. For noneducational achievement criteria, Wai, Lubinski, and Benbow (2005) had previously shown that validity remains high even among the most extremely talented individuals. Altogether, this evidence suggests that if any nonlinear ability–performance relationships exist, they are more likely to be characterized by an exponential curve rather than an asymptotic relationship.

GROUP DIFFERENCES ON COGNITIVE ABILITY MEASURES

One of the greatest points of concern in using cognitive ability measures in the United States is the potential for adverse impact. In this section, we review mean group differences on cognitive ability measures and discuss their implications for adverse impact. We also review findings regarding predictive fairness and discuss group differences in international contexts.

Group Difference in Central Tendency and Dispersion

In the United States, Title VII of the Civil Rights Act prohibits employment discrimination on the basis of race, color, religion, sex, and national origin, and the Age Discrimination in Employment Act addresses age discrimination. Historically, disadvantaged racial and ethnic groups in the United States are African Americans (Blacks), Hispanic/Latino Americans, Asian Americans, and Native Americans/Pacific Islanders (see Chapter 29, this volume, for an international perspective). Women and older adults have also been historically disadvantaged in many contexts. If selection decisions result in selection ratios for subgroups of protected classes that are less than 80% of those for the better-performing group, presence of adverse impact is concluded, and the burden of proof shifts to the employer to establish the job relatedness of the selection tools utilized, typically using criterion-related validity evidence (see also Chapter 28, this volume).

However, it is important to remember that adverse impact (or lack thereof) in the employee selection process is the result of a selection system and not only a single test. That is, adverse impact is the end result of the magnitude of group differences, selection ratios, use of different selection tools in combination, and the manner in which scores are combined and utilized. Sackett and Roth (1996) used a series of Monte Carlo simulations to investigate the effects of multistage selection strategies on minority hiring. The important features of selection systems that contributed to the level of minority hiring included subgroup differences on the predictors, intercorrelations among the predictors in the selection system, the overall selection ratio, and the selection strategy used (i.e., top-down, hurdle, etc.).

For cognitive ability tests, group differences have been examined in dozens of primary studies and have been meta-analytically summarized. In these studies, the measure of group differences is typically Cohen's *d*, which expresses the differences between the means of two groups in terms of standard deviation units. In meta-analyzing these effect sizes, *d* values from individual studies are pooled and averaged to obtain an overall effect size that reflects the magnitude of group differences in the population at large. Corrections for unreliability in cognitive ability measures are typically not applied, because selection decisions are based on observed scores.[3] In general, *d* values of .80 or greater are considered large effects, those around .50 are moderate, and those below .20 are small (Cohen, 1977). (From a theoretical perspective, *d* values under .20 are often trivial; however, under extreme conditions, such as when the majority group selection ratio is under 1%, even small differences in the .10 to .20 range can lead to violation of the four-fifths rule and thus constitute adverse impact.)

In the I-O psychology literature, it is widely believed and reported that sex differences in cognitive ability are nonexistent (e.g., see Table 1 in Ployhart & Holtz, 2008). Table 11.3 offers a more precise and detailed view in summarizing sex differences on cognitive variables. Differences in verbal and mathematical abilities are negligible. Women score moderately higher than men on one particular verbal ability marker—speech production. Largest sex differences are found on visual-spatial measures such as mental rotation and spatial perception (meta-analytic *d* values in favor of men are in the .40 to .70 range) as well as figural reasoning and technical aptitude (Irwing & Lynn, 2005; Lynn & Irwing, 2004; Schmidt, 2011). Thus, given selection ratios of 50% or lower for men, cognitive tests with visual-spatial items or technical aptitude questions (e.g., mechanical comprehension, electronics information) can result in adverse impact against women. However, when general mental ability scores (extracted from a battery of different tests) are considered, sex differences have been shown to be either negligible (Colom, Juan-Espinosa, Abad, & García, 2000; Deary, Irwing, Der, & Bates, 2007) or favoring females to a small degree (Keith, Reynolds, Patel, & Ridley, 2008). Thus, organizations concerned with gender diversity would be better off including general mental ability tests over tests of those specific abilities in their assessment systems.

Underrepresentation of women in science, technology, engineering, and math (STEM) fields has spurred inquiries into whether men's and women's variances are comparable on cognitive tests. If men are more variable than women, groups of individuals selected may reflect greater proportions of men at the high end of the ability distribution, even if there are no mean subgroup differences. Hyde's (2014) summary of the literature indicates male-to-female variance ratios in the 1.03–1.16 range for verbal abilities, 1.05–1.20 range for mathematics abilities, and 1.27 for spatial ability. The 27% greater variability among men on spatial abilities may partly explain lower female high achievement and accomplishments in STEM fields (Lubinski, 2010).

More so than sex differences in cognitive ability, race and ethnic group differences have consumed attention in employee selection research and practice, especially in the North American context (Hough, Oswald, & Ployhart, 2001). Table 11.4 summarizes race and ethnic group differences on cognitive ability based on the largest meta-analysis of the employment literature (Roth, Bevier, Bobko, Switzer, & Tyler, 2001). On average, Blacks score 1.00 and Hispanics .83

TABLE 11.3

Meta-Analyses of Sex Differences on Cognitive Ability Measures

Cognitive Variable	*Study*	*k*	*d*
Vocabulary	Hyde & Linn (1988)	40	−.02
Reading comprehension	Hyde & Linn (1988)	18	−.03
Speech production	Hyde & Linn (1988)	12	−.33
Mathematics computation	Hyde, Fennema, & Lamon (1990)	45	−.14
Mathematics concepts	Hyde, Fennema, & Lamon (1990)	41	−.03
Mathematics problem solving	Hyde, Fennema, & Lamon (1990)	48	.08
Spatial perception	Linn & Petersen (1985)	62	.44
Spatial perception	Voyer, Voyer, & Bryden (1995)	92	.44
Mental rotation	Linn & Petersen (1985)	29	.73
Mental rotation	Voyer, Voyer, & Bryden (1995)	78	.56
Mental rotation	Maeda & Yoon (2012)	70	.57
Mental rotation—untimed tests	Voyer (2011)	23	.51
Mental rotation—short time limits	Voyer (2011)	7	1.03
Mental rotation—long time limits	Voyer (2011)	6	.85
Spatial visualization	Linn & Petersen (1985)	81	.13
Spatial visualization	Voyer, Voyer, & Bryden (1995)	116	.19
Figural reasoning (matrices)	Lynn & Irwing (2004)	10	.30
Figural reasoning (matrices—standard)	Irwing & Lynn (2005)	10	.10
Figural reasoning (matrices—advanced)	Irwing & Lynn (2005)	11	.20

k = number of studies summarized in meta-analysis; *d* = standardized group mean score difference. Positive effect sizes indicate males scoring higher on average.

TABLE 11.4

Race and Ethnic Group Mean Score Differences in General Mental Ability Among Job Applicants

Group Comparison	*Setting*	*Job Complexity*	*N*	*k*	*d*
White–Black	Industrial	Across complexity levels	375,307	11	1.00
	Industrial (within-job studies)	Low	125,654	64	.86
		Moderate	31,990	18	.72
		High	4,884	2	.63
	Military	Across complexity levels	245,036	1	1.46
White–Hispanic	Industrial	Across complexity levels	313,635	14	.83
	Military	Across complexity levels	221,233	1	.85

k = number of studies summarized in meta-analysis; *d* = standardized group mean score difference; *N* = total sample size. Positive effect sizes indicate Whites scoring higher on average.
Source: Data from Tables 2, 4, and 7 of Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P., *Personnel Psychology, 54*, 297–330, 2001.

standard deviation units lower than Whites on general mental ability (GMA) measures used in employee selection. Group differences on measures used in military settings are somewhat larger, especially for the White–Black comparison. One explanation for this finding could be the greater heterogeneity among military job applicants. Cognitive ability differences between Black and White applicants to high-complexity jobs are smaller than among applicants to lower-complexity jobs, most likely because of severe self-selection as well as higher minimum requirements with regard to educational credentials. Among applicants to medium- and low-complexity jobs, White–Black and White–Hispanic differences in cognitive ability tests are large and almost certain to result in adverse impact if cognitive ability were the only predictor used in employee selection. This finding is at the root of the validity-diversity dilemma that most U.S. organizations face today (Kehoe, 2008; Kravitz, 2008; Ployhart & Holz, 2008; Potosky, Bobko, & Roth, 2008; Sackett, De Corte, & Lievens, 2008). The situation is slightly better among applicants to high-complexity jobs, in which group mean-score differences in cognitive ability are only moderate ($d = .63$), and thus carry slightly less severe implications for adverse impact.

Data on Asian American–White and Native American–White cognitive ability differences among job applicants are scant. Ability profiles and subgroup differences for Asian Americans and Native Americans remain mostly uninvestigated, especially when the job applied to is held constant (i.e., within-job examinations). The broader psychological literature indicates slightly higher scores among Asian Americans compared with Whites (Gottfredson, 1997), but again, systematic data on job applicants are scarce. The response categories used for demographic data collection in psychological research often subsume individuals from very heterogeneous race and ethnic backgrounds in a single category, which complicates comparisons, especially with regard to the White–Asian comparisons. (As an illustration: the most recent U.S. census lists 14 national and ethnic categories of interest as well as one residual "other" category; 7 of the 15 categories represent Asian groups, and in total, 57 multi-race combinations could possibly be endorsed; "The Asian Population: 2010," U.S. Census Bureau, 2012.) The educational literature reports sizable lower scores among Native Americans when compared with Whites (Humphreys, 1988). We were able to locate only one study that compared Native North Americans and Whites in a job context (Vanderpool & Catano, 2008). In this study, individuals from Canadian Aboriginal Peoples scored much lower on verbal ability tests than on nonverbal tests.

It is important to stress that although race and ethnicity are protected categories in the United States, the characteristics that define disadvantaged groups elsewhere are diverse (cf. Myors et al., 2008). Furthermore, constructs such as race and ethnicity are also often confounded with national origin or immigrant/refugee status. Cognitive ability test scores of disadvantaged groups around the globe remain largely unstudied in employee selection settings, although exceptions can be found in a handful of countries where race, ethnicity, or immigrant status have been examined (e.g., Australia, Israel, New Zealand, South Africa, Taiwan, Turkey, and the Netherlands). This research appears to point to consistently lower scores of disadvantaged groups (e.g., Aborigines in Australia, Canada, New Zealand, and Taiwan; Blacks in South Africa; immigrants in the Netherlands and Sweden; Sackett & Shen, 2008, Salgado, in press). Ongoing mass refugee movements, resulting in more than 60 million refugees worldwide (United Nations High Commissioner for Refugees, December 2015), will offer I-O psychologists and societies around the world both opportunities and challenges. If cognitively oriented assessments are utilized in assessing and placing refugees into jobs, potential subgroup differences must be attended to.

Large-scale international comparisons using the same cognitive ability test in employment settings are rare, if not nonexistent. The first two authors of this chapter were involved in content development for a computer adaptive figural reasoning test for use in employee selection around the globe (Dilchert & Ones, 2007). The data from hundreds of thousands of applicants allow us an extraordinary global look at group differences on the same, nonverbal reasoning measure. Among the nearly 60,000 U.S.-based applicants who completed the tests in the first few months after implementation, group differences were in the expected direction, with all minority groups except Asian Americans scoring lower than Whites. However, typically observed group differences were reduced. When analyzed on the country level, the 10 countries in which job applicants scored highest on average were Southeast Asian (4) and European (6). An analysis of the data by cultural clusters revealed Confucian Asia and southern and northern Europe scoring higher than other cultural regions. Regardless of the underlying mechanisms, observed differences between applicants from different cultural regions applying to the same organization present a challenge and an opportunity to actively shape their workforce on the basis of diversity and talent goals.

One area of group differences that has received little attention in the adverse impact literature is that of cognitive ability differences between younger and older adults. One large-scale examination of cognitive abilities across the working life span (Avolio & Waldman, 1994) offers some insights into the magnitudes of age-related declines in cognitive ability. Avolio and Waldman (1994) reported mean scores of 25,140 White, Black, and Hispanic job applicants who had taken the General Aptitude Test Battery (GATB) of the U.S. Employment Service and broke these scores down by age groups. When computing *d* values based on these data, one notices that age group differences in cognitive ability (both general and specific) start as early as age 35, but they become particularly notable for the 45–55 and 55–65 age groups (differences to the 20–34 year comparison groups range from approximately .80 to 1.5 standard deviation units). Verbal ability, however, shows the smallest declines across older age groups. Investigations of age differences among job applicant samples are still rare in the scholarly literature. One recent exception is the work by Klein, Dilchert, Ones, and Dages (2015), who reported age-differences among job applicants to managerial and executive positions, and did so separately for different ability tests. Their results, which also generalized in two representative, longitudinal U.S. general population samples, showed that certain crystallized verbal abilities actually increased over individuals' working lives, but that declines in general mental ability, as well as (most drastically) inductive reasoning, are notable as early as the early forties. Table 11.5 summarizes these findings. When cognitive ability measures are used in employee selection, younger applicants generally stand to get selected at greater rates than older applicants. The disparity in selection ratios can be particularly severe if applicant pools include individuals from the entire age spectrum of adults. However, we now know that the choice of ability test matters. Group differences (and thus the threat of adverse impact) are less severe on general mental ability tests/scores compared to fluid ability or inductive reasoning. Moreover, some crystallized verbal abilities might even present an advantage for older adults. But it is important to recall that crystallized verbal ability differences are largest in race/ethnic group comparisons (see above). Organizations that are concerned with diversity should closely examine both the demographic makeup of their applicant pools and the new knowledge acquisition, adaptability, and verbal ability requirements of specific jobs to strategically address age and race/ethnic diversity through proper test choice.

Evidence from the individual differences literature suggests that rates of cognitive decline are slower for those who have higher initial baseline ability (Deary, MacLennan, & Starr, 1998), higher levels of education (Deary et al., 1998; Rhodes, 2004), and those employed in complex or enriched jobs that presumably use their cognitive abilities to a greater extent (Schooler, Mulatu, & Oates, 1999). In the future, the aging workforces of most industrialized countries will certainly necessitate greater attention to the consequences of cognitive ability test use for workforce diversity with regard to age.

Validity Differences and Predictive Bias

Thus far, we have discussed only group mean score differences on cognitive ability tests. Another salient issue is that of differential validity and differential prediction. Differential validity refers to differences in criterion-related validity coefficients of various subgroups. It indicates the degree to which the pre-employment test similarly/differentially relates to a given criterion. While differential validity compares the magnitudes of criterion-related validities between groups of interest, differential prediction simultaneously compares slopes and intercepts of regression lines for such groups. A healthy body of literature in employee selection has led to the conclusion that there is no predictive bias against Blacks in the United States (Rotundo & Sackett, 1999; Schmidt, 1988).

Hunter, Schmidt, and Hunter (1979) and Schmidt, Pearlman, and Hunter (1980) have quantitatively summarized dozens of validation studies using the GATB with Blacks and Hispanics, respectively. Hunter et al.'s (1979) analysis demonstrated that, on average, validities for Whites were .01 correlational points higher than those for Blacks in predicting objective performance criteria and .04 correlational points higher for predicting subjective ratings of job performance (k = 866 non-independent validity pairs).

TABLE 11.5

Age Differences in Cognitive Ability

Cognitive Variable	*Age Group Comparison*	*N*	*d*
U.S. job applicants who completed the General Aptitude Test Battery[a]			
	Reference Group: 20–34 Years	13,746	–
GMA	35–44	4,305	0.33
	45–54	2,825	0.55
	55–65	1,161	0.80
Verbal ability	35–44	4,305	0.26
	45–54	2,825	0.35
	55–65	1,161	0.49
Numerical ability	35–44	4,305	0.36
	45–54	2,825	0.59
	55–65	1,161	0.71
Spatial ability	35–44	4,305	0.30
	45–54	2,825	0.55
	55–65	1,161	0.88
Form perception	35–44	4,305	0.56
	45–54	2,825	1.04
	55–65	1,161	1.53
Clerical ability	35–44	4,305	0.40
	45–54	2,825	0.69
	55–65	1,161	0.90
Job applicants to executive positions who completed multiple cognitive ability tests[b]			
	Reference Group: 20–34 Years	662	–
GMA	35–44	1,167	0.12
	45–54	1,098	0.23
	55–64	371	0.32
Verbal ability	35–44	1,167	-0.36
	45–54	1,098	-0.49
	55–64	371	-0.76
Figural reasoning	35–44	1,167	0.09
	45–54	1,098	0.20
	55–64	371	0.36
Inductive reasoning	35–44	1,167	0.51
	45–54	1,098	0.80
	55–64	371	1.03

N = sample size; *d* = standardized group mean-score difference. Positive effect sizes indicate younger individuals scoring higher on average; the reference group for computation of *d* values was 20–34 years of age for all effect sizes.

[a] Based on data presented in Table 3 of Avolio, B. J., & Waldman, D. A., *Psychology and Aging, 9*, 430–442, 1994. Means and standard deviations for the reference group were obtained by sample-size weighting means and pooling standard deviations for 20- to 24-year-old and 25- to 34-year-old age groups across race and ethnic groups.

[b] Based on data presented in Table 2 of Klein, R. M., Dilchert, S., Ones, D. S., & Dages, K. D., *Journal of Applied Psychology, 5*, 1497–1510, 2015. Group sample sizes were combined across subgroups to enable age-group comparisons consistent with those computed based on Avolio & Waldman data. *d* values were computed by sample-size weighting effects across the subgroups that were combined for each row.

In the 2010 edition of this Handbook, we stressed the need for an update of the existing literature on differential validity. Berry, Clark, and McClure (2011) presented an updated meta-analysis, incorporating previous data, reporting that in employment settings observed validities for Whites and Blacks differed by .03 (k = 143 independent studies incorporating data from 20,399 Whites and 10,350 Blacks). Observed validity differences reported in 93 military studies were drastic: White validities were double Black validities. Roth et al. (2014) suggested differential range restriction as a factor that clouds observed validity comparisons. In response, Berry, Cullen, and Meyer (2014) produced a new set of meta-analytic validity estimates corrected for differential range restriction. In civilian settings, range restriction values (u values) were .89 for Whites and .85 for Blacks. Differences in range-restricted validity remained at .03 correlational points lower for Blacks. In military studies, much of the differential validity was concluded to be due to differential range restriction, as the range restriction corrected validity difference between Whites and Blacks shrank to .07 correlational points (White validity higher). Using up-to-date meta-analytic operational validity estimates, Berry and Zhao (2015) concluded that there is "strong evidence that cognitive ability tests generally overpredict job performance of African Americans" (p. 162).

Research on Hispanic Americans is meager. Across 1,128 pairs of validity coefficients from 19 studies, Schmidt et al. (1980) showed White validities on average to be .02 correlational points higher than Hispanic validities. Berry et al. (2014) reported that across 35 studies in civilian employment settings, appropriately range-restriction-corrected validity differences were .02 (Whites higher). Differential validity analyses were available only in educational settings, indicating .02 operational validity points higher for Whites. Differential validity of cognitive ability tests in organizational settings has not been reported for Asian and Native Americans in the peer-reviewed literature.

Rothstein and McDaniel (1992) reported an examination of differential validity by sex for cognitive ability tests. Using 59 pairs of male and female correlations (N = 5,517 and 9,428, respectively), they found observed validities to be on average .03 correlational points higher for women (validities corrected for range restriction and unreliability in the criteria were .05 correlational points higher). The higher validity for women was more marked in lower-complexity jobs and female-dominated occupations. In male-dominated occupations, the validity was higher for predicting performance among men. For the prediction of academic success, the reverse (but weak) pattern was established in a recent meta-analysis (Fischer, Schult, & Hell, 2013). We were unable to locate differential validity investigations for older versus younger adults. Future research should examine differential validity for hitherto unexamined groups in employment settings (Asians, Native Americans, older adults). Some exceptions not withstanding (e.g., the Netherlands and South Africa), studies from other parts of the world (as well as those for other minority groups) are sparse and need to be conducted as well.

We would like to stress that continued research on differential validity would be valuable. Labor force participation and occupational distributions of women, Blacks, Hispanics, and a multitude of racial, ethnic, and religious groups are much different today than even 10–20 years ago. Changes in the nature of many jobs (e.g., greater complexity, greater technological demands) as well as changes in the social milieu in many organizations (e.g., emergence of workforce diversity as a core value, mass immigrations across the globe) may manifest themselves in cognitive ability–criteria relations. Research must also examine whether differential validity is found for criteria other than overall job performance. In our opinion, studies on organizational citizenship behaviors, task performance, and leadership criteria may constitute priorities. The only study that examined Black–White differential validity of a cognitive ability test for predicting nontraditional performance criteria investigated incidents of detected counterproductive behaviors (interpersonal and those targeted at the organization) and found no evidence of differential validity (Dilchert et al., 2007). However, replications of these results, as well as investigations among other minority groups, are certainly warranted.

FUTURE CHALLENGES FOR RESEARCH AND PRACTICE

In this chapter, we have identified specific areas in need of additional research attention as well as some challenges for the use of cognitive ability tests in applied settings. The high validity of cognitive measures makes them attractive for use in employee selection. Their ability to enhance productivity and offer substantial economic utility to organizations is indisputable. However, many applied psychologists are concerned, and understandably so, that various groups (e.g., Blacks, Hispanics/Latinos, other disadvantaged ethnic groups, and older applicants) on average score lower than the majority applicants, often resulting in differential selection ratios for different groups. Our literature is filled with suggestions on ways to reduce the likelihood for adverse impact. Thoughtful description and evaluation of various proposed alternatives is not possible in this short chapter but is available in various papers (e.g., Campbell, 1996; De Corte, Lievens, & Sackett, 2007, 2010; Hough et al., 2001; Hunter & Schmidt, 1982; Ployhart & Holtz, 2008; Potosky, Bobko, & Roth, 2005; Sackett, Schmitt, Ellingson, & Kabin, 2001; Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). Frankly, we believe that structural and procedural proposals to reduce adverse impact are stopgap measures that are not sufficient for dealing with profound group differences observed in occupational settings. Although the exact definition of what constitutes protected classes may differ, societies around the world are now facing similar issues—this fact has been corroborated by the adoption of antidiscrimination Directives 2000/43/EC and 2000/78/EC in the European Community (country-specific laws passed as a result still differ considerably with regard to prohibited grounds and protected classes; while some countries expanded the Directives by including color, national origin, or language, others are less broad in their protection). It will require the collective wisdom of scientists across disciplines to evaluate whether some group differences on individual differences traits can be reduced, and if so, how. In the meantime, I-O psychologists need to face the challenges that these group differences pose in applied settings. To this end, the responsibility is equally distributed among (a) scientists, who need to address the areas of concern summarized above; (b) test publishers, who need to continuously collect and make available data regarding group differences and predictive fairness of their tests; and (c) individual practitioners, who need to educate themselves on the past and current research as well as its implications for their specific purpose.

Another, somewhat easier challenge is that of enhancing the acceptability of cognitive ability measures among applicants to high-complexity jobs. As Lubinski (2004) pointed out, cognitive ability can be assessed in all shapes and forms: "Variegated conglomerations of information and problem-solving content, not necessarily tied to an educational program, which may involve fresh as well as old learning (acquired in or out of school), may be used to assess general intelligence" (p. 98). However, it is our opinion that when new formats and approaches are used to address issues of applicant reactions and face validity, intellectual honesty still mandates an acknowledgment of the construct being measured. The proliferation of "new" abilities and claims that such abilities are independent of traditional intelligence are insincere and harmful to the professional reputation of our field. "Gamification" of selection processes and tools might result in higher applicant engagement, but thorough assessment development and professional principles of measurement should not be abandoned in this pursuit.

We have also observed that sometimes preconceived notions of cognitive test acceptability can cloud our judgment. Our work with nonverbal figural reasoning tests, arguably an item type that on the surface does not appear extraordinarily related to most real-world tasks, yielded some surprising findings. Data show that such items, especially when compared to those with verbal content, are received very positively by applicants. Although contextualization is certainly a viable method of achieving face validity, items need not always be contextualized to invoke positive applicant reactions.

EPILOGUE

This chapter aimed to offer a broad and forthright overview of cognitive ability tests and their use in employee selection. Other excellent overviews of the topic may be found in Drasgow

(2003, especially with regard to structural issues); Ree, Carretta, and Steindl (2001, especially with regard to broader life correlates); Ones, Viswesvaran, and Dilchert (2004, 2005; especially with regard to validity for learning criteria); Ones et al. (2012), Dilchert (in press), and Salgado (in press, especially with regard to a criterion-related validity in organizational settings). Debates and exchanges over the use of cognitive ability tests in selection settings can also be found in special issues of *Human Performance* (Viswesvaran & Ones, 2002), *Human Resource Management* (Vol. 25[1]), and *Industrial and Organizational Psychology* (Vol. 12[5]).

Cognitive ability is the capacity to learn, solve problems, and adapt to environments. Abstract thinking and logic reasoning determine success in various life domains by allowing us to not only rely on skills acquired through past experience, but also to react to novel situations through knowledge and insights acquired in mental simulations. Cognitive ability continues to be the single best determinant of work performance. We believe that the benefits associated with cognitive ability test use in employee selection far outweigh potential concerns. More importantly, the changing nature of work in most developed economies (increasing job complexity, fewer traditional employment relationships, increasing job switching) means that cognitive ability should be an increasingly important human capital variable. Advances in technology, such as the ubiquitous availability of mobile devices as well as increasing Internet access even in remote regions, provide immense opportunities to improve both the science and practice of employee assessment using cognitive ability tests. Although few fundamental things might have changed in the last 30 years of cognitive ability assessment, the near future promises exciting developments.

NOTES

1. For the 100th anniversary of that article's publication, the *Journal of Personality and Social Psychology* published a special section attesting to the impact of cognitive ability on a multitude of life domains (Deary, Whiteman, Starr, Gottfredson, 2004a; Kuncel, Hezlett, & Ones, 2004; Lubinski, 2004; Plomin & Spinath, 2004; Schmidt & Hunter, 2004; Whalley, & Fox, 2004).
2. In fact, meta-analysis has changed the nature of epistemological inquiry in all sciences. Thirty years after its inception, 40,000 peer-reviewed publications have used or discussed meta-analytic methods, garnering hundreds of thousands citations (Christensen, Selzer, Beatty, & Ones, 2009).
3. However, corrections for attenuation due to range restriction and unreliability in the criterion are advisable when comparing results across studies differing in their levels of range restriction and unreliability.

REFERENCES

Alonso, A., Viswesvaran, C., & Sanchez, J. I. (2008). The mediating effects of task and contextual performance. In J. Deller (Ed.), *Research contributions to personality at work* (pp. 3–17). Munich, Germany: Rainer Hampp.

Arneson, J. J. (2007). *An examination of the linearity of ability—Performance relationships among high scoring applicants*. Unpublished doctoral dissertation, Minneapolis, MN: University of Minnesota.

Arthur, W., Jr., Glaze, R. M., Villado, A. J., & Taylor, J. E. (2010). The magnitude and extent of cheating and response distortion effects on unproctored Internet-based tests of cognitive ability and personality. *International Journal of Selection and Assessment, 18*(1), 1–16. http://dx.doi.org/10.1111/j.1468–2389.2010.00476.x

Avolio, B. J., & Waldman, D. A. (1994). Variations in cognitive, perceptual, and psychomotor abilities across the working life span: Examining the effects of race, sex, experience, education, and occupational type. *Psychology and Aging, 9*, 430–442.

Barrett, G. V., Phillips, J. S., & Alexander, R. A. (1981). Concurrent and predictive validity designs: A critical reanalysis. *Journal of Applied Psychology, 66*, 1–6.

Barros, E., Kausel, E. E., Cuadra, F., & Díaz, D. A. (2014). Using general mental ability and personality traits to predict job performance in three Chilean organizations. *International Journal of Selection and Assessment, 22*, 432–438.

Berry, C. M., Gruys, M. L., & Sackett, P. R. (2006). Educational attainment as a proxy for cognitive ability in selection: Effects on levels of cognitive ability and adverse impact. *Journal of Applied Psychology, 91*, 696–705.

Berry, C. M., Sackett, P. R., & Landers, R. N. (2007). Revisiting interview-cognitive ability relationships: Attending to specific range restriction mechanisms in meta-analysis. *Personnel Psychology, 60*, 837–874.

Berry, C. M., & Zhao, P. (2015). Addressing criticisms of existing predictive bias research: Cognitive ability test scores still overpredict African Americans' job performance. *Journal of Applied Psychology, 100*(1), 162–179. http://dx.doi.org/10.1037/a0037615

Bertolino, M., & Steiner, D. D. (2007). Fairness reactions to selection methods: An Italian study. *International Journal of Selection and Assessment, 15*, 197–205.

Burke, E. (April 2008). Remarks. In N. Tippins (Chair), *Internet testing: Current issues, research, solutions, guidelines, and concerns*. Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Campbell, J. P. (1990). The role of theory in industrial and organizational psychology. In M. D. Dunnette & L. M. Hough (Eds.), *Handbook of industrial and organizational psychology* (Vol. 1, pp. 39–73). Palo Alto, CA: Consulting Psychologists Press.

Campbell, J. P. (1996). Group differences and personnel decisions: Validity, fairness, and affirmative action. *Journal of Vocational Behavior, 49*, 122–158.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. New York, NY: Cambridge University Press.

Cattell, R. B. (1971). *Abilities: Their structure, growth, and action*. Oxford, England: Houghton Mifflin.

Chan, D., & Schmitt, N. (1997). Video-based versus paper-and-pencil method of assessment in situational judgment tests: Subgroup differences in test performance and face validity perceptions. *Journal of Applied Psychology, 82*, 143–159.

Chan, D., & Schmitt, N. (2004). An agenda for future research on applicant reactions to selection procedures: A construct-oriented approach. *International Journal of Selection and Assessment, 12*, 9–23.

Chan, D., Schmitt, N., Jennings, D., Clause, C., & Delbridge, K. (1998). Applicant perceptions of test fairness: Integrating justice and self-serving bias perspectives. *International Journal of Selection and Assessment, 6*, 232–239.

Chartered Institute of Personnel and Development. (2007). *2007 recruitment, retention, and turnover survey*. London, UK: Chartered Institute of Personnel and Development.

Christiansen, F., Seltzer, B. K., Beatty, A., & Ones, D. S. (April 2009). *Thirty years of meta-analysis: Assessing its impact on the sciences*. Poster presented at the annual conference of the Society for Industrial and Organizational Psychology, New Orleans, Louisiana.

Cohen, J. (1977). *Statistical power analysis for the behavioral sciences*. San Diego, CA: Academic Press.

Collins, J. M., Schmidt, F. L., Sanchez-Ku, M., Thomas, L., McDaniel, M. A., & Le, H. (2003). Can basic individual differences shed light on the construct meaning of assessment center evaluations? *International Journal of Selection and Assessment, 11*, 17–29.

Coward, W., & Sackett, P. R. (1990). Linearity of ability-performance relationships: A reconfirmation. *Journal of Applied Psychology, 75*, 297–300.

Deary, I. J., MacLennan, W. J., & Starr, J. M. (1998). Is age kinder to the initially more able?: Differential ageing of a verbal ability in the healthy old people in Edinburgh study. *Intelligence, 26*, 357–375.

Deary, I. J., Whiteman, M. C., Starr, J. M., Whalley, L. J., & Fox, H. C. (2004). The impact of childhood intelligence on later life: Following up the Scottish mental surveys of 1932 and 1947. *Journal of Personality and Social Psychology, 86*, 130–147.

De Corte, W., Lievens, F., & Sackett, P. R. (2007). Combining predictors to achieve optimal trade-offs between selection quality and adverse impact. *Journal of Applied Psychology, 92*, 1380–1393.

De Corte, W., Sackett, P. R., & Lievens, F. (2010). Selecting predictor subsets: Considering validity and adverse impact. *International Journal of Selection and Assessment, 18*(3), 260–270. http://dx.doi.org/10.1111/j.1468–2389.2010.00509.x

Detterman, D. K., & Daniel, M. H. (1989). Correlations of mental tests with each other and with cognitive variables are highest for low IQ groups. *Intelligence, 13*, 349–359.

Dilchert, S. (In press). Cognitive ability in occupational settings. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran, *International handbook of industrial, work, and organizational psychology*. Thousand Oaks, CA: Sage.

Dilchert, S., & Ones, D. S. (April 2007). Influence of figural reasoning item characteristics on group mean-score differences. In A. S. Boyce & R. E. Gibby (Chairs.), *Global cognitive ability testing: Psychometric issues and applicant reactions*. Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, New York, NY.

Dilchert, S., & Ones, D. S. (2009). Assessment center dimensions: Individual differences correlates and meta-analytic incremental validity. *International Journal of Selection and Assessment*, *17*, 254–270.

Dilchert, S., Ones, D. S., Davis, R. D., & Rostow, C. D. (2007). Cognitive ability predicts objectively measured counterproductive work behaviors. *Journal of Applied Psychology, 92*, 616–627.

Drasgow, F. (2003). Intelligence and the workplace. In W. C. Borman, D. R. Ilgen, & R. J. Klimoski (Eds.), *Handbook of psychology: Industrial and organizational psychology* (Vol. 12, pp. 107–130). Hoboken, NJ: John Wiley & Sons.

Foster, D. (April 2008). Remarks. In N. Tippins (Chair.), *Internet testing: Current issues, research, solutions, guidelines, and concerns*. Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Ghiselli, E. E., & Kahneman, D. (1962). Validity and non-linear heteroscedastic models. *Personnel Psychology, 15*, 1–11.

Gibby, R. E. (April 2008). Online and unsupervised adaptive cognitive ability testing: Lessons learned. In R. E. Gibby & R. A. McCloy (Chairs.), *Benefits and challenges of online and unsupervised adaptive testing*. Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Gibby, R. E., & Boyce, A. S. (April 2007). *Global cognitive ability testing: Psychometric issues and applicant reactions*. Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, New York, NY.

Gottfredson, L. S. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history and bibliography. *Intelligence, 24*, 13–23.

Gottfredson, L. S. (2002). Where and why g matters: Not a mystery. *Human Performance, 15*, 25–46.

Gottfredson, L. S. (2004a). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology, 86*, 174–199.

Gottfredson, L. S. (2004b). Life, death, and intelligence. *Journal of Cognitive Education and Psychology, 4*, 23–46.

Gonzalez-Mulé, E., Mount, M. K., & Oh, I-S. (2014). A meta-analysis of the relationship between general mental ability and nontask performance. *Journal of Applied Psychology, 99*(6), 1222–1243. http://dx.doi.org/10.1037/a0037547

Gustafsson, J-E. (2002). Measurement from a hierarchical point of view. In H. I. Braun, D. N. Jackson, & D. E. Wiley (Eds.), *The role of constructs in psychological and educational measurement* (pp. 73–95). Mahwah, NJ: Lawrence Erlbaum.

Hausknecht, J. P., Day, D. V., & Thomas, S. C. (2004). Applicant reactions to selection procedures: An updated model and meta-analysis. *Personnel Psychology, 57*, 639–683.

Hawk, J. (1970). Linearity of criterion-GATB aptitude relationships. *Measurement and Evaluation in Guidance, 2*, 249–251.

Hense, R., Golden, J. H., & Burnett, J. (2009). Making the case for unproctored internet testing: Do the rewards outweigh the risks? *Industrial and Organizational Psychology, 2*, 20–23.

Hoffman, B. J., Monahan, E., Lance, C. E., & Sutton, A. (2015). A meta-analysis of the content, construct, and criterion-related validity of assessment center exercises. *Journal of Applied Psychology, 100*, 1143–1168.

Hough, L. M., Oswald, F. L., & Ployhart, R. E. (2001). Determinants, detection and amelioration of adverse impact in personnel selection procedures: Issues, evidence and lessons learned. *International Journal of Selection and Assessment, 9*, 152–194.

Hull, C. L. (1928). *Aptitude testing*. Yonkers-on-Hudson, NY: World Book.

Hülsheger, U. R., Maier, G. W., & Stumpp, T. (2007). Validity of general mental ability for the prediction of job performance and training success in Germany: A meta-analysis. *International Journal of Selection and Assessment, 15*, 3–18.

Humphreys, L. G. (1988). Trends in levels of academic achievement of Blacks and other minorities. *Intelligence, 12*, 231–260.

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

Hunter, J. E., & Schmidt, F. L. (1982). Ability tests: Economic benefits versus the issue of test fairness. *Industrial Relations, 21*, 122–158.

Hunter, J. E., Schmidt, F. L., & Hunter, R. (1979). Differential validity of employment tests by race: A comprehensive review and analysis. *Psychological Bulletin, 86*, 721–735.

Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist, 60*, 581–592.

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65*, 373–398.

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin, 107*, 139–155.

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104*, 53–69.

Johnson, W., & Bouchard, T. J. (2005). The structure of human intelligence: It is verbal, perceptual, and image rotation (VPR), not fluid and crystallized. *Intelligence, 33*, 393–416.

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2007). Replication of the hierarchical visual-perceptual-image rotation model in de Wolff and Buiten's (1963) battery of 46 tests of mental ability. *Intelligence, 35*, 69–81.

Johnson, W., te Nijenhuis, J., & Bouchard, T. J. (2008). Still just 1 g: Consistent results from five test batteries. *Intelligence, 36*, 81–95.

Kane, H. D., Oakland, T. D., & Brand, C. R. (2006). Differentiation at higher levels of cognitive ability: Evidence from the United States. *Journal of Genetic Psychology, 167*, 327–341.

Kehoe, J. F. (2008). Commentary on Pareto-optimality as a rationale for adverse impact reduction: What would organizations do? *International Journal of Selection and Assessment, 16*, 195–200.

Kluger, A. N., & Rothstein, H. R. (1993). The influence of selection test type on applicant reactions to employment testing. *Journal of Business and Psychology, 8*, 3–25.

Krantowitz, T. M. (2014). *2014 Global assessment trends report*. Surrey, UK: CEB-SHL.

Kravitz, D. A. (2008). The diversity-validity dilemma: Beyond selection—The role of affirmative action. *Personnel Psychology, 61*, 173–193.

Kriek, H., & Dowdeswell, K. (2010). Adverse impact in South Africa. In J. L. Outz (Ed.), *Adverse impact: Implications for organizational staffing and high stakes selection* (pp. 375–399). New York. Routledge.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology, 86*, 148–161.

Kuncel, N. R., & Sackett. P. R. (2013). Resolving the assessment center construct validity problem (as we know it). *Journal of Applied Psychology, 99*, 38–47.

Lee, S. (2005). *Cross-cultural validity of personality for predicting job performance of Korean engineers*. Unpublished doctoral dissertation, Columbus: Ohio State University.

Lent, R. H., Aurbach, H. A., & Levin, L. S. (1971). Predictors, criteria, and significant results. *Personnel Psychology, 24*, 519–533.

Lievens, F., & Burke, E. (2011). Dealing with the threats inherent in unproctored Internet testing of cognitive ability: Results from a large-scale operational test program. *Journal of Occupational and Organizational Psychology, 84*(4), 817–824. http://dx.doi.org/10.1348/096317910X522672

Lievens, F., De Corte, W., & Westerveld, L. (2015). Understanding the building blocks of selection procedures: Effects of response fidelity on performance and validity. *Journal of Management, 41*, 1604–1627.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex-differences in spatial ability: A meta-analysis. *Child Development, 56*, 1479–1498.

Loehlin, J. C. (2000). Group differences in intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence* (pp. 176–193). New York, NY: Cambridge University Press.

Lubinski, D. (2004). Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General Intelligence', Objectively Determined and Measured". *Journal of Personality and Social Psychology, 86*, 96–111.

Lubinski, D. (2010). Spatial ability and STEM: A sleeping giant for talent identification and development. *Personality and Individual Differences, 49*(4), 344–351. http://dx.doi.org/10.1016/j.paid.2010.03.022

Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence, 32*, 481–498.

McCloy, R. A., Campbell, J. P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology, 79*, 493–505.

McDaniel, M. A., Hartman, N. S., Whetzel, D. L., & Grubb, W. (2007). Situational judgment tests, response instructions, and validity: A meta-analysis. *Personnel Psychology, 60*, 63–91.

McGrew, K. (2009). Editorial: CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1–10.

Mead, A. D., & Drasgow, F. (1993). Equivalence of computerized and paper-and-pencil cognitive ability tests: A meta-analysis. *Psychological Bulletin, 114*, 449–458.

Meriac, J. P., Hoffman, B. J., & Woehr. D. J. (2014). A conceptual and empirical review of the structure of assessment center dimensions. *Journal of Management, 40*, 1269–1296.

Muller, J., & Schepers, J. (2003). The predictive validity of the selection battery used for junior leader training within the South African National Defense Force. *South African Journal of Industrial Psychology, 29*, 87–98.

Myors, B., Lievens, F., Schollaert, E., Van Hoye, G., Cronshaw, S. F., Mladinic, A., et al. (2008). International perspectives on the legal environment for selection. *Industrial and Organizational Psychology, 1*, 206–246.

Nathan, B. R., & Alexander, R. A. (1988). A comparison of criteria for test validation: A meta-analytic investigation. *Personnel Psychology, 41*, 517–535.

Nikolaou, I., & Judge, T. A. (2007). Fairness reactions to personnel selection techniques in Greece: The role of core self-evaluations. *International Journal of Selection and Assessment*, *15*, 206–219.

Nimura, H., Ishashiro, S., & Naito, J. (2000). A meta-analysis and validity generalization study of a personnel test and a general cognitive test for measuring managerial aptitude. *Japanese Journal of Administrative Science, 13*, 159–167.

Olea, M. M., & Ree, M. J. (1994). Predicting pilot and navigator criteria: Not much more than g. *Journal of Applied Psychology, 79*, 845–851.

Ones, D. S. (October 2004). *Validity of cognitive ability tests in selection: Quantitative summaries of meta-analyses*. Cattell Award address given at the annual conference of the Society for Multivariate Experimental Psychology, Naples, FL.

Ones, D. S. (April 2016). "g" versus specific abilities. In L. Lin (Chair.), *IGNITE session: I/O hot topics debate—dual or duel?* Symposium session at the annual conference of the Society for Industrial and Organizational Psychology, Anaheim, CA.

Ones, D. S., & Dilchert, S. (2004). *Practical versus general intelligence in predicting success in work and educational settings: A first-order and a second-order meta-analysis*. Paper presented at the University of Amsterdam, Amsterdam, The Netherlands.

Ones, D. S., Dilchert, S., & Viswesvaran, C. (2012). Cognitive abilities. In N. Schmitt (Ed.), *Oxford handbook of personnel assessment and selection* (pp. 179–224). New York, NY: Oxford University Press.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2004). Cognitive ability in selection decisions. In O. Wilhelm & R. W. Engle (Eds.), *Handbook of understanding and measuring intelligence* (pp. 431–468). Thousand Oaks, CA: Sage.

Ones, D. S., Viswesvaran, C., & Dilchert, S. (2005). Cognitive ability in personnel selection decisions. In A. Evers, O. Voskuijl, & N. Anderson (Eds.), *Handbook of selection* (pp. 143–173). Oxford, England: Blackwell.

Owens, W. (1966). Age and mental abilities: A second adult follow-up. *Journal of Educational Psychology, 57*, 311–325.

Plomin, R., & Spinath, F. M. (2004). Intelligence: Genetics, genes, and genomics. *Journal of Personality and Social Psychology, 86*, 112–129.

Ployhart, R. E., & Holtz, B. C. (2008). The diversity-validity dilemma: Strategies for reducing racioethnic and sex subgroup differences and adverse impact in selection. *Personnel Psychology, 61*(1), 153–172.http://dx.doi.org/10.1111/j.1744–6570.2008.00109.x

Potosky, D., Bobko, P., & Roth, P. L. (2005). Forming composites of cognitive ability and alternative measures to predict job performance and reduce adverse impact: Corrected estimates and realistic expectations. *International Journal of Selection and Assessment, 13*, 304–315.

Potosky, D., Bobko, P., & Roth, P. L. (2008). Some comments on Pareto thinking, test validity, and adverse impact: When "and" is optimal and "or" is a trade-off. *International Journal of Selection and Assessment, 16*, 201–205.

Ree, M. J., & Carretta, T. R. (1996). Central role of g in military pilot selection. *International Journal of Aviation Psychology, 6*, 111–123.

Ree, M. J., Carretta, T. R., & Steindl, J. R. (2001). Cognitive ability. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology: Vol. 1: Personnel psychology* (pp. 219–232). Thousand Oaks, CA: Sage.

Reeve, C. L., & Lam, H. (2007). Consideration of g as a common antecedent for cognitive ability test performance, test motivation, and perceived fairness. *Intelligence, 35*, 347–358.

Reeves, T. J., & Bennett, C. E. (2004). *We the people: Asians in the United States—Census 2000 special report* (CENSR-17). Washington, DC: U.S. Department of Commerce, Economics and Statistics Administration, U.S. Census Bureau.

Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin card sorting test: A meta-analytic review. *Psychology and Aging, 19*, 482–494.

Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology, 54*, 297–330.

Roth, P. L., Le, H., Oh, I-S., Van Iddekinge, C. H., Buster, M. A., Robbins, S. B., & Campion, M. A. (2014). Differential validity for cognitive ability tests in employment and educational settings: Not much more than range restriction? *Journal of Applied Psychology, 99*(1), 1–20. http://dx.doi.org/10.1037/a0034377

Rothstein, H. R., & McDaniel, M. A. (1992). Differential validity by sex in employment settings. *Journal of Business and Psychology, 7*, 45–62.

Rotundo, M., & Sackett, P. R. (1999). Effect of rater race on conclusions regarding differential prediction in cognitive ability tests. *Journal of Applied Psychology, 84*, 815–822.

Ryan, A. M., Boyce, A. S., Ghumman, S., Jundt, D., Schmidt, G., & Gibby, R. (2009). Going global: Cultural values and perceptions of selection procedures. *Applied Psychology: International Review, 58*, 520–556.

Ryan, A. M., & Greguras, G. J. (1998). Life is not multiple choice: Reactions to the alternatives. In M. D. Hakel (Ed.), *Beyond multiple choice: Evaluating alternatives to traditional testing for selection* (pp. 183–202). Mahwah, NJ: Lawrence Erlbaum.

Ryan, A. M., 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.

Ryan, A. M., Reeder, M., Golubovich, J., Grand J., Inceoglu, I., Bartram, D., Derous, E., Nikolaou, I., & Yao, X. (under review). Culture and testing practices: Is the world flat?

Sackett, P. R., Borneman, M. J., & Connelly, B. S. (2008). High stakes testing in higher education and employment: Appraising the evidence for validity and fairness. *American Psychologist, 63*, 215–227.

Sackett, P. R., De Corte, W., & Lievens, F. (2008). Pareto-optimal predictor composite formation: A complementary approach to alleviating the selection quality/adverse impact dilemma. *International Journal of Selection and Assessment, 16*, 206–209.

Sackett, P. R., & Ostgaard, D. J. (1994). Job-specific applicant pools and national norms for cognitive ability tests: Implications for range restriction corrections in validation research. *Journal of Applied Psychology, 79*, 680–684.

Sackett, P. R., & Roth, L. (1996). Multi-stage selection strategies: A Monte Carlo investigation of effects on performance and minority hiring. *Personnel Psychology, 49*, 549–572.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist, 56*, 302–318.

Sackett, P. R., & Shen, W. (April 2008). *International perspectives on the legal environment for selection*. Symposium conducted at the annual conference of the Society for Industrial and Organizational Psychology, San Francisco, CA.

Salgado, J. F. (In press). Using ability tests in selection. In H. Goldstein, E. Pulakos, J. Passmore, & C. Semedo (Eds.), *The Wiley Blackwell handbook of the psychology of recruitment, selection, and retention*. Hoboken, NJ: John Wiley & Sons.

Salgado, J. F., & Anderson, N. (2002). Cognitive and GMA testing in the European Community: Issues and Evidence. *Human Performance, 15*, 75–96.

Salgado, J. F., & Anderson, N. (2003). Validity generalization of GMA tests across countries in the European Community. *European Journal of Work and Organizational Psychology, 12*, 1–17.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., & de Fruyt, F. (2003). International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology, 56*, 605.

Salgado, J. F., Anderson, N., Moscoso, S., Bertua, C., de Fruyt, F., & Rolland, J. P. (2003). A meta-analytic study of general mental ability validity for different occupations in the European Community. *Journal of Applied Psychology, 88*, 1068–1081.

Salgado, J. F., Viswesvaran, C., & Ones, D. S. (2001). Predictors used for personnel selection: An overview of constructs, methods and techniques. In N. Anderson, D. S. Ones, H. K. Sinangil, & C. Viswesvaran (Eds.), *Handbook of industrial, work and organizational psychology: Vol. 1: Personnel psychology* (pp. 165–199). London, England: Sage.

Schaie, K. W. (1994). The course of adult intellectual development. *American Psychologist, 49*, 304–313.

Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior, 33*, 272–292.

Schmidt, F. L. (2011). A theory of sex differences in technical aptitude and some supporting evidence. *Perspectives on Psychological Science, 6*, 560–573.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62*, 529–540.

Schmidt, F. L., & Hunter, J. E. (1981). New research findings in personnel selection: Myths meet realities in the 1980s. *Management, 23*, 23–27.

Schmidt, F. L., & Hunter, J. (2004). General mental ability in the world of work: Occupational attainment and job performance. *Journal of Personality and Social Psychology, 86*, 162–173.

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

Schmidt, F. L., Hunter, J. E., Outerbridge, A. N., & Goff, S. (1988). Joint relation of experience and ability with job performance: Test of three hypotheses. *Journal of Applied Psychology, 73*, 46–57.

Schmidt, F. L., Le, H., Oh, I-S., & Shaffer, J. (2007). General mental ability, job performance, and red herrings: Responses to Osterman, Hauser, and Schmitt. *Academy of Management Perspectives, 21*, 64–76.

Schmidt, F. L., Pearlman, K., & Hunter, J. E. (1980). The validity and fairness of employment and educational tests for Hispanic Americans: A review and analysis. *Personnel Psychology, 33*, 705–724.

Schmitt, N., Rogers, W., Chan, D., Sheppard, L., & Jennings, D. (1997). Adverse impact and predictive efficiency of various predictor combinations. *Journal of Applied Psychology, 82*, 719–730.

Schneider, W. J., & McGrew, K. (2012). The Cattell-Horn-Carroll model of intelligence. In D. Flanagan & P. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (3rd ed., pp. 99–144). New York, NY: Guilford.

Schooler, C., Mulatu, M. S., & Oates, G. (1999). The continuing effects of substantively complex work on the intellectual functioning of older workers. *Psychology and Aging, 14*, 483–506.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 15*, 201–293.

Stern, W. (1911). *Die differentielle Psychologie in ihren methodischen Grundlagen [Differential psychology and its methodological foundations]*. Leipzig, Germany: J. A. Barth.

Taylor, P., Keelty, Y., & McDonnell, B. (2002). Evolving personnel selection practices in New Zealand organisations and recruitment firms. *New Zealand Journal of Psychology, 31*, 8–18.

Thadeu, S. H., & Ferreira, M. C. (2013). The validity of psychological assessment in a selection process in the area of public safety. *Revista Iberoamericana de Diagnóstico y Evaluación*, 117–145.

Tiffin, J., & Vincent, N. L. (1960). Comparison of empirical and theoretical expectancies. *Personnel Psychology, 13*, 59–64.

Tippins, N. (2015). Technology and assessment in selection. *Annual Review of Organizational Psychology and Organizational Behavior, 2*, 551–582.

United Nations High Commissioner for Refugees. (December 2015). *UNHCR mid-year trends 2015*. Retrieved from http://www.unhcr.org/statistics/unhcrstats/56701b969/mid-year-trends-june-2015.html

U.S. Census Bureau. (2010). *The Asian population: 2010*. Issued March 2012.

Vanderpool, M., & Catano, V. M. (2008). Comparing the performance of Native North Americans and predominantly White military recruits on verbal and nonverbal measures of cognitive ability. *International Journal of Selection and Assessment, 16*, 239–248.

Verhaeghen, P. (2003). Aging and vocabulary score: A meta-analysis. *Psychology & Aging, 18*, 332–339.

Viswesvaran, C., & Ones, D. S. (2002). Special issue: Role of general mental ability in industrial, work, and organizational psychology. *Human Performance, 15*.

Voyer, D., Voyer, S., & Bryden, M. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117*, 250–270.

Wai, J., Lubinski, D., & Benbow, C. P. (2005). Creativity and occupational accomplishments among intellectually precocious youths: An age 13 to age 33 longitudinal study. *Journal of Educational Psychology, 97*, 484–492.

Wiernik, B. M., Wilmot, M. P., & Kostal, J. W. (2015). How data analysis can dominate interpretations of dominant general factors. *Industrial and Organizational Psychology, 8*, 438–445.

Wonderlic Inc. (2002). *Wonderlic personnel test and scholastic level exam user's manual*. Libertyville, IL: Author.