CHAPTER 3

GENERAL COGNITIVE ABILITY

Michael A. McDaniel, George C. Banks

The success of an organization depends on the effective performance of its employees. For this reason, managers, supervisors, and other organizational leaders have a vested interest in the human capital that comprises their organization. Thus, leadership within an organization should strive to select and develop the most talented individuals. One way in which talented individuals can be identified is through the evaluation of their general cognitive ability. In this chapter, we review the research on cognitive ability that is of relevance to applied human resource researchers and practitioners.

General cognitive ability is the ability that consistently differentiates individuals on mental abilities regardless of the cognitive task or test (Jensen, 1998). Thus, general cognitive ability can be measured in a variety of ways with a variety of tests. For the purpose of this review, general cognitive ability is synonymous with traditional conceptions of intelligence such as IQ and Charles Spearman's general intelligence factor (*g*). This review is restricted to cognitive ability conceptions of intelligence and does not review other assessments that have adopted the word *intelligence*, such as emotional intelligence and practical intelligence. We also exclude measures and models of intelligence for which there is little research support.

Many well-known tests measure general cognitive ability for workplace assessment. Some general cognitive ability tests that are commonly used in workplace assessment are presented in the Appendix. Some of the many vendors of such tests are better than others. We list a few test publishers whose cognitive ability tests are well constructed, widely used, and for which there is excellent documentation.

Dominant Models of General Cognitive Ability

Multiple models describe the nature of general cognitive ability, but for the purposes of this overview, we focus on the most widely recognized and accepted models in the field. We begin with a review of Spearman's general intelligence factor, or *g*. We then summarize Cattell's model of crystallized and fluid intelligence and John Carroll's three-stratum model of cognitive ability.

General Intelligence Factor Model

Spearman (1904) presented a model of intelligence explained by two factors, the general cognitive ability factor (*g*) and the specific factor, abbreviated as *s* (Carroll, 1993; Jensen, 1998). The *s* allows an individual to be more effective at one cognitive ability test than another one when there is task- or domain-specific knowledge. Thus, *s* is unique from *g*, which is responsible for an individual's effectiveness on all tests of cognitive ability. Spearman was able to derive his model of the *g* factor from a method known as factor analysis (which he also invented). Through factor analysis, a researcher can determine correlations between specific variables in a cognitive test and a common variable that is present throughout tests (Jensen, 1998). Spearman named the common variable in this case *g*. Through factor analysis, he concluded that *g* is the common factor that differentiates individuals on all tests of cognitive ability. The primary reason that Spearman's model has prevailed as one of the dominant models of cognitive ability is that empirical research has produced strong results demonstrating the importance of the *g* factor in all cognitive ability tests (Jensen, 1998).

Cattell's Crystallized and Fluid Intelligences

Raymond Cattell, a student of Charles Spearman, proposed that crystallized and fluid intelligence are two underlying components of the *g* factor (Cattell, 1943; Jensen, 1998). Cattell's model, like Spearman's, also has substantial research support. Fluid intelligence is the ability to solve novel problems through reasoning. Crystallized intelligence is the ability to rely on prior experience and knowledge to solve problems. Individual differences in fluid intelligence contribute to individual differences in crystallized intelligence. Thus, it is not surprising that fluid and crystallized intelligence measures are highly correlated. Fluid intelligence tends to peak in early adulthood and then declines with age.

Jensen (1998, p. 123) offered the following two items:

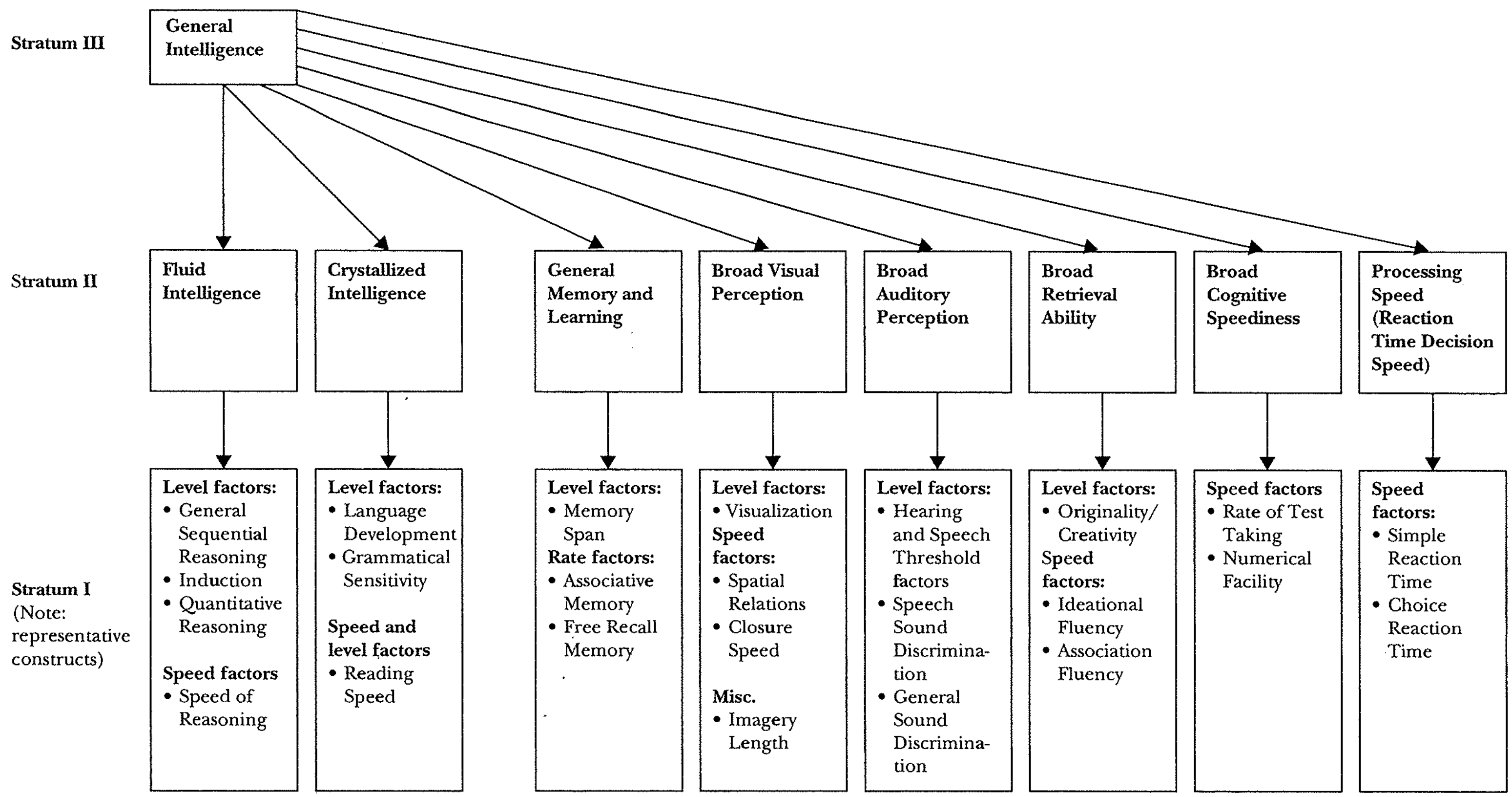
1. Temperature is to cold as height is to: (a) hot (b) inches (c) size (d) tall (e) weight.
2. Bizet is to Carmen as Verdi is to: (a) Aïda (b) Elektra (c) Lakmé (d) Manon (e) Tosca.

The first item primarily places demands on fluid intelligence to see the logical relationships (d is the answer). The second item primarily places demands on crystallized intelligence, particularly on knowledge related to opera (a is the answer).

Carroll's Three-Stratum Theory

Carroll's three-stratum theory (Carroll, 1993) complements the other two models of cognitive ability. In extensive series of factor analyses, Carroll offered evidence for three layers, or strata, of cognitive ability that attempt to explain a narrow, broad, and more general level of cognitive ability (Jensen, 1998). Carroll devised his model to describe Spearman's *g* factor as his stratum III or more general level. Carroll's broad level, stratum II, incorporates Cattell's fluid and crystallized intelligence, as well as six other factors. The third level Carroll proposed is the narrow level, or stratum I. In Carroll's model, this lowest level addresses sixty-nine abilities (Jensen, 1998). Figure 3.1 graphically displays this model. Note the space gap in stratum II between crystallized intelligence and general memory and learning. This gap was intentional to show the relative importance of the stratum 2 abilities

Figure 3.1. Carroll's Three-Stratum Theory of Cognitive Ability



Source: Adapted with permission from Carroll (1993), Figure 15.1 (p. 627). Copyright Cambridge University Press.

to general intelligence. Fluid and crystallized intelligence are the most highly related factors to general intelligence. The other factors are related, but to a lesser degree. The order of the lesser contributors is also important and reflects their contribution to general intelligence. For example, memory contributes more to general intelligence than does broad visual perception.

Cognitive Ability and the World of Work

Cognitive ability and the world of work spans several topics. The knowledge base on cognitive ability is intricately tied to a body of knowledge called validity generalization, which we examine first in this section. Next, we examine the value of cognitive ability in the prediction of training performance and job performance. We then address two issues relevant to the use of cognitive ability tests: the linearity of prediction and the relative value of general versus specific abilities in prediction. We follow this with a discussion of evidence relevant to three demographic subgroup issues: race/demographic mean differences in cognitive ability, validity for different subgroups, and the failure of efforts to reduce subgroup mean differences by altering items.

Validity Generalization

Beginning in the 1920s, it was observed that different applications of the same general cognitive ability test yielded different validity results, that is, the magnitude of the relationship between the test and job performance varied (Schmidt & Hunter, 1998). This caused some to speculate that there were some unknown characteristics of the situation in which the test was used that caused the general cognitive ability test to be predictive of job performance in one situation (for example, a teller job in a bank on Oak Street), but not predictive or less predictive in a different situation (a teller job in a bank on Elm Street). This speculation became known as the situational specificity hypothesis (Schmidt & Hunter, 1998). Because the situational characteristics presumably causing these validity differences were not identified through job analysis, it became the custom that the validity of a general cognitive ability test was examined in each setting in which the test was used.

In the late 1970s, Schmidt and Hunter questioned the legitimacy of the situational specificity hypothesis. They demonstrated that random sampling error, and not some unknown situational moderator, was the primary reason that the validity of general cognitive ability tests varied across applications (Schmidt & Hunter, 1977). In addition, they documented how measurement error made general cognitive ability tests appear less valid than they were. Also, differences across studies in the extent of measurement error, in addition to random sampling error, caused predictive validity of general cognitive ability tests to vary across situations. Similarly, they documented how range restriction made general cognitive ability tests appear to be less valid than they were and that differences across studies in the degree of range restriction caused the predictive validity of general cognitive ability tests to vary across situations.

In brief, the seminal work of Schmidt and Hunter demonstrated that the validity of general cognitive ability tests was quite stable across settings once one adjusted the data for random sampling error and differences across studies in measurement error and range restriction. Schmidt and Hunter also demonstrated that the true predictive validity of general cognitive ability tests had been substantially underestimated due to a failure to consider the effects of measurement error and range restriction.

Prediction of Training Performance

No one can perform successfully in a job without the necessary job knowledge and job skills. Although an organization can select applicants who have the needed knowledge and skills, many jobs require that new employees be trained after they are hired. These jobs include entry jobs in the police and fire services and jobs in the military. The acquisition of skill and knowledge depends on learning, and the rate of learning is related to general cognitive ability (Jensen, 1998). It should not be surprising that general cognitive ability tests are excellent predictors of success in training programs for newly hired employees. Two types of studies are most useful for summarizing the value of general cognitive ability tests in the prediction of training performance.

The first type of particularly valuable studies are validity generalization studies that have examined the value of general cognitive ability in the prediction of training performance in specific job families. For example, Barrett, Polomsky, and McDaniel (1999) reported a validity of .77 for general cognitive ability tests and firefighter training performance. Pearlman, Schmidt, and Hunter (1980) reported a validity of .71 for general cognitive ability tests and clerical training performance. Hirsh, Northrop, and Schmidt (1986) reported a validity of .71 for a general cognitive ability composite for police and detective training performance. There are many additional studies, including ones that that have examined semiprofessional occupations (Trattner, 1985) and blue-collar jobs where apprenticeship training was common (Northrop, 1988). Lilienthal and Pearlman (1983) reported general cognitive ability validities for training performance in entry-level aid and technician occupations in the health, science, and engineering fields. Hunter and Hunter (1984) reported training validities for a large number of jobs grouped by job complexity (the cognitive demands of the jobs). These validities varied from .50 to .65.

The second type of valuable studies are primary validity studies that use military data. Such studies are of particular value because they use well-developed tests of general cognitive ability, everyone who enters the military receives some job training, and the sample sizes are huge. Olea and Ree (1994) reported validities for general cognitive ability tests and pilot and navigator training in the U.S. Air Force. General cognitive ability predicted training performance (validity = .31) for the pilots and .46 for the navigators. Earles and Ree (1992) examined the validity of general cognitive ability in 150 military training programs using data from 88,724 U.S. Air Force recruits. The general cognitive ability test showed impressive validities across all 150 training programs.

Prediction of Job Performance

Schmidt and Hunter (1998), in their review of eighty-five years of employment testing research, offered .51 as the validity of general cognitive ability and job performance. This value corresponds to the validity of general cognitive ability for medium complexity jobs as reported by Hunter and Hunter (1984), who documented that the validity of cognitive ability tests increases as the complexity (cognitive demands) of the jobs increases. The validity for the most complex jobs is .56, and for the least complex jobs the validity is .23. Ree, Earles, and Teachout (1994) documented the validity of general cognitive ability for seven diverse U.S. Air Force jobs, reporting an average validity across jobs of .42. Many additional studies evaluate the validity of general cognitive ability for the prediction of job performance (see McDaniel, 2007). Without any doubt, general cognitive ability is a large-magnitude, robust predictor of job performance.

Prediction Linearity

A common misperception is that beyond a certain level, general mental ability has no practical importance in the prediction of training and job performance. This is simply untrue. The relationship between general cognitive ability and both training and job performance is linear (Coward & Sackett, 1990). This means that on average, higher-scoring applicants will have better job performance than lower-scoring applicants. Thus, to maximize the productivity of those hired, applicants can be rank-ordered from high to low on test scores and hired in that order.

General Versus Specific Cognitive Abilities

Although this chapter concerns general cognitive ability measures, it is useful to understand the relations between general cognitive ability and specific cognitive abilities in the prediction of training performance. Examples of specific cognitive ability tests are arithmetic reasoning, reading comprehension, and word knowledge. Two lines of evidence are useful in examining the contribution of specific abilities relative to general cognitive ability in the prediction of training and job performance. One line of evidence concerns the relationship between the validity of a specific cognitive ability test and its relationship with general cognitive ability (the loading of the specific cognitive ability test on the general cognitive ability factor). Two large U.S. military studies have examined this with respect to training performance validity (Ree & Earles, 1992, 1994). These studies clearly showed that the validity of a specific cognitive ability test increases to the extent that the test has a large loading on the general cognitive ability factor. That is, the larger the general cognitive ability loading of the specific cognitive ability measures, the larger the validity of the specific cognitive ability. A second line of evidence is the extent to which specific cognitive abilities predict training and job performance over and above general cognitive ability. Ree and Earles (1994) and Ree et al. (1994) showed that specific cognitive abilities do little in the prediction of training and job performance over and above general cognitive ability. In summary, with respect to validity in the prediction of training and job performance, prediction is a function of general cognitive ability and specific abilities have little to no additional value.

Differences on Cognitive Ability Among Ethnic/Race Demographic Groups

There are large differences (about one standard deviation difference) between mean levels of general cognitive ability for whites and blacks in the U.S. population (Jensen, 1998). Although smaller than the black-white differences in mean levels of cognitive ability, the mean differences between Hispanics and whites are also large. Asian-white mean differences in cognitive ability are usually small and often favor Asians. Unfortunately, the mean race differences between blacks and whites have remained fairly constant for at least ninety years (Jensen, 1998). Although some researchers argue that the black-white mean gap is narrowing, others argue that such claims are incorrect. Regardless of whether the gap is constant or somewhat narrowing, the black-white and Hispanic-white mean differences are currently large enough to result in disparate hiring rates (for example, hiring a larger proportion of the white applicants than the black or Hispanic applicants).

For at least two reasons, the differences between whites and blacks in job-applicant settings are smaller than the one standard deviation difference in the population. One reason is that those at the lowest levels of general cognitive ability are not job applicants. Second, jobs often have educational or experience requirements. Job applicants who meet the educational and experience requirements will be more equal in general cognitive ability than a random sample of the population. Thus, one would expect large mean ethnic differences for jobs with low educational and experience requirements and smaller mean ethnic differences for jobs with extensive educational and experience requirements. Thus, for example, mean ethnic differences should be smaller for those whose highest educational credential was a four-year college degree than for those who did not graduate from high school. Roth, BeVier, Bobko, Switzer, and Tyler (2001) provided the best estimates of mean ethnic differences in general mental ability among applicant groups. The mean difference is expressed as standardized mean differences (d) in a standard deviation metric. Thus, if $d = 1$, the white mean on general cognitive ability is one standard deviation higher than the black mean. Based on 125,654 applicants, the mean d for low-complexity jobs was .86. For medium-complexity jobs (31,990 applicants), the mean d was .72. Based on 4,884 applicants, the mean d for high-complexity jobs was .63. This pattern of larger mean differences for lower-complexity jobs is consistent with the expected mean differences varying as a function of educational and experience requirements. At all complexity levels, these differences are large and will cause disparate hiring rates by race if hiring is solely based on scores on general cognitive ability tests in a race-blind manner. Adding other predictors to a cognitive ability test battery may reduce the disparate hiring rates.

Single-Group Validity, Differential Validity, and Differential Prediction

The large mean differences between whites and blacks in general cognitive ability raise the possibility that general cognitive ability tests are biased against blacks. The hypothesis of single-group validity holds that a general cognitive ability test has validity for one group (for example, whites) but zero validity for another group (for example, blacks). The hypothesis of differential validity holds that a test has validity for one group but a substantially different validity for another group. After substantial research examined this issue, the National Academy of

Sciences concluded that single group and differential validity is very uncommon (Wigdor & Garner, 1982).

Over time researchers realized that the critical issue was differential prediction because it is possible to have a test with equal validity for two groups, but the optimal regression line to predict job performance might not be the same. Differential prediction may occur in two ways: the regression slopes might be different, or the regression intercepts might be different. The conclusion from the literature is that different slopes by race do not occur (Bartlett, Bobko, Mosier, & Hannan, 1978).

The finding of different intercepts by race is more common. However, the error in prediction favors minority groups. That is, when a common regression line is used to predict job performance for all races, the job performance of Hispanics and blacks is overpredicted on average (Hartigan & Wigdor, 1989). Therefore, when differential prediction occurs in employment tests, it is to the advantage of black and Hispanic applicants and the disadvantage of white applicants. One could obtain accurate prediction for all groups by using different regression lines for each group, but that would be a violation of the 1991 Civil Rights Act for U.S. employers.

Altering Item Types to Reduce Mean Ethnic Differences

Some speculate that mean ethnic differences may be reduced in employment tests by altering the tests in some way. Outside of cognitive ability testing, one can reduce the mean ethnic differences by reducing the extent to which the test correlates with general cognitive ability. For example, Whetzel, McDaniel, and Nguyen (2008) showed that the magnitude of ethnic differences in situational judgment tests was largest for those tests most highly correlated with cognitive ability. Video is often used to reduce the cognitive loading of employment tests because it reduces the reading demands of the test.

For cognitive ability tests, the issue of test format and mean ethnic score differences is typically discussed as an issue of cultural loading. The idea is that test items may reflect the culture and experience of majority group applicants more so than those of minority groups, and the resulting score differences are functions of the cultural loading and not differences in general cognitive ability. Jensen (1990) reviewed efforts to build culture-reduced measures of general cognitive ability. One such effort was the Davis-Eells Games measure. The authors of the measure sought to remove verbal, abstract, and academic content from their test items. Their items consisted of one-frame cartoons that displayed pictures of people in familiar situations. The test examiner would ask a question about the cartoon, and the correct answer would be based on a reasonable inference about the situation displayed in the cartoon. Jensen noted that the test did not result in smaller mean ethnic differences. The same conclusion was drawn for all of the culturally reduced tests reviewed. Thus, there is no evidence that one can alter the magnitude of mean ethnic differences in general cognitive ability tests by changing item types.

Guiding Practice

In this section we address issues related to guiding practice. We focus on the diversity-validity dilemma, content validity considerations in selection measures, and validity generalization in the context of the *Uniform Guidelines, Principles,* and *Standards.*

The Diversity-Validity Dilemma

Most employers seek to simultaneously obtain a diverse workplace and hire the best available applicants (Pyburn, Ployhart, & Kravitz, 2008) because both goals can potentially improve organizational effectiveness. When attempting to address a diversity-validity dilemma, an organization must begin by evaluating underlying assumptions of effectiveness. Once it establishes a set of criteria, it can proceed to operationalize practices and procedures linked to effective hiring (Banks, 2008).

Merit selection is the hiring of the best available applicants. General cognitive ability is the best single predictor of training and effective job performance. It also has the largest white-black mean differences. This potentially causes racial diversity and merit selection to be competing goals. Pyburn et al. (2008) referred to this as the *diversity-validity dilemma.* Employers respond to this dilemma in various ways. Some use predictors with lower validity and lower adverse impact than general cognitive ability tests to reduce ethnic mean differences in hiring rates. Although this promotes diversity, it can make the organization less effective because the lower validity tests are not as efficient in identifying the best employees. Other employers use general cognitive ability tests and accept the social and legal consequences of hiring fewer minorities. Pyburn et al. (2008) reviewed the legal issues facing employers due to the diversity-validity dilemma.

Ployhart and Holtz (2008) reviewed sixteen approaches for reducing demographic subgroup differences. The two most relevant to cognitive ability are using a predictor other than cognitive ability and manipulating the scoring of tests. Both approaches involve reducing the influence of general cognitive ability in the selection process, and most conclude that validity will or may suffer when compared to that of general cognitive ability. Concerning the first approach, one can use something other than general cognitive ability (say, a personality test); however, such predictors typically have lower validity, and often much lower validity.

An example of the second approach is to add other predictors to a selection system already containing general cognitive ability. Ployhart and Holtz (2008) noted that this strategy may reduce mean race differences and yield validity higher than general cognitive ability alone. For example, McDaniel, Hartman, Whetzel, and Grubb (2007) showed that adding a situational judgment test or a Big Five personality assessment to a measure of general cognitive ability could yield a larger validity than general cognitive ability alone. Another example of the second approach concerns banding where tests are grouped into categories (for example, well qualified, qualified, not qualified) and it is asserted, incorrectly, that all applicants in the same category are equally qualified. We note that any banding should serve to reduce the validity of the general cognitive ability test on average, because it obscures real score differences among applicants. Ployhart and Holtz noted that banding does little to reduce subgroup differences unless preferences are given to the subgroup (for example, the hiring manager is forced or influenced to hire the minority in the band before hiring any of the majority group members in the band). A final example of the second approach concerns manipulating the weighting of predictors to minimize subgroup differences. For example, in a selection battery to screen police officers, the reading examination was scored pass-fail, and the passing score was set at the score corresponding to the bottom 1 percent of the incumbent officers. This effectively stripped all cognitive ability variance from the test battery (Gottfredson, 1996).

In summary, the diversity-validity dilemma is a serious one with no ready solution. A general cognitive ability test is typically the best predictor of job performance. Typically it is also the predictor with the largest mean ethnic group differences. In general, strategies that serve to reduce the magnitude of the mean group differences in general cognitive ability reduce the validity of the test battery. One exception to this observation is the supplementing of a general cognitive ability test with a measure with lower mean subgroup differences, such as a measure of personality or of situational judgment. When the general cognitive ability measures and these other measures are optimally weighted in a selection composite, the validity will likely exceed that of the general cognitive ability measure alone, and the mean subgroup difference may be somewhat smaller. However, Potosky, Bobko, and Roth (2005) noted that the reductions in mean ethnic differences are often minimal.

Because there appears to be no magic bullet to reduce mean race differences in tests of general cognitive ability, employers who use these tests may wish to consider some forms of affirmative action to promote diversity without sacrificing the integrity of their selection system. Kravitz (2008) reviewed a variety of affirmative action efforts to increase diversity and recommended efforts that do not entail preferences for minorities in hiring.

Content-Validity Considerations in the Selection of Measures

Content validity concerns the extent to which the content of the test represents the content of the job. General cognitive ability can be measured with a variety of tests. From the perspective of applicant acceptance, it may be useful to identify cognitive ability tests that reflect job content. For example, logic-based measurement approaches for measuring verbal ability (Colberg, 1985) are offered as methods to measure cognitive ability in a manner to reflect job content. Such measures likely have a positive effect on an applicant's reactions to the test. Likewise, reading comprehension items can be based on material that is typically read on the job.

The key to building a general cognitive ability measure from a set of content-valid specific abilities is to use a diverse set of such tests so that they have sufficient breadth of coverage to yield a composite measuring general cognitive ability. If specific ability tests such as memory or spatial ability are given, it is preferable to combine several of them so that the resulting test battery will be a good assessment of general cognitive ability. The use of a single specific ability test is likely to yield lower validity because it does not fully assess general cognitive ability.

Validity Generalization and the *Uniform Guidelines, Principles,* and *Standards*

Substantial research has shown that cognitive ability tests predict job performance at some level for all jobs. This conclusion is primarily based on validity generalization studies. Landy (2003) and McDaniel (2007) reviewed the status of the validity generalization with respect to the *Uniform Guidelines on Employee Selection Procedures* (Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice, 1978). One notable issue is that validity generalization is not addressed in the *Uniform Guidelines.* In part, this oversight is due to the bulk of the validity generalization research being published after the propagation of the *Uniform Guidelines.* However, with a careful reading of the *Uniform Guidelines,* one might have hope that this oversight would be resolved because the *Uniform Guidelines* state:

> The provisions of these guidelines relating to validation of selection procedures are intended to be consistent with generally accepted professional standards for evaluating standardized tests and other selection procedures, such as those described in the Standards for Educational and Psychological Tests prepared by a joint committee of the American Psychological Association, the American Educational

> Research Association, and the National Council on Measurement in Education. . . . and standard textbooks and journals in the field of personnel selection [section 5C].

As McDaniel (2007) noted, the Society of Industrial and Organization Psychology (SIOP) contacted the agencies that authored the *Uniform Guidelines* shortly after their release to detail the ways in which the guidelines were inconsistent with professional practice and guidance. Unfortunately, the letter did not result in a revision of the *Uniform Guidelines*. Thus, although the *Uniform Guidelines* state that they are intended to be consistent with professional standards, the federal agencies that are responsible for them have not called for their revision during the past thirty years.

Professional organizations with relevance to employment testing have acknowledged the significant scientific status of validity generalization analyses (McDaniel, 2007). One set of testing guidelines, the *Standards for Educational and Psychological Testing* is published jointly by three professional organizations (American Educational Research Association, the American Psychological Association, & the National Council on Measurement in Education, 1999). The other major set of testing guidelines, the *Principles for the Validation and Use of Personnel Selection Procedures*, is published by SIOP (2003). Both sets of guidelines comment favorably on the use of validity generalization as a means for establishing the validity of an employment test. The *Principles* state that validity generalization findings for cognitive ability tests are "particularly well established" (p. 28). Thus, although the *Uniform Guidelines* are mute with respect to validity generalization, the professional guidelines support the value of validity generalization in documenting the validity of general cognitive ability tests and other personnel selection tests.

Copus (2006) provided a critique of employment discrimination enforcement efforts of the U.S. Office of Federal Contract Compliance Programs (OFCCP) that rely on the *Uniform Guidelines* and provided compelling arguments for greater acceptance of validity generalization results in the enforcement field. Copus argued that OFCCP's perspective on test validation is inconsistent with professional standards in two key ways. First, Copus criticized the OFCCP for requiring that an employer conduct a validation study when the use of the test results in a disparate hiring rate for women or minorities. Second, Copus asserted that OFCCP inappropriately requires fairness studies to determine if tests have the same degree of job relatedness for majority and minority applicants. Copus noted that the validity generalization evidence for the validity of cognitive ability tests is so overwhelming that local validation studies, even in the presence of disparate impact, are unneeded. He wrote that "validity generalization research in lieu of local studies fully satisfies the professional standards set forth in both the APA *Standards* and SIOP *Principles*" (p. 6). Copus then reviewed the research literature on validity generalization and discredits OFCCP's requirement for fairness studies by reviewing the literature on differential validity and prediction.

McDaniel (2007) speculated on why the *Uniform Guidelines* have not been revised to reflect current scientific knowledge as expressed in professional documents such as the *Principles* and *Standards*:

> A primary use of the *Uniform Guidelines* is to pressure employers into using suboptimal selection methods in order to hire minorities and Whites at approximately the same rates. If employers do not hire minorities at about the same rates as Whites, the *Uniform Guidelines* are invoked by enforcement agencies and plaintiffs to require the employer to prepare substantial validity documentation. . . . In other areas of federal regulations and guidelines, such regulations and guidelines are often updated to reflect scientific knowledge and professional practice. It is well past the time for the *Uniform Guidelines* to be revised [pp. 168–169].

Conclusion

This chapter has reviewed research and practice in the use of general cognitive ability tests in workplace assessment. We examined models of general cognitive ability and detailed research and applied issues in the use of general cognitive ability tests in workplace assessment. Cognitive ability tests show substantial validity for all jobs and exceed the validity of other tests. These tests are not biased against minorities in the prediction of job performance, although one can expect blacks and Hispanics on average to score lower than whites and Asians. The mean differences occur because the tests accurately measure the mean differences in cognitive ability in the population. There are many inexpensive, commercially available cognitive ability tests. Any merit-based selection system should include such tests.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (1999). *Standards for educational and psychological testing.* Washington, DC: American Educational Research Association.

Banks, G. P. (2008). *The issue of race: A resolution for the 21st century.* New York: Seabury Press.

Barrett, G. V., Polomsky, M. D., & McDaniel, M. A. (1999). Selection tests for firefighters: A comprehensive review and meta-analysis. *Journal of Business and Psychology, 13,* 507–514.

Bartlett, C. J., Bobko, P., Mosier, S. B., & Hannan, R. (1978). Testing for fairness with a moderated multiple regression strategy: An alternative to differential analysis. *Personnel Psychology, 31,* 233–241.

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

Cattell, R. (1943). The measurement of adult intelligence: Present practice in adult intelligence testing. *Psychological Bulletin, 40,* 153–193.

Colberg, M. (1985). Logic-based measurement of verbal reasoning: A key to increased validity and economy. *Personnel Psychology, 38,* 347–359.

Copus, D. (2006, March 27). *Validation of cognitive ability tests.* Letter to Charles James, Office of Federal Contract Compliance Programs. Morristown, NJ: Ogletree Deakins.

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

Earles, J. A., & Ree, M. J. (1992). The predictive validity of the ASVAB for training grades. *Educational and Psychological Measurement, 52,* 721–725.

Equal Employment Opportunity Commission, Civil Service Commission, Department of Labor, & Department of Justice. (1978). *Uniform guidelines on employee selection procedures. Federal Register, 43*(166), 38290–39315.

Gottfredson, L. S. (1996). Racially gerrymandering the content of police tests to satisfy the U.S. Justice Department: A case study. *Psychology, Public Policy, and Law, 2,* 418–446.

Hartigan, J. A., & Wigdor, A. K. (Eds.). (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery.* Washington, DC: National Academy Press.

Hirsh, H. R., Northrop, L. C., & Schmidt, F. L. (1986). Validity generalization results for law enforcement occupations. *Personnel Psychology, 39,* 399–420.

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

Jensen, A. R. (1990). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

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

Landy, F. J. (2003). Validity generalization: Then and now. In K. R. Murphy (Ed.), *Validity generalization: A critical review* (pp. 155–195). Mahwah, NJ: Erlbaum.

Lilienthal, R. A., & Pearlman, K. (1983). *Validity of federal selection tests for aide/technicians in the health, science, and engineering fields.* Washington, DC: National Technical Information Service. (NTIS no. PB83202051).

McDaniel, M. A. (2007). Validity generalization as a test validation approach. In S. M. McPhail (Ed.), *Alternative validation strategies* (pp. 159–180). San Francisco: Jossey Bass.

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

Northrop, L. C. (1988). *Situational specificity and validity generalization results for apprentice occupations.* Washington, DC: U.S. Office of Personnel Management. (NTIS no. PB88213608).

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

Pearlman, K., Schmidt, F. L., & Hunter, J. E. (1980). Validity generalization results for tests used to predict job proficiency and training criteria in clerical occupations. *Journal of Applied Psychology, 65,* 373–406.

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,* 153–172.

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.

Pyburn, K. M. Jr., Ployhart, R. E., & Kravitz, D. A. (2008). The diversity-validity dilemma: Overview and legal context. *Personnel Psychology, 61,* 143–151.

Ree, M. J., & Earles, J. A. (1992). Intelligence is the best predictor of job performance. *Current Directions in Psychological Science, 1,* 86–89.

Ree, M. J., & Earles, J. A. (1994). The ubiquitous predictiveness of *g.* In M. G. Rumsey, C. B. Walker, & J. H. Harris (Eds.), *Personnel selection and classification* (pp. 127–135). Mahwah, NJ: Erlbaum.

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

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

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

Society for Industrial and Organizational Psychology. (2003). *Principles for the validation and use of personnel selection procedures* (4th ed.). Bowling Green, OH: Author.

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

Trattner, M. H. (1985). *Estimating the validity of aptitude and ability tests for semiprofessional occupations using the Schmidt-Hunter interactive validity generalization procedure.* Washington, DC: U.S. Office of Personnel Management. (NTIS no. PB86169463).

Whetzel, D. L., McDaniel, M. A., & Nguyen, N. T. (2008). Subgroup differences in situational judgment test performance: A meta-analysis. *Human Performance, 21,* 291–309.

Wigdor, A. K., & Garner, W. R. (Eds.). (1982). *Ability testing: Use, consequences, and controversies.* Washington, DC: National Academy Press.