ASSOCIATION FOR
PSYCHOLOGICAL SCIENCE

Fact and Fiction in Cognitive Ability Testing for Admissions and Hiring Decisions

Current Directions in Psychological Science
19(6) 339-345

DOI: 10.1177/0963721410389459
http://cdps.sagepub.com
SAGE

Nathan R. Kuncel[1] and Sarah A. Hezlett[2]
[1] University of Minnesota and [2] Personnel Decisions Research Institutes

Abstract

Standardized measures of intelligence, ability, or achievement are all measures of acquired knowledge and skill and have consistent relationships with multiple facets of success in life, including academic and job performance. Five persistent beliefs about ability tests have developed, including: (a) that there is no relationship with important outcomes like creativity or leadership, (b) that there is predictive bias, (c) that there is a lack of predictive independence from socioeconomic status, (d) that there are thresholds beyond which scores cease to matter, and (e) that other characteristics, like personality, matter as well. We present the evidence and conclude that of these five beliefs, only the importance of personality is a fact; the other four are fiction.

Keywords

standardized tests, intelligence, cognitive ability, admissions tests, test bias, job performance, academic success

Standardized tests of cognitive abilities, such as college admissions tests, are some of the strongest and most consistent predictors of performance in educational and work settings. It therefore is not surprising that tests are widely used to aid admission and personnel selection decisions. This has rightly made standardized tests of cognitive abilities the focus of intense scrutiny, and a positive outcome of this attention has been the accumulation of a great deal of knowledge. The downside has been the widespread proliferation of erroneous ideas. The purpose of this article is to summarize what is known about the relationship between tests of cognitive ability and performance while addressing common concerns and misperceptions.

We begin by clarifying the nature of standardized tests of cognitive abilities. Then we discuss the scope of research on such tests, focusing on their predictive relationships with performance in college, graduate school, and jobs. Next, we consider why scores on cognitive ability tests predict performance, and we address concerns about the complex relationship between scores on cognitive ability tests and socioeconomic status (SES). We then briefly examine personality and other alternate predictors of academic and work performance. Finally, we tackle some of the most controversial and misunderstood aspects of such tests' use: adverse impact and test bias.

Definition

Although the surface features of standardized tests of cognitive abilities are familiar to those who have taken them, what they measure may sometimes seem obscure. Most assess a combination of reasoning, verbal, and quantitative skills or discipline-specific knowledge, which are correlated and fit into a hierarchical structure with a single overarching general ability. This means that those who do well on a test with one kind of content (e.g., mathematics) will tend to do well on tests with different content (e.g., verbal skills). At the same time, people reliably demonstrate relative strengths and weaknesses in different domains.

Standardized tests of cognitive abilities are grounded in the psychometric approach to intelligence, which has focused on understanding individuals' ability to reason, plan, solve problems, think abstractly, learn and adapt, and process and comprehend complex ideas and information (Ones, Visweswaran, & Dilchert, 2005). This does not mean that cognitive tests are pure measures of individuals' *innate* ability. Although highly stable over the course of decades (e.g., Deary, Whalley, Lemmon, Crawford, & Starr, 2000), test scores reflect developed abilities and are a function of innate talent, learned knowledge and skills, and environmental factors that influence knowledge and skill acquisition, such as prior educational opportunities.

Corresponding Author:
Nathan R. Kuncel, 75 East River Road, University of Minnesota, Minneapolis MN 55455
E-mail: kunce001@umn.edu

Predictive Validity

Researchers have conducted thousands of studies in educational and employment settings to answer the basic question "Do cognitive tests predict performance?" This literature is so large that individual studies can be found and selectively cited to support the argument that cognitive test scores are nearly perfect predictors of performance measures or the argument that cognitive test scores are unrelated to performance in school and work. Large-scale studies and meta-analyses offer the most accurate estimates of the typical relationship between tests and performance (Kuncel & Hezlett, 2007), and they identify factors that affect the strength of the relationship between scores and performance, often called moderators.

In considering these relationships, a good question is "What correlation is large enough to matter?" Moderate relationships between predictors and criteria often are inappropriately discounted. For example, correlations of .30 have been dismissed as accounting for less than 10% of the variance in the criteria. However, this relationship is sufficiently large that hiring or admitting individuals who score better on the test can double the rate of successful performance. For example, 67% of individuals who score in the top quintile on a predictor will have above-average performance, compared to only 33% of individuals who score in the bottom quintile (Sackett, Borneman, & Connelly, 2008). Similarly, consider the hypothetical, but realistic, scenario presented in Table 1, in which 60% of individuals who score above average on an admissions test finish graduate school and only 40% of those who score below average attain their degrees. Degree attainment is multiply determined and what the test measures is clearly not the sole determinant of success, yet selecting individuals with above- versus below-average scores results in a 20% swing in an important outcome that has individual, organizational, and societal implications. Note that this situation reflects a correlation of only .20 (4% of the variance). Thus, even if the average correlation between scores on standardized tests of cognitive abilities and desired outcomes were as low as .20, those scores would provide valuable information. For most academic and work outcomes, the predictive power of test scores is much higher and improves prediction accuracy beyond other measures like prior grades or interviews (Kuncel & Hezlett, 2007; Ones et al., 2005; Sackett et al., 2008).

Predicting academic performance

Research on the prediction of college and graduate school performance has assessed performance using diverse measures, including grades, degree attainment, faculty evaluations of performance, professional licensure, and even research productivity. Figure 1 summarizes results from a number of large-scale meta-analytic reviews of admissions tests scores' correlations with subsequent academic performance (Kuncel & Hezlett, 2007; Berry & Sackett, 2009). Several consistent patterns in the results from thousands of studies on hundreds of thousands of students are clear. First, test scores are positive predictors of diverse indices of academic performance but are less strongly correlated with motivationally determined outcomes. For example, meta-analytic estimates of the average correlation between scores on graduate admissions tests and graduate school GPA (corrected for range restriction and criterion unreliability) range between .35 and .46; comparable estimates for degree completion range from .13 to .39 (Kuncel & Hezlett, 2007). Second, scores on tests that are specific to a particular content area or discipline (e.g., GRE Subject tests) tend to be better predictors than scores on tests measuring broader general math and verbal skills. If students are going to pursue a specific course of study, all else equal, assessing knowledge and skill in that field will yield the most predictive power.

Table 1. Hypothetical Percent of Students Graduating or Dropping Out by Admissions Test Score

Score	School outcome	
	Graduate	Drop out
Above average	60	40
Below average	40	60

Predicting work performance and training success

The power of tests for predicting job performance has also been examined across a large number of studies and jobs. Figure 2 summarizes results from the largest of these review studies. Consistent with research in academic domains, scores on cognitive ability tests are strongly related to success in occupational training in both civilian and military jobs, with meta-analytic estimates ranging from the high .30s to 70s (Ones et al., 2005). Cognitive ability test scores also predict outcomes in all jobs including overall job performance, objective leadership effectiveness, and assessments of creativity. Looking across the results of multiple meta-analyses, estimates of the average validity of general mental ability for predicting job performance (corrected for range restriction and measurement error in the criterion) converge around .50 (Ones et al., 2005). The strength of the relationship between test scores and performance increases as training and jobs become more cognitively complex (Ones et al., 2005; Schmidt & Hunter, 1998).

Are there predictive limits?

Even if ability measures are correlated with academic and work performance, perhaps test scores only matter to a certain point. For example, people with very high scores may not perform any better than those with merely high scores. Under these circumstances, the relationship between test scores and performance would be curvilinear and there would be a "ceiling" on scores beyond which having a higher score would not correspond to increased performance.

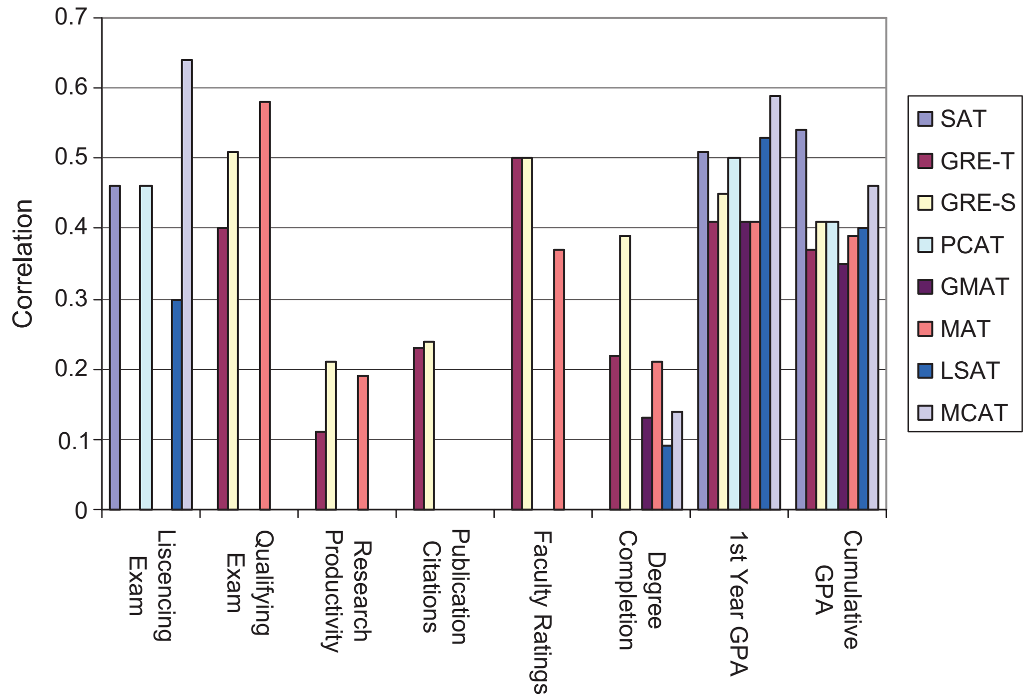

Fig. 1. Correlations between standardized admissions tests scores and measures of academic performance. Coefficients are corrected for measurement error and restriction of range where possible. Results presented are from Berry and Sackett (2009) and Kuncel and Hezlett (2007).

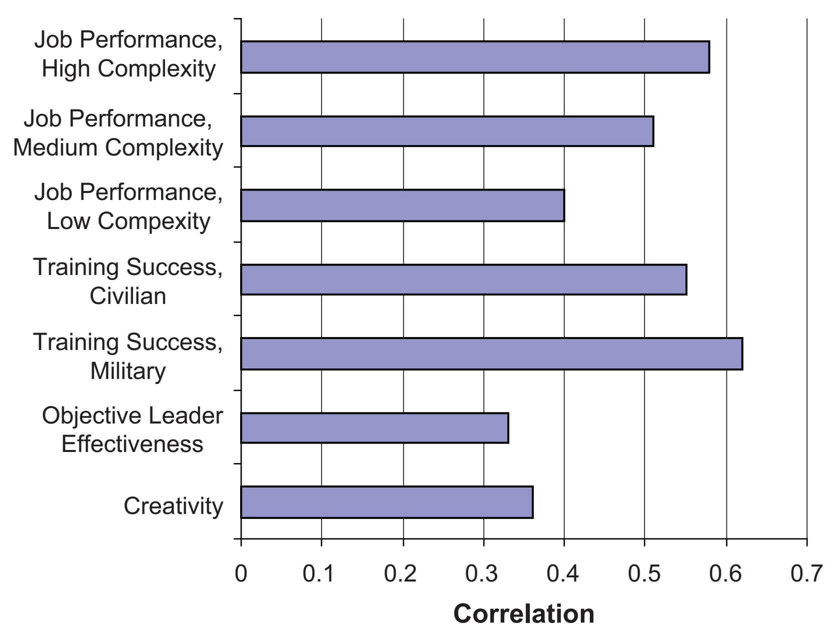

Fig. 2. Correlations between cognitive ability and measures of work performance. Coefficients are corrected for measurement error and restriction of range where possible. Results presented are from Ones, Viswesvaran and Dilchert (2005) and Kuncel, Hezlett, and Ones (2004).

This issue has been examined in a number of large-scale studies in both work and academic settings, and findings have led to a simple conclusion: More ability is associated with greater performance. College GPA is linearly related to SAT test scores across the entire range of test scores (Cullen, Hardison, & Sackett, 2004). Similarly, relationships between supervisors' ratings of employees' job performance were found to be linearly related to overall broad aptitude test scores in 174 studies on more than 36,000 workers (Coward & Sackett, 1990). Dramatically, these results hold even for individuals

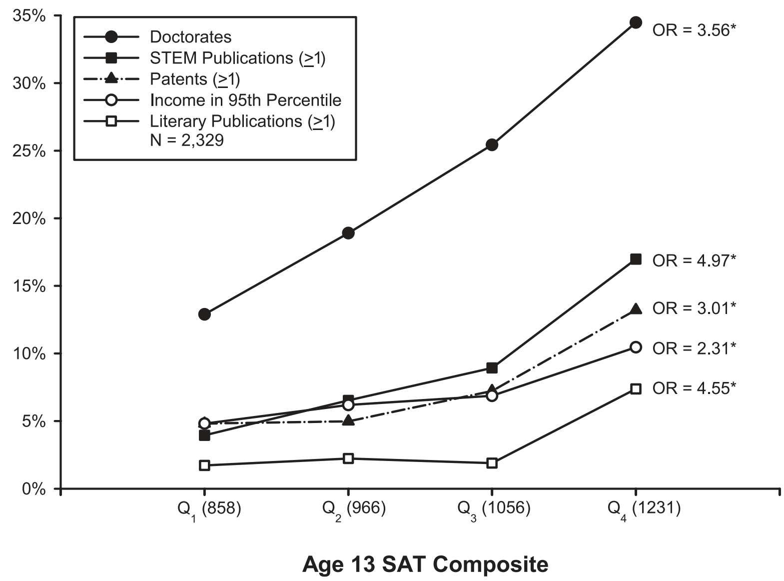

Fig. 3. Professional accomplishments across individual differences for the quartiles within the top 1% of general cognitive ability, 25+ years after identified at age 13. From Lubinski (2009).

in the highest cognitive ability range (Lubinski, 2009). Exceptional outcomes (doctoral-level degrees, scholarly publications, patents) are more frequent for those with higher scores *within the top 1%* of the ability distribution. Remarkably, those around the 99.13th percentile published less research and obtained fewer patents than those at the 99.88th percentile, even when controlling for type of institution and degrees earned (see Fig. 3). Note that to accurately examine this issue, measures must be without a ceiling, covering the full range of both cognitive ability *and* performance outcomes. For example, in predicting performance on the high jump, jumping ability ceases to be important beyond a certain point: if the bar is set only 12 inches off the ground!

Why Tests Predict

Although the predictive power of tests has been extensively demonstrated, the source of this predictive power has been a topic of debate. For example, concerns that test scores are merely a proxy for SES have recently been reignited by widely publicized but poorly understood data from the University of California. Overall, theory and research indicate that tests are valuable tools because an assessment of current skill and knowledge is predictive of what a person can do right now as well as how well a person is likely to learn and develop in the future.

Models of job performance

Cognitive ability is one of the major determinants of training outcomes (Colquitt, LePine, & Noe, 2000) and job performance (Campbell, McCloy, Oppler, & Sager, 1993). Meta-analytic research on training outcomes suggests that cognitive abilities influence knowledge and skill acquisition during training. In turn, employees who learn more during training do a better job of applying what they learn on the job and, ultimately, perform their jobs better (Colquitt et al., 2000). Similarly, models and research on job performance indicate that job knowledge and skills mediate the relationship between cognitive ability and job performance (McCloy, Campbell, & Cudeck, 1994; Schmidt, Hunter, & Outerbridge, 1986). Situational factors such as past experience, education, and training also affect current levels of job-related knowledge and skill. Based on this robust literature, it appears that scores on cognitive ability tests predict performance because they forecast the extent to which individuals both currently possess and will continue to acquire the specific knowledge and skills needed to perform effectively in educational and work settings (Ones et al., 2005).

Test scores and SES

An alternative scenario is that test scores and performance measures coincidentally correlate because they are both related to SES. One appropriate method of investigating this possibility is to use statistical techniques like multiple regression to examine whether test scores incrementally and substantially predict student performance after the contribution of SES to performance is accounted for. Recent results from the University of California system were stated to demonstrate that "after controlling for [SES] . . . the relationship between SAT I scores and UC grades virtually disappears" (Atkinson, 2005, p. 21). At first glance, this conclusion appears to be supported by the data (Model A in Table 2). The weight associated with the SAT I is virtually zero when SES measures are included in a regression predicting first-year college grade point average (GPA) from high school GPA and test scores. More detailed analysis of the data, however, reveals the correct interpretation of the results.

As shown in Table 2, the SAT I predicts first-year college GPA when SES variables are present and the SAT IIs and high school GPA are not (Model B). A comparison of additional models (C and D) shows that controlling for SES only slightly diminishes how well the SAT I predicts when it is used in combination with high school GPA. Therefore, the correct interpretation of the troublesome Model A is that a measure of verbal and quantitative skills (SAT I) does not contribute much to the prediction of grades beyond another measure largely focused on verbal and quantitative skills (SAT IIs). An extreme example of this effect would be using height in inches and height in centimeters in a prediction model. Therefore, the same data from the University of California show that test scores are not just a proxy for SES. They predict performance even after SES and high school GPA are taken into consideration.

Consistent results were obtained in a recent meta-analysis and large-scale analysis of primary data from multiple institutions that were directed toward evaluating different relationships among SES, test scores, and academic performance (Sackett, Kuncel, Arneson, Cooper, & Waters, 2009). As we would expect from previous results, SES was related to test scores to a modest degree. However, SES variables did not come close to eliminating the predictive power of tests, and tests provided incremental predictive information beyond grades. The modest relationship between test scores and SES is consistent with the idea that cognitive tests measure developed skills while SES effects are indirect. SES does not explain the relationship between test scores and subsequent performance.

Table 2. Predictive Power of Test Scores, High School Grades, and Socioeconomic Variables in Different Combinations (Regression Models) on 1st-Year College GPA in the University of California System

	Model A	Model B	Model C	Model D
Multiple-R	.48	.37	.46	.46
Standardized betas				
SAT I	.02	.34	.22	.25
High school GPA	.28		.30	.30
SAT II	.24			
Family income	.03	.01	.03	
Parental education	.06	.04	.05	

Note. Multiple-R is the estimate of the combined predictive power of all variables in the model and ranges from 0 to 1.0. Standardized betas are estimated weights for the set of variables included in that model predicting the outcome (1st-year gpa) when all variables are standardized to create weights in the same units (–1.0 to 1.0). Data from Geiser and Studley (2001).

Standardized Measures of Personality, Habits, and Attitudes

Cognitive abilities are not the sole determinants of performance in academic and work settings. The prediction of performance using standardized tests of cognitive ability can be incremented by adding measures of personality, values, interests, and habits to the admission or selection system, but only if they are carefully selected and developed. For example, the Big Five personality traits (Emotional Stability, Extroversion, Conscientiousness, Agreeableness, and Openness) demonstrate useful correlations with academic performance (Poropat, 2009) and job performance (Barrick & Mount, 1991). Conscientiousness and its facets are consistently connected with effective behaviors at school and work (corrected correlations with GPA of .20 and with overall job performance of .26). Measures that are developed explicitly for use in a particular setting can be even stronger predictors. For example, the corrected correlation between work-oriented integrity tests and job performance has been estimated to be .41 (Schmidt & Hunter, 1998). In the academic domain, measures designed to capture aspects of study habits, attitudes, and skills (which have moderate associations with personality traits) produce corrected correlations with grades earned of .40 or higher (Crede & Kuncel, 2007). This predictive power rivals that achieved by standardized tests of cognitive abilities. Furthermore, measures of personality, habits, and attitudes demonstrate low to zero correlations with cognitive ability and, therefore, produce useful incremental validity in predicting performance. Application of self-report measures of personality, habits, attitudes, or interests as predictors in high-stakes settings should be considered with some caution, given their greater susceptibility to coaching and faking than objective tests of cognitive abilities. However, other-report measures of personality (e.g., peers, friends) have recently been shown to be more strongly predictive of performance than are self-reports of personality and may, depending on the source, avoid some or all of the concerns about faking that arise from self-reports (Connelly & Ones, in press).

Predictive Bias

Research on the fairness of ability tests has drawn the conclusion that tests are not biased against women and minority groups (for reviews see Kuncel & Hezlett, 2007; Sackett et al., 2008). Given that, on average, some groups obtain lower scores, this conclusion can appear very confusing to those who are not familiar with the technical definition of bias. How can a test be unbiased *and* have group differences in test scores at the same time? The key to the problem is that predictive bias cannot be determined without assessing the nature of the relationship between test scores and subsequent performance across groups. This relationship is typically examined using moderated regression. A test is *not* biased if it accurately reflects a capability difference between groups and if the nature of the relationship between capability and performance is similar for all groups. Some examples help illustrate this concept.

Men are taller than women on average. Would using a metric ruler to measure applicants' height and then basing selection decisions on the results be biased? It depends. Height is largely unrelated to performance in many jobs and academic settings. Consequently, performance for women as a group would be underpredicted by height, regardless of the accuracy of the ruler, because women perform better on the job or at school than their lower scores (shorter average height) would indicate. Height, even if measured well, would be a biased predictor for admissions and most jobs. However, for predicting performance in a coed basketball league, we suspect it would not be biased. Even holding other gender differences constant (e.g., upper body strength), height is an important predictor of basketball performance and would be a legitimate consideration in selecting players. Of course, exceptional basketball playing skills are observed in both genders, and height is not the sole determinant of performance (a fact that argues for using multiple predictors in selecting players). However, height is important and its relationship with performance is likely to be similar for both genders. Teams comprised of both genders would consequently have more men on them, all else being equal, if height were used as a part of the selection process. This example illustrates that assessment of bias depends on the nature of a predictor's relationship with performance. Note that a particular ruler would be biased for selecting men and women for basketball teams if it failed to accurately reflect the height of both groups. For example, suppose a paper ruler was hung on the wall crookedly in a way that tended to add a few centimeters to most males' height and subtract a centimeter from most females' height. Even though height itself (the construct) remains an unbiased predictor of basketball performance, this particular ruler is biased.

Now consider a skill assessment that shows some large racial differences and has effects on occupational and other outcomes. For example, there are very large Black–White group differences in swimming skills, with white swimmers, on average, being more skilled. This difference is associated with large differences in employment in occupations requiring swimming skills, massive differential representation in the Olympics, and substantial differences in incidents of drowning (Mael, 1995). Swimming assessments are straightforward and typically test the ability to swim a certain distance. For a job requiring swimming skills, it is hard to argue that the assessment is biased. We would expect the skill test to be associated similarly with performance regardless of race. A person can either propel themselves in the water or not, and performance as a life guard or a member of a Coast Guard rescue team is dependent on being able to do so.

One could raise a legitimate societal concern that there are differential opportunities to learn swimming, as well as familial, social, environmental, peer, economic, and cultural factors that contribute to the difference. Indeed research has found support for some of these factors (Mael, 1995). However, this is not bias in the measure. The key points are that the swimming test is not the source of the difference and is a measuring a legitimate skill and a predictor of subsequent performance. We want life guards who can swim well. Addressing the skill disparity is a societal issue. Condemning the swimming test will not correct the societal issue any more than discarding a thermometer will make a fever go away.

Similarly, differences in cognitive ability test scores reflect some of the same broad effects that have been correlated with familial, social, environmental, peer, community, and economic influences (e.g., Phillips, Brooks-Gunn, Duncan, Kelbanov, & Crane, 1998). Bias is concerned with evaluating whether scores capture skill differences that have similar relationships to performance for all groups. The evidence suggests that average test score differences reflect developed-skill differences that are relevant for performance on the job or in school, not inherent biases in the tests.

Conclusion

Given the impact of high-stakes testing on individuals and society, it is important for scientists, citizens, and policymakers to critically examine standardized tests of cognitive ability and fully inform themselves about the scientific evidence on these selection tools. The vast body of accumulated knowledge about these tests is clear: They are among the strongest and most consistent predictors of performance across academic and work settings. Although the evidence indicates that the group differences reflected by standardized cognitive tests are not caused by the tests themselves, we need to decide how to address the causes of group differences and wrestle with their consequences. We should continue to strive to further understand the nature and development of cognitive abilities and seek additional assessments that supplement cognitive ability test scores to improve decision-making accuracy.

Recommended Readings

Deary, I.J. (2001). *Intelligence: A very short introduction*. New York: Oxford University Press. A clearly written general introduction to the nature of cognitive abilities and their relationship with life outcomes.

Kuncel, N.R., Hezlett, S.A., & Ones, D.S. (2004). (See References). A study and review of evidence that argues against the idea that book smarts are different from street smarts.

Declaration of Conflicting Interests

The authors declared that they had no conflicts of interest with respect to their authorship or the publication of this article.

References

Atkinson, R.C. (2005). College admissions and the SAT: A personal perspective. *APS Observer*, *18*(5), 15–22.

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

Berry, C.M., & Sackett, P.R. (2009). Individual differences in course choice result in underestimation of the validity of college admissions systems. *Psychological Science*, *20*, 822–830.

Campbell, J.P., McCloy, R.A., Oppler, S.H., & Sager, C.E. (1993). A theory of performance. In N. Schmitt & W.C. Borman (Eds.), *Personnel Selection in Organizations* (pp. 35–70). San Francisco: Jossey-Bass.

Colquitt, J.A., LePine, J.A., & Noe, R.A. (2000). Toward an integrative theory of training motivation: A meta-analytic path analysis of 20 years of research. *Journal of Applied Psychology*, *85*, 678–707.

Connelly, B.S., & Ones, D.S. (in press). An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity. *Psychological Bulletin*.

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

Crede, M., & Kuncel, N.R. (2007). Study habits, skills, and attitudes: The third pillar supporting collegiate academic performance. *Perspectives on Psychological Science*, *3*, 425–453.

Cullen, M.J., Hardison, C.M., & Sackett, P.R. (2004). Using SAT-grade and ability-job performance relationships to test predictions derived from stereotype threat theory. *Journal of Applied Psychology*, *89*, 220–230.

Deary, I.J., Whalley, L.J., Lemmon, H., Crawford, J.R., & Starr, J.M. (2000). The stability of individual differences in mental ability from childhood to old age: Follow-up of the 1932 Scottish Mental Survey. *Intelligence*, *28*, 49–55.

Geiser, S., and Studley, R. (2001). Supporting documents for "UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California." Oakland, CA: University of California Office of the President, Division of Academic Affairs. Downloaded August 21, 2009, from http://www.ucop.edu/sas/research/researchandplanning/supporting.htm

Kuncel, N.R., & Hezlett, S.A. (2007). Standardized tests predict graduate student's success. *Science*, *315*, 1080–1081.

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.

Lubinski, D. (2009). Exceptional cognitive ability: The phenotype. *Behavioral Genetics*, *39*, 350–358.

Mael, F.A. (1995). Staying afloat: Within-group swimming proficiency for Whites and Blacks. *Journal of Applied Psychology*, *80*, 479–490.

McCloy, R.A., Campbell, J.P., & Cudeck, R. (1994). A confirmatory test of a model of performance determinants. *Journal of Applied Psychology*, *78*, 493–505.

Ones, D.S., Viswesvaran, C., & Dilchert, S. (2005). Cognitive ability in personnel selection decisions. In A. Evers, N. Anderson, & O. Voskuijl (Eds.), *The Blackwell handbook of personnel selection.* Oxford, UK: Blackwell Publishing.

Phillips, M., Brooks-Gunn, J., Duncan, G.J., Kelbanov, P., & Crane, J. (1998). Family background, parenting practices, and the Black-White test score gap. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap.* Washington, DC: Brookings Institution Press.

Poropat, A.E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological Bulletin*, *135*, 322–338.

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

Sackett, P.R., Kuncel, N.R., Arneson, J., Cooper, S.R., & Waters, S. (2009). Socio-economic status and the relationship between admissions tests and post-secondary academic performance. *Psychological Bulletin*, *135*, 1–22.

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.

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.