Comprehensive Handbook of Multicultural School Psychology

Edited by

Craig L. Frisby
Cecil R. Reynolds

WILEY

John Wiley & Sons, Inc.

2005

CHAPTER 18

Implications of Cognitive Differences for Schooling Within Diverse Societies

LINDA S. GOTTFREDSON

THE ACHIEVEMENT GAP is the focus of much public concern today. Indeed, No Child Left Behind, a federal law passed in 2001, requires the public schools to eliminate all racial/ethnic differences in tested achievement by 2014, at which time all students are to perform to a high level of proficiency (Banchero & Little, 2002). The title of a recent book on raising minority achievement—*No Excuses* (Thernstrom & Thernstrom, 2003)—captures the spirit of the law, and schools will be punished if they fail to make adequate yearly progress in satisfying that law.

Forty years ago the federal government set out on a similar mission—to raise low educational achievement levels by raising low IQs in at-risk populations. Enlightened people of the time spoke about Black-White differences in IQ because preschool interventions were thought to be sure-fire means for eliminating them and the unequal school and life outcomes they cause. High expectations for the new interventions were met by clear failure (Brody, 1992, chap. 6; Rowe, 1997), followed by an effort to discredit IQ tests and the very notion of intelligence. The enlightened are now obliged to deny the existence of any real, important, and enduring racial differences in cognitive ability.

The science of mental abilities has advanced dramatically since the 1960s. Although it has not yet pinned down a definitive answer to what causes racial disparities in cognitive abilities, it has confirmed that they are, in fact, real, important, and enduring. As such, they also create enormous dilemmas for schools. They put school psychologists, like teachers, on the front lines of an intense political battle against human difference itself. This chapter examines the empirical roots of the struggle and its consequences for schools.

The chapter begins by describing common attempts to preempt discussion of group differences in cognitive ability by denying the facts that confirm them. As a prelude to describing group differences, key research findings on individual differences in cognitive ability are summarized. As shown, evidence from diverse studies—psychometric, biological, and sociological—dovetails at both the phenotypic and genotypic levels to create a coherent picture of how powerfully general intelligence influences our lives. The chapter then turns to differences in the distribution of phenotypic cognitive abilities by racial/ethnic group, both in the United States and worldwide. (Potential genotypic differences have been difficult to study for both technical and political reasons.) It focuses in particular on how large the group differences are and whether they are measured accurately, have changed over time, and can account for the large racial/ethnic gaps in educational achievement. The picture, once again, is of the coherence in results over different tests, times, and cultures. The chapter ends by showing how these data help explain the difficulties schools have in meeting the egalitarian goals set for them, and why their efforts often backfire in predictable ways.

THE PURSUIT OF DOUBT AND DIVERSION

Each advance in knowledge on group differences is met with efforts to sow doubt about its scientific merits or to divert attention from the interlocking pattern of results to which it contributes. Such efforts often involve isolating a finding from the full pattern of knowledge while keeping the pattern offstage. They serve to discount group differences in cognitive ability by promoting belief in one or more of seven falsehoods: (a) Intelligence does not exist; (b) even if intelligence does exist, we cannot measure it; (c) even if we can measure intelligence, we cannot measure it fairly; (d) even if we can measure intelligence fairly, it is not that important anyway; (e) even if intelligence is important, there are multiple independent forms of intelligence that are just as important; (f) even if single and important, intelligence is mostly just the product of social privilege rather than something that is substantially heritable; and (g) even if it is highly heritable, intelligence level is still quite malleable.

To illustrate, some purveyors of doubt would have us remain agnostic about the very existence of intelligence until everyone agrees on its definition (Armour-Thomas & GoPaul-McNicol, 1998; Valencia & Suzuki, 2001). Others assert that because different cultures describe intelligence somewhat differently, it must be a cultural artifact—nothing more than what a society chooses to value (Armour-Thomas & GoPaul-McNicol; Sternberg & Grigorenko, 2001). Others advance the irrelevant truth that intellectual life is too complex, and abilities too varied and numerous, to be captured by a single number (Armour-Thomas & GoPaul-McNicol), which ultimately promotes nihilism. Despite voluminous evidence that mental tests are not biased against American Blacks or other native English speakers, others continue to assert that the question of test bias is still an "open issue" (Valencia & Suzuki). A recently popularized claim is that we do not know whether different races perceive test items in the same way (Helms, 1992, on cultural equivalence), despite long-standing evidence that the superficial characteristics of items do not matter as long as test takers are fluent in the language (see Jensen, 1980, 1998, on the indifference of the indicator).

The following sections present evidence, some of it new, which contradicts all seven falsehoods. Great puzzles remain (e.g., the secular increases in IQ scores), to be sure, and work has barely begun on some of the biggest questions (e.g., finding the genes for intelligence). But contemporary efforts to promote confusion by sowing seeds of doubt and diversion only stall more effective coping with the ability differences that roil society today.

WHICH COGNITIVE ABILITIES ARE MOST FUNDAMENTAL?

Research has turned our understanding of mental abilities literally upside-down in recent decades. Illustrative of its time, a 1973 review of group differences in IQ (Dreger,

1973) asserted that the "gross IQ" is just the average of the distinct intellectual functions that an IQ battery's various subtests measure and is thus not a particularly informative or tractable measure of intellectual functioning. Researchers have since learned, however, that the way IQ tests are built is no guide to what they actually measure or to how abilities themselves are constructed. As shown in the following pages, the most specific abilities are actually the most psychometrically complex and general intelligence is the most unitary.

There Are Many Abilities, but Only One General Mental Ability Factor (*g*)

Psychometricians have spent a century charting the relations among—the "structure of"—abilities, and concluded that mental abilities are most usefully distinguished according to their generality, that is, the range of tasks on which they enhance performance. Carroll's (1993) monumental reanalysis of data from hundreds of prior factor analyses crystallized this new consensus.

Figure 18.1 shows Carroll's three-stratum hierarchical model of human cognitive abilities. The many dozens of narrow Stratum I abilities include, for example, *Reading Decoding (RD), Closure Speed (CS), Ideational Fluency (FI),* and *Memory for Sound Patterns (UM).* Stratum I abilities all intercorrelate to various degrees and, when factor analyzed, yield the familiar group factors of ability, including verbal ability, spatial aptitude, and short-term memory. These constitute the broad Stratum II ability factors, of which Carroll identified eight.

The Stratum II factors also correlate among themselves (when oblique rotation is allowed), and in turn yield a more general third-order factor. Carroll confirmed that only one factor emerges at the Stratum III level. It is called *g*, short for the general mental ability factor, and it accounts for a third to half of the variance in scores on any broad battery of mental tests. Stratum II abilities are so highly correlated with *g* that Carroll describes them as different flavors of *g*. The Stratum II ability called fluid *g* (reasoning) often cannot be distinguished from *g* itself (Gustafsson, 1988). *g* corre-

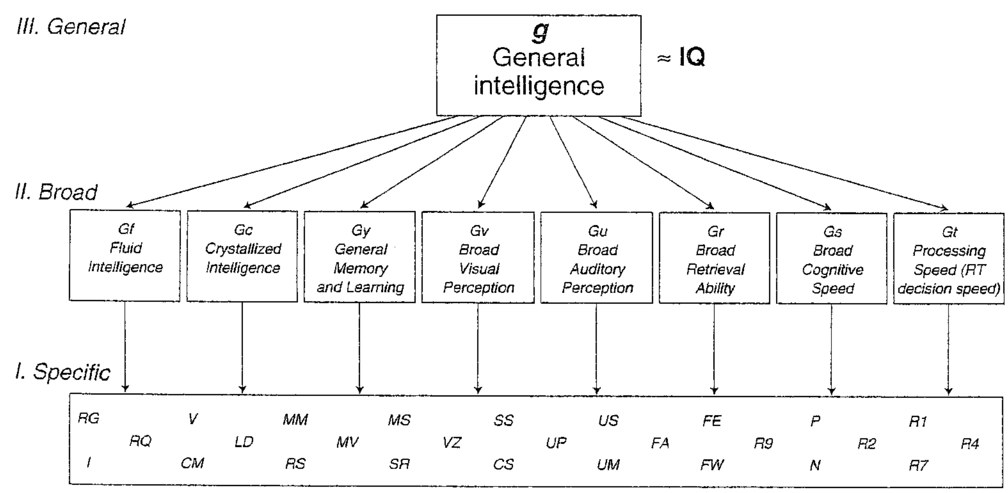


Figure 18.1 Hierarchical model of human abilities.
Source: Adapted from Carroll (1993, p. 626) and reprinted with permission of Cambridge University Press

sponds well with what most people think of as intelligence, and many researchers have adopted it as their working definition of intelligence.

g Is the Major Building Block of All Mental Abilities

A key insight embodied in the hierarchical model is that mental abilities are built top-down rather than bottom-up, as once thought. That is, rather than the broad abilities being aggregates of many specific ones, the former constitute the core of the more specific abilities. In fact, *g* is the major component of all tests, whatever their manifest content, format, or intended purpose. No one has ever succeeded in developing a useful test of a Stratum I or II ability that does not also measure mostly *g*. Thus, higher stratum abilities are not just the simplest psychometrically but also the most fundamental.

The Hierarchical Model Incorporates Theories of Multiple Intelligences

By providing a unified model of cognitive abilities (Deary, 2000), the hierarchical model reconciles many seemingly conflicting theories of intelligence. For instance, Howard Gardner's (1983) theory of multiple intelligences posits eight or so independent intelligences. Carroll (1993, p. 641) points out, however, that four of Gardner's yet-unmeasured intelligences (linguistic, logical-mathematical, spatial, and musical) resemble abilities at the Stratum II level (*Crystallized Intelligence, Fluid Intelligence, Visual Perception*, and parts of *Auditory Perception*). Gardner has ignored the Stratum III level altogether and, in several cases, applied the label *intelligence* to traits mostly outside the mental realm. His interpersonal, intrapersonal, and bodily-kinesthetic intelligences embody noncognitive traits long studied under the rubrics of personality, emotion, and psychomotor ability. They are important traits but not, as the label *intelligence* connotes, comparable to—or substitutable for—*g* as alternative modes of higher level thinking.

Sternberg's (1997) triarchic theory proposes three independent intelligences, all presumably at the Stratum III level: analytical (*g*), creative, and practical. When data from Sternberg's Triarchic Abilities Test are properly analyzed, however, its three scales collapse into one general factor (Brody, 2003). The evidence said to support Sternberg's single most researched triarchic ability, practical intelligence, is either irrelevant or misleading (Gottfredson, 2003).

HOW WELL DO MENTAL TESTS MEASURE *g*?

The ability to separate *g* from the vehicles used to measure it has been tremendously important because *g* provides an independent external criterion against which to compare and validate mental tests (Jensen, 1998, chap. 10). By the same token, it frees the study of general intelligence from the particular tests used to measure it.

All Tests Converge on the Same True *g*

Every broad test battery yields a strong *g*, but that does not necessarily mean they all yield the same *g*, especially when extracted with different factoring procedures or from different test batteries. The similarity of any two such factors can be assessed by calculating a coefficient of congruence between the factor loadings, for the same subtests, on the two separately derived *g* factors. The coefficient can range from −1.0 to 1.0, where .90 is high factor similarity and .95 is practical identity.

Although some factor-extraction methods are more theoretically appropriate than others, all yield virtually indistinguishable *g* factors when the battery taps a wide variety of abilities and when samples are representative (Jensen, 1998, pp. 81–83; Jensen & Weng, 1994; Ree & Earles, 1991). Less research has examined the similarity of *g* factors derived from different test batteries, but here, too, the evidence is for high similarity (Thorndike, 1986). It is clear, however, that a *g* factor can be contaminated, skewed, or "flavored" by specific ability factors (Carroll, 1993, p. 596) when it is extracted from a battery containing many tests of one type (e.g., quantitative reasoning) but few or none of others (e.g., language). Purer *g*s are derived when batteries sample abilities more broadly and evenly so that the specific factors cancel each other out.

All Populations Converge on the Same True *g*

Essentially the same *g* factor is obtained from diverse races and cultures in North America, Europe, Asia, and Africa (see reviews in Jensen, 1998, pp. 87–88; Rushton & Jensen, 2005). For instance, congruence was over .99 when the *g* factor derived from Japanese samples taking a Japanese translation of the Weschler was compared to the *g* factor obtained from the test's (American) standardization sample. It was likewise high (averaging .995) for Blacks and Whites in 17 studies (Jensen, 1998, p. 375), one of 3-year-olds (where the coefficient was .98; Peoples, Fagan, & Drotar, 1995, p. 76). Most recently, Jensen (2003) found coefficients of .94 to .99 for Black-White comparisons in each of Grades 3 through 8. The fact that *g* factors derived from different test batteries, populations, and procedures converge on a common "true" *g* means that *g* constitutes a readily available, common yardstick for validly comparing diverse populations.

Individual IQ Tests Are Very Good, But Imperfect, Measures of *g*

The overall score from any particular IQ test battery (e.g., the Wechsler Intelligence Scale for Children–Fourth Edition's [WISC-IV's] Full-Scale IQ [FSIQ], or the Woodcock-Johnson III's [WJ III's] General Intelligence Ability [GIA] provides a good measure of *g*. A battery's composite IQ score usually correlates about .95 with its own *g* factor (formula in Jensen, 1998, pp. 103–104) and probably at least .8 to .9 with true *g* (pp. 90–91). IQ scores are not, however, pure measures of *g* because they always capture some non-*g* components of ability from Stratum I and Stratum II, the mix of which can differ from one IQ test battery to another. The impurities are too small to affect the practical utility of well-validated batteries in applied settings, but, when ignored, they can muddy results from research exploring the nature of *g*.

Mental Tests Differ in How Well They Measure *g* (Their *g*-Loadedness)

Although *g* is the major component—the spine—of all mental tests, tests vary considerably in their *g*-loadedness. This can be seen, for example, in the subtests of the Wechsler series of IQ test batteries (Sattler, 2001, pp. 238, 342, 389). Vocabulary, Information, Similarities, Arithmetic Reasoning, Block Design, and Comprehension tend to be the best measures of *g* (median correlation of .76 with their battery's *g* factor); Object Assembly, Picture Completion, Picture Arrangement, and Symbol Search are somewhat weaker measures of *g* (median .65); and Digit Span, Mazes, and Coding are weaker yet (median .54). The ability to assess the *g* loading of tests and tasks is essential, as will be seen, for understanding why some tests predict group differences and life outcomes better than others.

IQ Test Batteries Measure the Same Ability Constructs in All Races

As noted, research has shown cross-group comparability in *g* factors. In the last decade researchers have turned to confirmatory factor analysis (CFA) for assessing more fully and precisely the construct validity of IQ batteries (Bickley, Keith, & Wolfle, 1995; Keith, 1997). The increasing use of hierarchical, multisample CFA (MCFA) has been especially valuable because it provides a more direct, systematic, and statistically rigorous way of determining whether a test battery yields the same general and specific ability factors (and the same factor loadings on them) for different ages, races, and sexes. Reinforcing earlier research on test bias, MCFA research has consistently found that the IQ test batteries studied thus far measure the same constructs equally well in different races, and at all ages examined.[1] Group differences in scores on the major test batteries therefore represent real differences in the underlying abilities they measure (Jensen, 1980; Neisser et al., 1996; Reynolds, Lowe, & Saenz, 1999).

Cognitive Tests Predict Achievement Equally Well for All Racial/Ethnic Groups

Tests of cognitive ability predict major life outcomes (e.g., standardized academic achievement, job performance, and status attainment) equally well for all racial/ethnic groups studied sufficiently to make a determination (e.g., Jencks et al., 1979; Jensen, 1980; Schmidt, 1988). This fact is consistent with research concluding there are no race-specific (e.g., Black-only) influences on academic development in childhood (Jensen, 1998, pp. 557–559), such as culturally distinct learning styles (Frisby, 1993). The same factors that account for achievement differences between siblings account for the differences between races (Rowe, Vazsonyi, & Flannery, 1995).

WHAT IS *g*?

Intelligence researchers no longer debate which abilities should be labeled *intelligence* (e.g., *g* alone, the entire three-stratum cognitive structure, or the full panoply of human competencies). Instead, they are working to understand the empirical phenomenon they have isolated, *g*, by pursuing it down into the deepest recesses of the brain and out to the furthest reaches of social life.

g Is a Biological Phenomenon, Not a Statistical Artifact

g is probably not an ability per se, but a property of the brain—perhaps more efficient information processing owing to greater speed, accuracy, number, and connectedness of neurons. Many aspects of brain structure and function are correlated

[1] Results are the same for both Blacks and Whites across three age groups between ages 7 and 12.5 on the Kaufman Assessment Battery for Children (K-ABC; Fan, Willson, & Reynolds, 1995; Keith et al., 1995); Blacks, Whites, and Hispanics across four age groups between ages 2 and 17 on the DAS (Keith, Quirk, Shartzer, & Elliott, 1999); Blacks and Whites in the standardization sample of the WISC-III (Kush et al., 2001); and a screening test at Grades K through 1, for Blacks, Whites, and Hispanics (Tu, Scott, Mason, & Urbano, 1999). The structure of test-session behavior is likewise the same for White, Black, and Hispanic children in two age groups between ages 6 and 16 in the linking sample for the Wechsler Intelligence Scale for Children–Third Edition (WISC-III) and the Wechsler Individual Achievement Test (WIAT; Konold, Glutting, Oakland, & O'Donnell, 1995). See Caretta and Ree (1995) for adults.

with *g*, at both the phenotypic and genotypic levels (Gray & Thompson, 2004; Jensen, 1998). The heritability of individual differences in IQ increases from 40% in the early elementary grades to 80% by middle age (Plomin, DeFries, McClearn, & McGuffin, 2001). High within-race heritabilities lend plausibility, but not proof, to the hypothesis that group mean differences in IQ are also substantially heritable (Rushton & Jensen, 2005). The heritability of Stratum II abilities can be traced almost entirely to the genes they share with *g* (Plomin et al., chap. 10). *g* is more heritable and less malleable than the more specific abilities. Very specific skills can often be readily taught, perhaps precisely because they are so narrow in application, but more broadly applicable (more generalizable or transferable) abilities are more resistant to intervention.

g Is a Highly General Ability to Process Complex Information

The *g* factor is measured well by tests that differ dramatically in purpose, content, and format, which means that *g* must represent some highly generic capacity for processing information of virtually any type. It is not the mere accumulation of bits of knowledge, but a capacity that facilitates learning and applying any body of knowledge. In school settings, it may be conceived as the ability to profit from instruction, especially from incomplete instruction (Snow, 1996). IQ tests are sometimes dismissed as measuring only a narrow academic ability of little practical import (e.g., Steinberg, 2002, pp. 76–77; Sternberg, 1997, p. 11) because their items tend to pose tasks that are esoteric (matrices) or seemingly academic in nature (reading and calculating). Such content does not represent the essence of *g*, but only the sorts of psychometric tasks that measure it with greatest reliability and validity (Gottfredson, 2002).

WHAT DOES *g* PREDICT BEST, AND WHY?

The "specificity doctrine" (Jensen, 1984) reigned until dethroned about 20 years ago by validity-generalization research (Schmidt, 1988). That doctrine held that performance on different school and work tasks was best predicted by different mixes of abilities, which might shift according to even slight changes in the performance setting. General intelligence was accordingly assumed to have only limited and unpredictable value for explaining differences in human performance. *g*'s predictive value is actually, like *g* itself, highly general and it rises and falls in tandem with the complexity of life's tasks.

The Ability of Mental Tests to Predict Performance Inside and Outside Schools Rests Almost Exclusively on Their Large *g* Components

The better a test measures *g*, the better it predicts performance in school or work. Large studies in both the United States and Europe find that specific cognitive abilities add little or nothing to their prediction beyond that contributed by *g* (e.g., Jencks et al., 1979, chap. 4; Ree, Earles, & Teachout, 1994; Salgado, Anderson, Moscoso, Bertua, & de Fruyt, 2003; Schmidt & Hunter, 1998; Thorndike, 1986). For instance, specific academic abilities (e.g., mathematical) predict performance in all subjects (e.g., reading, social studies, math) about equally well, and predict it better the more *g*-loaded they are (Jencks et al., chap. 4). Reschly (1997) also argues that only the general factor of IQ batteries has demonstrated treatment validity.

Like Mental Tests, Life Tasks Differ in Their Demands for *g* (*g*-Loadedness)

The active ingredient in life tasks is the same as in IQ tests: the complexity of the information processing the task demands for learning or performing it well (Gottfredson, 1997). Task complexity includes the amount, abstractness, unpredictability, and ambiguity of information to be processed, the degree of inference it requires, and its embeddedness in distracting information. *g* correlates most with learning proficiency when the material to be learned is hierarchical (builds on prior learning), is meaningful, and requires insight rather than rote memorization (e.g., algebra vs. arithmetic, vocabulary vs. spelling; Jensen, 1998, pp. 320–328).

g Is a Strong Predictor of Standardized Academic Achievement

The median correlations between IQ and school achievement hover around .6 for standardized tests in the three Rs (see Table 18.1). The IQ-achievement correlation goes up to .8, however, when different forms of achievement are aggregated into a single composite (Jensen, 1998, pp. 323–324). This value rivals the median correlation observed among the major individually administered IQ test batteries (.85) and exceeds the correlation between those batteries' composite IQs and their own more academic subtests (.76; see Table 18.1). IQ-achievement correlations are smaller at higher levels of education, but this is just an artifact of more restriction in range on IQ in the higher grades (e.g., college vs. elementary students; Jensen, 1980, p. 319).

g Is the Best Single Predictor of Many Life Outcomes

The *g* factor correlates more strongly than any other single trait or circumstance, including socioeconomic status (SES), with a wide range of socioeconomic successes (e.g., level of education, occupation, and income) and failures (e.g., chronic welfare use, incarceration, and bearing illegitimate children; see review in Gottfredson, 2002). Grades in elementary and secondary school, years of education, and job level attained by midlife all correlate moderately highly with IQ measured in childhood and adolescence (.5–.7; see especially Jencks et al., 1979; also Jensen, 1980, chap. 8; Lynn & Vanhanen, 2002). These correlations are comparable to those discussed earlier between IQ and standardized achievement in particular school subjects. Other life outcomes correlate less well with IQ (e.g., .3–.4 for income in midlife; Jencks et al., chap. 4) or very little (.2 for law-abidingness; Gordon, 1997, p. 219). Attributes that correlate more highly with *g* and thus serve as better surrogates for *g* (e.g., years of education is better than income level) also predict other life outcomes better, especially when their demands are more complex (Hunter, 1986; Jencks et al.; see also Gottfredson, 2004, on predicting health).

IQ's Predictive Value Differs Systematically According to Both Task Complexity and How Tightly Rewards Are Tied to Good Task Performance

Some life tasks (e.g., education) require more complex learning, reasoning, and problem solving than others (e.g., law-abidingness). That is, some are more *g*-loaded. Job performance provides a good example, because its correlation with mental ability rises from about .2 in the simplest, most routine jobs to almost .8 in the most complex, self-directed, and critical occupations (Gottfredson, 1997; also Salgado, Anderson, Moscoso, Bertua,

Table 18.1

Maximum and Minimum Achievement Gaps (d_{ach}) Predicted for Blacks, Hispanics, and Asians, Relative to Whites

Outcome Being Correlated With IQ	Outcome's Correlation With: IQ[a] Range (1)	Median (2)	g[b] (Est.) (3)	Column (3) Is Corrected for Unreliability[c] (4)	Minimum Exp. d_{ach} If d_{IQ} Equals[c]: 1.20 Black (5)	.90 Hispanic (6)	|.30|[d] Asian (7)
Composite IQ from one of 5 major IQ test batteries							
Correlated with each other's composite IQ[e]	.61–.93	.85	.92	.96	**1.15**	**.86**	**–.29**
Correlated with scores on own most academic subtests[f]	.68–.83	.76	.83	.87	**1.04**	**.78**	**|.26|**
Standardized academic achievement in specific subjects, correlated with composite IQ from 5 major IQ batteries[g]							
Math/Arithmetic	.32–.81	.63	.67	.70	**.84**	**.63**	**–.21**
Reading	.30–.76	.61	.66	.69	**.82**	**.62**	**.21**
Language	.48–.68	.57	.62	.64	**.77**	**.58**	**.21**
Writing	.47–.68	.56	.61	.63	**.76**	**.57**	**.19**
Spelling	.15–.63	.42	.46	.48	**.57**	**.43**	**.14**

Note. Based on typical correlations between IQ and achievement and mean differences in IQ of d_{IQ} = 1.20, .90, and –.30. Maximum predicted d_{ach} are in headings of columns (5)–(7), and minimum predicted are entered in those columns. Math/Arithmetic, Reading, Language, and Writing = the three Rs.

[a] IQ = Full-scale IQ or equivalent unless specified otherwise.

[b] Entry in first row is the square root of data in column (2). Entries for other rows are data in column (2) multiplied by .92, the estimated *g*-loading of the 5 batteries' FSIQ.

[c] Used mean reliabilities from data in Sattler (2001): .96 for 5 major test batteries, .85 for their verbal subtests, and .92 for achievement tests, with data coming from Sattler (pp. 225, 338, 379, 380, 461, and 512 for IQ tests; pp. 583, 586, and 591 for achievement tests). Calculated disattentuated correlations using formula in Jensen (1980, p. 514), setting the condition that intercept differences are owing entirely to unreliability. This is the same condition required for estimating the *minimum* expected *g*-determined d_{ach}. The minimum expected d_{ach} in columns (5)–(7) were obtained by multiplying the disattentuated correlation in column (4) by the d_{IQ} in the heading of those columns, 1.20, .90, or .30. These d_{IQ} provide the *maximum* predicted d_{ach} for each outcome. The maximum effect sizes for achievement, d_{ach}, listed for Blacks (1.20) and Hispanics (.90) are based on the d_{IQ} documented for Black and Hispanic 18- to 23-year-olds in the largest recent national study, namely, the 1980 Armed Forces Vocational Aptitude Battery (ASVAB) standardization sample (see the entry in Table 18.2 for Armed Forces Aptitude Batteries). Asians might be represented by a d_{IQ} of |.30|, with a d_{IQ} –.30 for verbal ability and .30 for nonverbal ability.

[d] East Asians tend to score about .3 *SD* above Whites in IQ. This entry is listed as an absolute value, however, because East Asians tend to score below Whites on verbal tests. In addition, different subgroups of Asians range from above average to below average in mean IQ.

[e] Data from Sattler (2001, pp. 229, 384, 462, 514–515). Batteries include the Differential Abilities Scales (DAS), Stanford-Binet Intelligence Scale, Fourth Edition (SB-IV), Wechsler Preschool and Primary Scale of Intelligence (WPPSI), Wechsler Intelligence Scale for Children–Third Edition (WISC-III), and Wechsler Adult Intelligence Scale–Revised (WAIS-R).

[f] Data taken from Sattler (2001, pp. 238, 342, 389, 466, 517) for the vocabulary, information, similarities, arithmetic/quantitative, and comprehension subtests. These represent the major verbal subtests in the Wechsler series.

[g] Data from Sattler (2001, pp. 229, 384, 462, 514–515), and the achievement tests are the Wechsler Individual Achievement Test (WIAT), Wide Range Achievement Test–Revised (WRAT-R), Wide Range Achievement Test–Third Edition (WRAT-III), and Woodcock-Johnson–Revised (WJ-R). Data published mostly in the 1990s and late 1980s.

de Fruyt, & Rolland, 2003). IQ's predictive value can also vary because other (non-*g*) traits and circumstances of individuals affect certain life outcomes more than others. For instance, conscientiousness has no discernible effect on IQ test performance, but it can greatly affect the grade point averages (GPAs) that students accumulate over their educational careers, just as it enhances long-term performance in all jobs (Schmidt & Hunter, 1998). Correlations with IQ are also typically .1 to .2 lower when performance is measured subjectively rather than objectively. Not only are teachers', trainers' and employers' ratings of performance less reliable psychometrically, but they are also based less exclusively on actual performance and are instead swayed by the personality, deportment, and other non-performance-related traits of ratees. Likewise, although higher *g* enhances one's competitiveness for gifted education, admission to college, and a job, actually being selected depends on the institution's definition and measurement of merit, both of which are often hotly debated and can change over time.

HOW DO GROUPS DIFFER IN COGNITIVE ABILITY?

This chapter focuses on racial/ethnic differences in the single Stratum III ability, *g*, because it is better researched and far more important than the specific abilities for understanding group differences in achievement. Most research in the United States has compared Blacks and Whites, so that is necessarily the focus here. There are fewer data for Hispanic, Asian, and Native Americans, and they are more ambiguous for Hispanics and Asians because higher proportions of these groups are nonnative speakers of English.

GROUP DIFFERENCES IN THE DISTRIBUTION OF IQ ARE THE RULE, NOT THE EXCEPTION, WORLDWIDE

Virtually all individuals fall within a range of eight *SD*s (IQs of 40–160). Every racial/ethnic group's IQ distribution spans this range and is approximately normal, but each has a small excess at the lower tail owing to various genetic anomalies and environmental insults. Groups tend to differ somewhat in variability in IQ. Those that are more variable (have larger *SD*s) have flatter and more spread-out IQ bell curves, whereas groups with smaller *SD*s have more peaked bell curves because members are bunched closer around their group's average. The differences in variability generally are not large, for example, the *SD* being 13.03 IQ points for American Blacks and 14.67 for Whites in the Wechsler Adult Intelligence Scale–Revised (WAIS-R) standardization sample (Reynolds, Chastain, Kaufman, & McLean, 1987, p. 330). Small differences in variation can, however, have meaningful effects on the relative representation of groups at the tails of the IQ distribution.

The more consequential difference for most purposes is that the IQ bell curves for different groups tend to be centered at different points along the IQ continuum. Studies in the United States and other developed nations converge on mean IQs of roughly 85, 100, and 106 for Blacks, Whites, and East Asians, respectively. In sub-Saharan Africa, Black Africans, Coloreds (mixed-race individuals), and Whites average about 70, 85, and 100, respectively (Rushton & Jensen, 2005). American Blacks are also mixed-race individuals, the mean degree of admixture with European Whites being about 25% (Reed, 1971). Hispanic and Native Americans average around IQ 90 or a bit higher, which is roughly the median for countries worldwide (Lynn & Vanhanen, 2002, pp. 60–62). Subgroups within the White/European, Black, Hispanic, Asian, and indigenous categories also differ among themselves in average IQ, depending on ancestral origin. For example, Mexican Americans score higher on the average than do Puerto Ricans. Racial/ethnic differences in the distribution of IQ are surprisingly uniform regardless of host nation or culture (Rushton & Jensen, 2005).

Group Differences in IQ Represent Group Differences in *g* Itself

Several types of research converge on the conclusion that the IQ gaps between groups represent gaps in *g* itself. The *Spearman hypothesis,* named after its originator (Spearman, 1927), holds that if racial differences in IQ reflect primarily differences in *g*, then the races in question should differ most, on the average, on mental tests that are most *g* loaded. The hypothesis has been supported in over 20 independent studies from age 3 through middle age and including all major racial/ethnic groups in the United States and South Africa as well as various immigrant groups in the Netherlands (Jensen, 1998, chap. 11; Jensen, 2003; Rushton & Jensen, 2005). The relation between tests' *g* loadings and the size of Black-White differences on them can be illustrated with the WISC-R standardization sample. *g* loadings are high for Comprehension (.7) but low for Coding (.4), as noted earlier, and standardized Black-White differences on the two subtests are also large (.90) versus small (.30; see Table 18.2).

Moreover, the *g* factor alone accounts for more than four times as much of the total between-group variance in test scores as do the three largest non-*g* factors combined (Jensen, 1998, p. 379). This means that group differences in mental test scores stem mostly from differences in *g*. Studies of *reaction time,* which refers to quickness (in milliseconds) of apprehending simple perceptual stimuli (e.g., a light or sound), show that racial differences in reaction time also track the tasks' *g* loadings. Accordingly, groups with the highest average *g* also have the fastest choice reaction times: Asian children tend to be the fastest, Black children the slowest, and Whites intermediate (Jensen, pp. 389–402).

IQ Tests Somewhat Underestimate Group Differences in *g*

Although IQ tests provide good estimates of *g*, the non-*g* Stratum I and Stratum II contaminants in IQ scores mean that racial differences in IQ usually underestimate racial differences in *g*. To illustrate, the standardized Black-White gap on the major IQ batteries is usually 1.0 to 1.1 *SD*s, which represents an average Black-White difference of 15 to 17 IQ points (when the test's $SD = 15$). When *g* scores are estimated from those same batteries, the gap in *g* is closer to 1.3 *SD*s (Jensen, 1998, p. 377, 2003). The Black-White gap in *g* is thus close to 20 points in the IQ metric.

All Groups Follow the Same Developmental Path, but Some Just More Slowly

Like other aspects of growth, mental age increases with chronological age until adolescence, at which time growth starts to level off. Lower IQs represent a slower *rate* of cognitive development. (Recall that the IQ represents cognitive ability relative to agemates, not some absolute level of ability.) Spearman's hypothesis therefore predicts that between-race differences in IQ will mimic within-race differences in mental development. Pursuing this line of inquiry, Jensen (1998) found that Black elementary school children tend to perform less like White children of the same chronological age

Table 18.2

Standardized Mean IQ Differences (d_{IQ}), Relative to Whites, of American Blacks, Hispanics, Native Americans, and Asians on Different Tests, at Different Ages, and in Different Years

	White		Black			Hispanic			Native American			Asian		
Year[a]/Age	Mean	*SD*	Mean	*SD*	d_{IQ}	Mean	*SD*	d_{IQ}	Mean	*SD*	d_{IQ}	Mean	*SD*	d_{IQ}
380 studies in Shuey (1966)[b]														
1922–1944 (ages 2–6)					**.60**									
School children														
Individually administered tests					**.93**									
Group tests: verbal					**.87**									
Group tests: nonverbal					**1.07**									
High school					**.73**									
All samples					**.87**									
1945–1966 (ages 2–6)					**1.07**									
School children														
Individually administered tests					**.93**									
Group tests: verbal					**1.07**									
Group tests: nonverbal					**.87**									
High school					**1.27**									
All samples					**1.07**									
Coleman et al. (1966) Equality of Educational Opportunity Report[c]														
1965 data														
1st grade IQ														
Verbal	53.2		45.4		**.78**	44.9/46.5[d]		**.83/.67**[d]	47.8		**.54**	51.6		**.16**
Nonverbal	54.1		43.4		**1.07**	45.8/50.1		**.83/.40**	53.0		**.11**	56.6		**–.25**
12th grade IQ														
Verbal	52.1		40.9		**1.12**	43.1/43.8		**.90/.83**	43.7		**.84**	49.6		**.25**
Nonverbal	52.0		40.9		**1.11**	43.3/45.0		**.87/.70**	47.1		**.49**	51.6		**.04**
12th grade achievement tests														
Reading	*51.9*		*42.2*		***.97***	*42.6/44.2*		***.93/.77***	*44.3*		***.76***	*48.8*		***.31***
Math	*51.8*		*41.8*		***1.00***	*43.7/45.5*		***.81/.63***	*45.9*		***.59***	*51.3*		***.05***
General Info	*52.2*		*40.6*		***1.16***	*41.7/43.3*		***1.05/.89***	*44.7*		***.75***	*49.0*		***.32***

	National Longitudinal Study (NLS) of High School Class of 1972[e]													
1972 Data														
12th grade	208	31	169	28	**1.16**									
	WISC-R Standardization Sample[f]													
1974 (ages 6–16.5)	102.3		86.4		**1.06**									
FSIQ	103.2	13.8	87.8	13.1	**1.03**									
Verbal IQ	102.0	14.2	87.8	13.2	**.95**									
Performance IQ	102.2	14.1	87.2	13.4	**.94**									
Subtests														
Comprehension	10.4	2.8	7.8	2.5	**.90**									
Block Design	10.4	2.9	7.7	2.7	**.90**									
Vocabulary	10.4	2.9	7.9	2.8	**.85**									
Object Assembly	10.7	3.0	7.9	3.0	**.82**									
Information	10.4	2.9	8.1	2.7	**.77**									
Similarities	10.3	3.0	7.9	2.9	**.80**									
Picture Completion	10.4	2.9	8.1	3.0	**.76**									
Picture Arrangement	10.4	2.9	8.1	3.0	**.76**									
Mazes	10.4	3.1	8.4	3.2	**.67**									
Arithmetic	10.4	2.8	8.6	2.8	**.58**									
Coding	10.2	3.3	8.9	2.9	**.45**									
Tapping Span	10.1	2.9	9.1	3.0	**.32**									
Digit Span	10.1	3.0	9.2	3.2	**.30**									
	WISC-R: 98 California School Districts[g]													
1975 (ages 5–11)	103.2	13.4	87.8	13.1	**1.03**									
	SB-IV Standardization Sample[h]													
1986														
Ages 2–6	104.7	14.7	91.0	13.2	**.86**	94.9	14.8	**.61**	99.7	19.9	**.31**	88.0	17.8	**1.04**
Ages 7–11	102.6	15.6	92.7	13.2	**.62**	93.8	13.2	**.55**	94.1	17.3	**.53**	103.6	13.1	**–.06**
Ages 12 to 18–23	103.5	15.8	86.1	15.1	**1.09**	94.9	13.4	**.54**	94.7	17.7	**.55**	99.9	15.4	**.23**

(*continued*)

Table 18.2 *Continued*

	White		Black			Hispanic			Native American			Asian		
Year[a]/Age	Mean	*SD*	Mean	*SD*	d_{IQ}	Mean	*SD*	d_{IQ}	Mean	*SD*	d_{IQ}	Mean	*SD*	d_{IQ}
DAS Standardization Sample[i]														
1986 data														
Ages 2.6–3.5	103.0	14.1	91.5	13.4	**.77**	96.0	14.4	**.47**						
Ages 3.6–5.11	103.6	13.7	86.7	13.3	**1.23**	93.6	12.0	**.67**						
Ages 6–17	102.7	14.3	89.2	14.1	**.90**	93.9	13.4	**.59**				107.2	14.1	**–.30**
WISC-III Standardization Sample[j]														
1991 (ages 6–16)	103.5		88.6		**.99**	94.1		**.63**						
PPVT-R: Children of National Longitudinal Study of Youth (NLSY) Mothers														
1986–1994 data (1980 for mothers)														
Ages 3–4[k]	52		40		**1.20**									
Ages 5–6[l]	98.9	15.2	81.9	14.6	**1.13**									
Mothers' AFQT[l]	99.7	12.4	85.7	10.8	**.93**									
Armed Forces Aptitude Batteries														
1917–1918 (WWI) data: Recruits[m]					**1.16**									
1944–45 (WWII) data: AGCT for recruits[n]					**1.25**									
1980 data: AFQT standardization sample (NLSY)[o]														
Ages 18–23	522	86.9	401	94.5	**1.21**	429	105.8	**.93**						
WAIS-R Standardization Sample[p]														
1981														
Ages 16–19	100.8	14.1	86.9	14.5	**.93**									
Ages 20–34	101.8	15.1	87.0	11.6	**.99**									
Ages 35–54	101.4	14.8	86.6	13.2	**.99**									
Ages 55–74	101.4	14.6	87.0	13.0	**.96**									
Total	101.4	14.7	86.9	13.0	**.97**									

			GATB: Job applicants[q]									
1940–1970 data	102		84		**.90**	91		**.55**	84	**.90**	98	**.10**
			Wonderlic Personnel Test: Job applicants[r]									
1970: Ages 16–72	23.3	7.5	15.8	7.1	**.94**	17.0	7.6	**.79**				
1983: Ages 16–72	23.5	7.2	16.2	6.9	**.95**	17.2	6.6	**.82**				
1992: Ages 16–72	22.8	6.8	16.2	6.4	**.85**	17.3	6.7	**.77**				

Note. Effect size is calculated as the group mean difference (e.g., White–Black) divided by the total *SD* for the battery in question, including all racial/ethnic groups. I note when *d*s are based on medians rather than means. Negative effect sizes indicate that the minority mean was higher than the White mean.

[a] Except where otherwise specified, "Year" refers to year of publication and not year of data collection.

[b] Shuey (1966), as reported in Eitelberg (1981, p. 12).

[c] Coleman et al. (1966, p. 20). Mean = 50, *SD* = 10; effect sizes based on medians, not means.

[d] First entry under "Hispanics" is for Puerto Rican Americans, and the second is for Mexican Americans.

[e] Osborne (1982, p. 260). IQ = "ability index," which is sum of NLS tests of vocabulary, reading comprehension, mathematics, and letter groups (inductive reasoning).

[f] WISC-R = Wechsler Intelligence Scale for Children–Revised. From Jensen and Reynolds (1982, p. 425). For IQ scales, mean = 100 and *SD* = 15; for subtests, 10 and 3, respectively.

[g] Jensen and Figueroa (1975, p. 885). Mean = 100, *SD* = 15. Authors used mean weighted *SD* to calculate effect size of 1.15 (rather than the *SD* = 15 used here to yield effect size of 1.03).

[h] SB-IV = Stanford-Binet Intelligence Scale, Fourth Edition. From Thorndike, Hagen, and Sattler (1986, pp. 34–36). Mean = 100, *SD* = 16.

[i] DAS = Differential Ability Scales. From Lynn (1996, p. 272).

[j] Sattler (2001, p. 232). Mean = 100, *SD* = 15.

[k] PPVT-R = Peabody Picture Vocabulary Test–Revised. From Jencks and Phillips (1998, p. 2). Mean = 50, *SD* = 10; effect sizes based on differences in medians, not means.

[l] Phillips, Brooks-Dunn, Duncan, Klebanov, and Crane (1998, p. 108). Mean = 100, *SD* = 15. NLSY mothers of foregoing 5- to 6-year-olds. Young mothers and their children are overrepresented.

[m] Loehlin, Lindzey, and Spuhler (1975, pp. 143, 408–409). Based on a variety of tests (Army Alpha, etc.) put on one scale.

[n] Jensen (1998, p. 376).

[o] AFQT = Armed Forces Qualifying Test. From Laurence, Eitelberg, and Waters (1982, p. 43). Mean = 500, *SD* = 100. Same sample as used in NLSY (e.g., Herrnstein & Murray, 1994). "White" includes all racial/ethnic groups other than Blacks and Hispanics.

[p] WAIS-R = Wechsler Adult Intelligence Scale–Revised. From Reynolds et al. (1987, p. 330). Mean = 100, *SD* = 15.

[q] GATB = General Aptitude Test Battery. From U.S. Department of Labor (1970, Table 17-12). Mean = 100, *SD* = 20; "Hispanic" refers to Mexican American.

[r] Wonderlic Personnel Test, Inc. (1999, p. 34). Mean = 22; *SD*s = 8.02 for 1970, 7.70 for 1983, and 7.10 for 1992.

than like Whites of the same *mental* age, who tend to be chronologically 2 years younger. The similarity of older Black children to chronologically younger Whites goes beyond merely obtaining the same total test scores to exhibiting the same psychometric particulars, such as the sophistication of errors they make (Jensen, pp. 365–366). These observations suggest that Blacks follow the same developmental path as Whites, just more slowly, and therefore eventually level off at a lower average mental age. (See Jensen, 2003, for an especially interesting analysis.)

DO GROUP DIFFERENCES IN IQ DIFFER BY AGE, SOCIAL CLASS, OR DECADE?

Table 18.2 shows the standardized mean IQ differences, or effect sizes (d_{IQ}), for the largest and most representative studies in the United States from the 20th century. It organizes them roughly by date, but places at the end the four data sets including only adults (ages 16+). Although these are the best available studies, it should be noted that their data are not entirely comparable.[2] Nor were data available for the most recent standardizations of the Weschler and Stanford Binet tests. The estimates in Table 18.2 are about 10% smaller than those reported elsewhere for the same data, but the difference is just an artifact of the method necessary for calculating them here.[3]

STANDARDIZED GROUP DIFFERENCES IN IQ (d_{IQ}) DO NOT APPEAR TO DIFFER BY AGE

The standardized differences in mean IQ from preschool through high school in Table 18.2 illustrate just about every possible age trend: rising—or falling—at higher ages/

[2] First, although only studies that are broadly representative of the specified race and age groups have been included, not all are nationally representative samples of the groups in question (e.g., for adults, they sometimes include only applicants to jobs). This is particularly a problem for the samples of children in the ages before school entry and after compulsory attendance, and especially for studies from the early decades of the 20th century when relatively few adolescents attended high school. Second, although the White and Black subsamples are all fairly large and those for Hispanics often adequate, the Native American and Asian samples are almost always small and thus plagued by much sampling error. Third, the ethnic composition of the Hispanic and Asian categories is quite heterogeneous, and their subgroups differ considerably in mean IQ: "Hispanic" includes Mexicans, Central Americans, Cubans, and Puerto Ricans (and can be of any race), and "Asian" includes Japanese, Chinese, Pacific Islanders, Cambodians, Vietnamese, Indonesians, Pakistanis, East Indians, and more (some of the latter being Caucasian). The composition of these two broad categories often differs greatly by locale and has changed much over the decades because of shifting immigration policy and periodic influxes of refugees. Fourth, it cannot be presumed that all the IQ tests listed are equally sound psychometrically (e.g., reliability, ceilings/floors on scores) or equally *g*-loaded. This means it cannot be assumed they would all yield the same effect sizes under the same conditions for the same true differences in *g*. Inadequate *g*-loading may be a problem especially in the early grades, because it can be difficult at these ages to measure *g* adequately with group-administered rather than individually administered IQ tests; but group tests are the only feasible way to test large samples.

[3] All methods begin by calculating the mean difference between two groups' test scores, and then standardize that difference by dividing it by a relevant measure of dispersion in the scores. The preferred measure of dispersion is usually the *N*-weighted average of the two groups' *SD*s. Standardized differences derived by this method are not comparable across studies, however, because the denominator fluctuates with the *N*s and *SD*s for the particular minority groups in the sample. The White *SD* would provide the most universally comparable denominator for American samples, but it often is not available, so this chapter uses the *SD* for the entire (multiracial) population in question, which for IQ tests is usually set to either 15 or 16 at each age. The *d* derived from the foregoing three ways of defining the reference *SD* become successively smaller, with the *d* used here being perhaps 10% smaller than those calculated with an *N*-weighted *SD* (which explains why some *d* in Table 18.2 are smaller than those published elsewhere for the same IQ test).

grades; and peaking—or troughing—in the intermediate ages/grades. Rather than signaling volatility in development across ages, this instability probably represents some combination of sampling error, differential representativeness of the samples, and different *g*-loadings from test to test. The cross-age consistency of effect sizes for adults on the WAIS-R (see Table 18.2) and Kaufman Adolescent and Adult Intelligence Test (KAIT; Kaufman, McLean, & Kaufman, 1995) provides evidence that the average Black-White gap in general cognitive ability is, in fact, stable by adolescence. The sparser data for Hispanics in Table 18.2 point in that direction, too, but no comparable data were located for Native Americans and Asian Americans. Of course, group averages conceal the many small and occasional moderate shifts in individual IQ during childhood (Moffitt, Caspi, Harkness, & Harkness, & Silva, 1993), but this within-group shifting in rank relative to age peers does not reposition groups along the IQ continuum.

Standardized Group Differences in IQ (d_{IQ}) Are Larger at Higher Social Class Levels

Jensen and Reynolds (1982) found that controlling for social class in the WISC-R standardization sample reduced the average Black-White IQ difference from 15 to 12 IQ points. The IQ effect size for all Blacks and Whites was 1.03 *SD*, but effect sizes increased from .12 for children in families at the lowest socioeconomic level to 1.20 at the highest (calculated from their tables in Jensen & Reynolds, pp. 885, 887). More advantaged family background is therefore accompanied by bigger, not smaller, group differences in children's cognitive ability. Moreover, there is little overlap between the two sets of averages. To illustrate, Black children from the most advantaged families average about the same IQ (IQ 89–95) as do the least advantaged Whites (IQ 85–94). The same general pattern is found for WAIS-R IQs at different education levels (Reynolds et al., 1987): Effect sizes are larger at higher education levels, and college-educated Blacks have no higher average IQs (95.9) than do White adults with only 9 to 11 years of education (98.0).

Group Differences in IQ (d_{IQ}) Did Not Change in the United States Over the 20th Century

Table 18.2 shows that, regardless of the decade of data collection or reporting, virtually all d_{IQ} for Blacks in Table 18.2 fall within the range of 1.00 ± .2. Hispanic effect sizes are almost always .70 ± .2, regardless of decade. The only effect sizes outside this range occur when different subgroups of Hispanics are distinguished: As seen in the 1966 Coleman data in Table 18.2, Puerto Rican Americans tend to score at least .15 *SD* below Mexican Americans. The IQ effect sizes for Native Americans (.50 ± .4) and Asian Americans (.15 ± .4) vary more, owing both to sampling error and to these groups' nonverbal scores' being considerably better than their verbal scores. (Some studies in Table 18.2 provided IQ scores only separately by verbal and nonverbal IQ.) Nonetheless, it appears that Native Americans probably score close to Mexican Americans, and Asian Americans close to Whites (sometimes surpassing them), in the various time periods. The data are too unstable, however, to conclude anything about possible trends over time for these smaller minority groups. The best hypothesis for all groups is still the null hypothesis, that is, that group differences in IQ remained the same throughout the century.

The secular increases in average IQ throughout the developed world during the

20th century have led many commentators to assume that the long-standing racial/ethnic gaps in cognitive ability must be quite malleable (Neisser, 1998, chap. 1). Whether, and to what extent, the secular increases in IQ represent increases in *g* itself is still a hotly contested issue (Neisser). As just seen, however, this debate is not relevant in the current context because the racial/ethnic IQ gaps, where they were measured reasonably well (Blacks vs. Whites), remained the same over the entire century. As Kaufman and Lichtenberger (2001, p. 101) have concluded, the Black-White difference of roughly 1 *SD* is "seemingly impervious to time."

HOW DO GROUPS DIFFER IN STANDARDIZED ACHIEVEMENT?

Just as group differences in cognitive ability are pervasive across time, place, and age, so too are group differences in standardized academic achievement. The latter appear to be somewhat more elastic, however.

GROUP DIFFERENCES IN STANDARDIZED ACHIEVEMENT (d_{ach}) ARE UBIQUITOUS, LARGE, AND STUBBORN

The *achievement gap* is the new shorthand for enduring, nationwide racial disparities that continue to pervade all forms of academic achievement, regardless of all attempts to eradicate them (Steinberg, 1996; Thernstrom & Thernstrom, 2003). The achievement gap bedevils even the most affluent, socially liberal communities that have struggled earnestly to equalize achievement across racial lines (Banchero & Little, 2002). Noguera (2001), for example, describes the early confidence and later disappointment of the Minority Student Achievement Network (MSAN), a consortium formed in 1999 by 14 advantaged communities to eradicate their districts' achievement gaps. He captures the sad experience of all MSAN members when he describes the district on whose school board he had served: Despite an "impressive track record of public support, Berkeley schools [continue to be] characterized by extreme disparities in academic outcomes. . . . The majority of White and Asian students score at or above the 80th percentile on most norm referenced tests, while the scores of Black and Latino students are generally closer to the 30th percentile" (cf. Jensen, 1991, on Berkeley's efforts).

Ogbu (2003) documents equally distressing disparities in another MSAN school district, the affluent suburb of Shaker Heights, Ohio. Its school system is "one of the best in the nation" and Black students report the schools "to be exceptionally good" (pp. xii, 12). The Shaker Heights community, which is one third Black, is "highly educated" and describes itself as "middle- and upper middle-class." In the 1960s it had already become "a model of a voluntarily self-integrated community" (p. xii). The school district possesses all the educational resources and interracial spirit that were once thought to hold the answer to closing the achievement gap, and its Black students do, in fact, outperform Blacks elsewhere in Ohio. And yet its achievement gaps are huge, as Ogbu describes in dispiriting detail (pp. 5–7). To cite a typical finding from four subjects in three grades, White 8th graders averaged 92% on the math proficiency test, but Blacks averaged 37%. Black students received 80% of the Ds and Fs in high school semester grades. The average GPA was 1.6 for all Blacks versus 2.87 for all Whites; for high school graduates, the GPAs averaged 2.22 and 3.34, respectively, for the two races. In 1992–1995, only 22 Blacks were among the 310 students ranked in the top 20% of their graduating class; but 295 out of the 325 students ranking in the bottom 20% were Black. College-bound Blacks averaged more than 1 *SD* below Whites on both the Scholastic Achievement Test (SAT) Verbal (485 vs. 600) and

SAT Math (471 vs. 598), and about 74% of Black graduates versus 90% of White graduates went to college.

A survey of students in all MSAN schools found few or no racial differences in student attitudes toward school, perceptions of teachers, or effort devoted to schoolwork (Ferguson, 2002). It did, however, reveal big self-reported racial differences in payoff for effort invested, such as degree of understanding the material read for school, frequency of understanding the teacher's lesson, and GPA. The most extensive and incisive empirical assessment of the Black-White difference in test scores in the United States (Jencks & Phillips, 1998) reported that the Black-White cognitive gap is large, exists prior to school entry, does not change appreciably during the elementary and secondary school years, and originates mostly in factors outside schools (cf. Steinberg, 1996; Thernstrom & Thernstrom, 2003).

Group Differences in Standardized NAEP Achievement (d_{ach}) Narrowed Somewhat in the 1980s

The best data on trends in standardized academic achievement come from the U.S. Department of Education's National Assessment of Educational Progress (NAEP), which has assessed large representative samples of American school children aged 9, 13, and 17 with the same tests for over three decades. Table 18.3 shows standardized mean differences in reading, mathematics, and science achievement for Blacks and Hispanics, relative to Whites. Assuming that the content and psychometric properties of the three NAEP achievement tests were successfully kept constant from year to year, then the achievement gaps for Blacks and Hispanics were larger in the 1970s than in the later two decades. The median effect sizes for the two spans of time (in brackets) in Table 18.3 indicate that achievement gaps narrowed 25% in reading but under 20% in math for both races (respectively, from 1.06 to .79 and 1.07 to .87 for Blacks, and from .88 to .66 and .85 to .71 for Hispanics). The already larger d_{ach} in science narrowed less for Blacks (15%) and not at all (or grew slightly) for Hispanics. There is no obvious trend in NAEP performance across the 1980s and 1990s, but recent reports suggest the gaps have started to widen again (Grissmer, Flanagan, & Williamson, 1998; Hedges & Nowell, 1998).

Group Differences in Standardized NAEP Achievement (d_{ach}) Are Larger in Science Than Math and Larger in Math Than Reading

Table 18.3 shows that beginning in the 1980s the NAEP achievement gaps tended to be smallest in reading, somewhat larger in math, and considerably larger in science for both Blacks and Hispanics: Median effect sizes in the three subjects were, respectively, .79, .87, and 1.04 (Blacks) and .66, .71, and .86 (Hispanics). Table 18.2 shows the same pattern for Black 12th graders in the 1965 Coleman data (where a test of general information replaces science).

Standardized Gaps in NAEP Achievement (d_{ach}) Do Not Differ by Age or Grade

The cross-sectional data in Table 18.3 reveal no clear differences in NAEP achievement gaps across the three age groups (9, 13, and 17) for the two minority groups studied (Blacks and Hispanics). Longitudinal data tell the same story (e.g., Phillips, Crouse, & Ralph, 1998). The cross-age stability of achievement gaps is consistent with the cross-age stability of IQ gaps.

Table 18.3

Achievement Gaps (d_{ach}) Observed (and Predicted) for Blacks and Hispanics in NAEP Reading, Math, and Science, at Ages 9, 13, and 17, in 1971–1999

	Observed d_{ach}[a]											
	Reading				Math				Science			
	9	13	17	Median	9	13	17	Median	9	13	17	Median
Blacks												
1971	1.04	1.08	1.15	1.06[b]								
1975	.92	1.02	1.19									
1977									1.22	1.10	1.23	1.22
1978					.88	1.08	1.07	1.07				
1980	.84	.91	1.19									
1982					.84	1.02	.98		1.03	1.04	1.25	
1984	.79	.74	.79									
1986					.74	.79	.93		.86	1.03	1.01	
1988	.71	.53	.55									
1990	.79	.58	.71	.79	.81	.68	.87	.87	1.02	1.02	1.04	1.04
1992	.83	.73	.86		.82	.93	.87		.97	1.16	1.07	
1994	.80	.77	.66		.74	.90	.89		.95	1.15	1.08	
1996	.74	.82	.69		.75	.92	.89		.88	1.05	1.03	
1999	.91	.74	.73		.82	.98	1.02		1.02	1.06	1.18	
Mean observed												
1970s	.98	1.05	1.17		.88	1.08	1.07		1.22	1.10	1.23	
1980s	.78	.73	.84		.79	.91	.96		.95	1.03	1.13	
1990s	.81	.73	.73		.79	.92	.87		.97	1.09	1.08	
Predicted gap[c]:		.82–1.20				.84–1.20						
Hispanics												
1971	—	—	—									
1975	.88	.83	.92	.88								
1977									.84	.98	.79	.84
1978					.59	.86	.85	.85				
1980	.82	.78	.58									
1982					.57	.63	.79		.98	.82	.95	
1984	.76	.65	.67									
1986					.63	.63	.79		.78	.90	.86	
1988	.58	.61	.64									
1990	.62	.68	.53	.66	.65	.70	.84	.71	.78	.86	.85	.86
1992	.65	.69	.61		.70	.63	.65		.86	.80	.76	
1994	.79	.75	.73		.81	.77	.72		.96	.92	.98	
1996	.64	.71	.70		.66	.81	.71		.76	.88	.83	
1999	.72	.63	.57		.76	.73	.72		.84	1.05	.69	
Mean observed												
1970s	.88	.83	.92		.59	.86	.85		.84	.98	.79	
1980s	.72	.68	.63		.60	.65	.81		.88	.86	.91	
1990s	.68	.69	.63		.72	.73	.73		.84	.90	.82	
Predicted gap[c]:		.62–.90				.63–.90						

Note. Dashes = sample size too small to permit reliable estimate.

[a] Effect sizes calculated with *SD*s for entire national sample of students that age in that year taking that test. Original data are from the National Center for Education Statistics (2000).

[b] Boxes show median effect size for the bracketed values.

[c] Predictions from columns (5) (Blacks) and (6) (Hispanics) in Table 18.4.

Group Differences in Absolute Levels of NAEP Proficiency Increase in the Higher Grades

Differences in *g* represent differences in the ability to profit from instruction and learn moderately complex material. Higher-*g* students therefore tend to accumulate knowledge at a faster rate than do lower-*g* peers, which means that gaps in their *absolute* levels of knowledge widen further in each successive grade. A slow learner who starts Grade 2 only one grade equivalent behind may be three or four grade equivalents behind by Grade 12.

As discussed earlier, Blacks average about 2 years behind Whites in mental development during the elementary grades. The latest available NAEP proficiency scores (National Center for Education Statistics, 2000) reveal that Blacks trail Whites in NAEP reading and math achievement by at least three grade levels by age 13 and at least four by age 17. The gap is even larger in science. The overall pattern is the same for Hispanics, except that Hispanics trail Whites by about one fewer years than do Blacks at these ages. Recall that standardized gaps in both NAEP achievement and IQ remain the same across age groups, so no new factors need be introduced to explain why groups with lower mean IQs seem to fall further behind in the higher grades.

Gaps in NAEP Achievement (d_{ach}) Within Social Classes Parallel Within-Class Gaps in IQ

Blacks and Hispanics tend to come from poorer families than Whites, but this cannot explain their lower average levels of achievement. As noted earlier, Black-White IQ gaps are only somewhat smaller when social class is controlled (12 vs. 15 IQ points), and they are largest at the highest social class levels. Gaps in achievement are likewise larger higher up the social ladder. Table 18.4 reveals the by-now familiar pattern, first for social class origin and then for social class destination. The top panel shows that Black children of college graduates perform no better in NAEP reading, on average, than do White children whose parents dropped out of high school; Hispanic children of college graduates perform no better than Whites whose parents completed high school only. This pattern is comparable to that reported earlier for IQ, where Black children with parents holding jobs in the highest several occupational strata scored no better on IQ tests than did White children whose parents work in the lowest strata.

The middle panel of Table 18.4 shows race and class differences in mean levels of functional literacy on the U.S. Department of Education's National Adult Literacy Survey (NALS). NALS literacy levels among Black college graduates are close to the literacy levels of Whites who obtained only high school diplomas. The 4-year mean difference in reading proficiency between Black and White 17-year-olds described earlier thus seems to follow the two racial groups into the college-age years and beyond. It is also consistent with the WAIS-R results reported earlier, in which Black adults with at least some college scored somewhat lower in IQ (95.9) than Whites who had dropped out of high school (98.0). Once again, Hispanics score somewhat better than Blacks, but still lag comparably educated Whites by several grade equivalents in NALS literacy.

The bottom panel of Table 18.4, for college-bound high school seniors, shows that SAT scores follow the same general pattern with regard to socioeconomic background. Black students from families earning over $100,000 in 1999 scored no better on the SAT than did Whites from families earning only $20,000 to 35,000 (SAT total scores of 1006 vs. 1010). Within-class SAT gaps relative to Whites were smaller for Hispanics and

Table 18.4

Mean Proficiency Levels of Whites, Blacks, Hispanics, and Asians on Three Achievement Tests

	1994 NAEP Reading (Grade 12)[a]			
	White	Black	Hispanic	Asian
	Parents' highest educational level (years completed)			
0–11	274	258	260	—
12	283	258	265	—
13–15	294	271	279	—
16+	302	272	283	—

	1992 NALS (Ages 16+)[b]							
	Prose Literacy				Quantitative Literacy			
	White	Black	Hispanic	Asian	White	Black	Hispanic	Asian
	Own educational level (years completed)							
0–8	202	159	135	—	195	140	128	—
9–11	243	213	200	—	242	197	196	—
12	278	242	242	209	279	232	240	227
13+ (no degree)	302	267	265	264	304	258	265	273
2-year degree	313	276	291	—	313	267	286	—
4-year degree	328	288	282	271	329	280	286	286
Graduate degree	341	298	312	301	338	285	312	314

	1999 SAT: Verbal + Math (College-bound seniors only)[c]			
	White	Black	Hispanic	Asian
Family income				
< 20,000	986	803	838	950
20–35,000	1010	851	900	1018
35–60,000	1033	888	955	1065
60–100,000	1072	928	1002	1124
>100,000	1131	1006	1063	1191

Note. NAEP = National Assessment of Educational Progress; NALS = National Adult Literacy Survey, SAT = Scholastic Achievement Test. Dashes indicate sample size too small to permit reliable estimate.

[a] Campbell, Donahue, Reese, and Phillips (1996, p. 37).

[b] Kirsch, Jungeblut, Jenkins, and Kolstad (1993, p. 127). Standard errors are 1.2–3.8 for Whites, 1.6–5.2 for Blacks, 3.5–9.1 for Hispanics, and 5.7–16.0 for Asians on these scales.

[c] Data provided by the College Board (1999).

usually reversed for Asians. That socioeconomically advantaged Blacks and Hispanics score no better than disadvantaged Whites and Asians is apparently the rule, not the exception, on all highly *g*-loaded tests.

CAN THE IQ GAPS ACCOUNT FOR THE ACHIEVEMENT GAPS?

If nothing except the IQ gaps were responsible for the achievement gaps, then the maximum standardized achievement gap to be expected between any two groups is simply their standardized gap in *g* (which is estimated from their IQ gaps), and the minimum to be expected is the aforementioned maximum multiplied by IQ's (disattenuated) validity for predicting academic achievement (see note c in Table 18.1). If observed gaps are clearly smaller than the predicted minima, then other factors are *neutralizing* some of *g*'s usual influence; if the gaps are larger than the predicted maxima, then something else is *adding* to the achievement gaps.

In the 1960s, Achievement Gaps (d_{ach}) Were Near the Maximum Predicted by IQ Gaps (d_{IQ})

The 1966 Coleman report provides both IQ and achievement test scores for over 645,000 students from six racial/ethnic groups in more than 3,000 schools. Despite its age, it also provides the most extensive data yet available for Native Americans and Asians. Table 18.2 shows that achievement gaps at that time were near the maximum expected because there was rough concordance between the standardized gaps in verbal and nonverbal IQ, on the one hand, and standardized gaps in reading and math achievement, on the other hand. As calculated from the table, the respective mean d_{IQ} and d_{ach} for 12th graders were 1.11 versus .98 (Blacks), .82 versus .78 (the two Hispanic groups), .66 versus .67 (Native Americans), and .14 versus .18 (Asians). The d_{ach} in general information tended to be larger than those for reading and math and, at least for Hispanics, somewhat larger than the group's IQ gap.

In the 1980s, Group Differences in Achievement (d_{ach}) Moved Closer to the Minima Predicted by Group Differences in IQ (d_{IQ}) but Remained Within Expected Ranges

There is no large study after 1970 comparable to the Coleman Report in reporting good data for both IQ and achievement. With proper caution, however, achievements gaps in one study can be compared to typical IQ gaps in others.[4] As shown in Table 18.1, the maximum expected achievement gaps are 1.20 for Black Americans, .90 for Hispanic Americans (and perhaps Native Americans, too), and up to |.30| for Asian Americans. The minimum standardized achievement gaps to be expected on good tests of the three Rs, solely on the basis of IQ gaps, are .76 to .84 for Blacks, .57 to .63 for Hispanics, and .14 to |.21| for Asians.

Table 18.3 shows the NAEP achievement gaps actually observed in the last three decades for Blacks and Hispanics, relative to Whites, at ages 9, 13, and 17. The median observed gaps in reading, math, and science for both Blacks (1.06, 1.07, 1.22) and Hispanics (.88, .85, .84) during the 1970s were 88% to 102% of the predicted maxima (1.20 for Blacks, .90 for Hispanics), which put them at most only a third of the way from the maxima toward the minima expected for achievement in the three Rs. After 1980, gaps in NAEP reading achievement for Blacks (.79 median) fell to the minimum expected (.77 language, .82 reading) but not quite that far for Hispanics (.66 median, where the minima expected are .62 in reading and .58 in language). The gaps in math moved one half (Hispanics) to three quarters (Blacks) of the way toward the expected minima, but

[4] One must be extremely cautious in comparing effect sizes for IQ (d_{IQ}) in one study to effect sizes for achievement (d_{ach}) in another. In addition to the usual problems of sampling error and differential psychometric quality of tests being paired, the samples may not be equally representative. That is why only large, nationally representative, and psychometrically sound data sets are used here to make such comparisons.

the d_{ach} in science remained near the maximum predicted for both racial/ethnic groups. Achievement gaps have thus narrowed most in the group (Blacks) and subject (reading) most intensely targeted by educational reforms. That none of the achievement gaps fell materially below the *g*-predicted minima may signal a natural lower bound for feasible reductions in achievement gaps absent any reductions in the IQ gaps. In short, achievement gaps seem to be somewhat more elastic than IQ gaps, but still tethered to them.

DO RATES OF EDUCATIONAL PLACEMENT AND PERSISTENCE ALSO MIRROR IQ GAPS?

Group disparities in placement and advancement are commonly discussed as part of the so-called *pipeline problem*, namely, that groups do not pass in equal proportion through the successive filters governing educational and occupational advance. Such disparities are generally described in the metric of over- and underrepresentation, so Figure 18.2 presents group IQ gaps in those terms, too. It provides estimated cumulative percentages of Whites, Blacks, Hispanics, and Asians at each 5-point step along the continuum of normal IQ (70–130). The bottom three rows estimate degree of over- and underrepresentation for each group, relative to Whites, above or below five specific IQ thresholds: IQ 75, 90, 100, 110, and 125. These three gradients of over- and underrepresentation along the IQ continuum are the relative rates to expect if *g* is the only factor creating group differences in placement and persistence. Deviations of observed from *g*-predicted differentials in progress can provide clues to non-*g* factors at work.

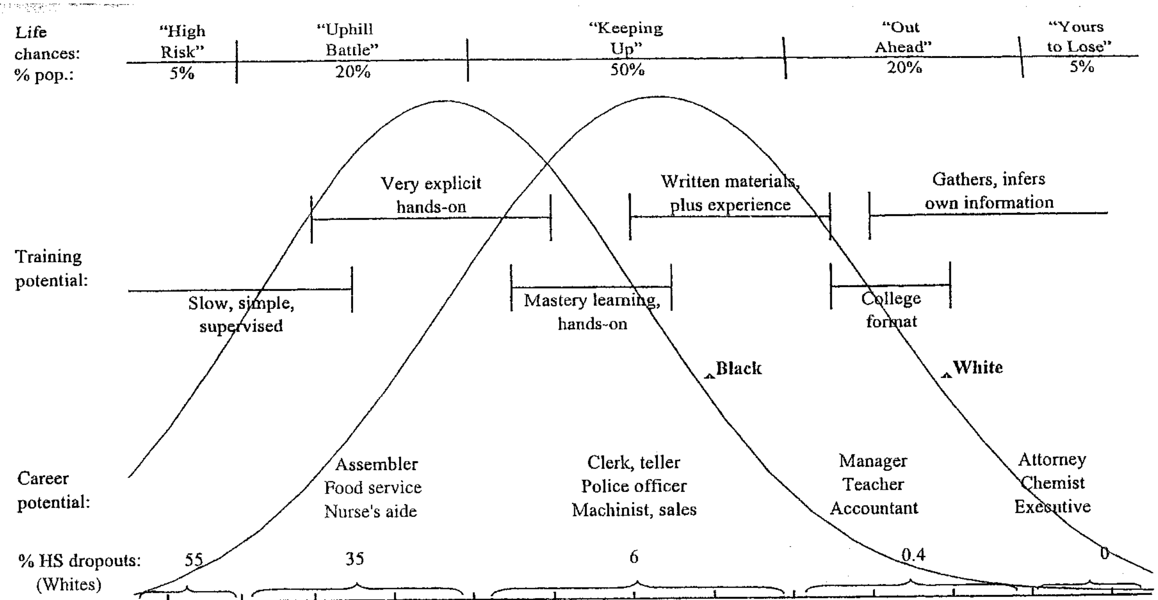
The upper half of the Figure 18.2 summarizes previous analyses of typical training potential, typical occupation level, and (for Whites) high school dropout rates for individuals in five successive segments of the IQ continuum. As indicated by the dropout rates, high *g* does not guarantee educational success or low *g* guarantee failure, but the odds of success in many realms of life improve steadily and substantially at successively higher IQ levels.

GROUP DIFFERENCES IN IQ REPRESENTATION SHIFT ALONG THE IQ CONTINUUM AND ARE LARGER FOR GROUPS SEPARATED FURTHER FROM WHITES IN MEAN IQ

The representation of lower-scoring groups relative to Whites shifts from overrepresentation at the left tail of the IQ distribution to underrepresentation at the right tail. For example, the proportion of Blacks relative to Whites shifts from 5:1 overrepresentation below IQ 75 to 1:30 underrepresentation above IQ 125. Disparities in representation above IQ 100 relative to Whites are largest for Blacks (1:3), smaller for Hispanics (1:2), and yet smaller (and reversed) for Asians (6:5).

GROUP DIFFERENCES IN IQ REPRESENTATION ARE MOST STRIKING AT THE TAILS OF THE IQ DISTRIBUTION

Whereas Blacks are overrepresented by less than 3:1 (41% vs. 18%) between IQs 76 and 90, they are overrepresented by almost 5:1 (18% vs. 4%) below IQ 75. The disparities are even greater at the right tail, with Black underrepresentation worsening from 1:7 for IQs 111 to 125, to 1:30 above IQ 125. The disproportions for Hispanics are less marked but still large at the two extremes: almost 4:1 (14% vs. 4%) under IQ 75 and 1:5 above IQ 125. If the IQ data for Asians are even roughly accurate, then the pattern is



WAIS IQ:	70	75	80	85	90	95	100	105	110	115	120	125	130
Pop. percentile:		**5th**	10th	16th	**25th**	37th	50th	63rd	**75th**	84th	90th	**95th**	
Cumulative %													
White	2	**4**	7	13	**22**	33	46	60	**72**	82	90	**95**	**97**
Black	10	**18**	30	44	**59**	73	84	92	**96**	98	99	**99+**	**99+**
Hispanic	8	**14**	23	34	**47**	60	73	82	**90**	94	97	**99**	**99+**
Asian	1	**2**	4	8	**14**	23	34	47	**60**	73	82	**90**	**94**

Ratio in IQ range	**IQ≤75**	**IQ≤90**	IQ>100	**IQ>110**	**IQ>125**
Black/White	**5:1**	**3:1**	1:3	**1:7**	**1:30**
Hispanic/White	**4:1**	**2:1**	1:2	**1:3**	**1:5**
Asian/White	**2:3**	**2:3**	6:5	**4:3**	**2:1**

Figure 18.2 The distribution of people and life chances along the IQ continuum.
Note: Cumulative percentages are based on mean Wechsler Adult Intelligence Scale (WAIS) IQs of 101.4 for Whites and 86.9 for Blacks, and *SD*s of 14.7 for Whites and 13.0 for Blacks (Reynolds, Chastain, Kaufman, & McLean, 1987, p. 330). Means used for Hispanics and Asians were, respectively, 91 and 106, and an *SD* of 15 was used for both. The estimates for Whites and Blacks are good, reasonable for the conglomerate Hispanic category, but more questionable for Asians. Percentiles for IQ scores were estimated by use of cumulative normal probability tables. Minority/White ratios were calculated before percentiles were rounded. Adapted from Gottfredson (1997, Fig. 3, p. 117). Copyright 1997 by Elsevier Science. Reprinted with permission.

reversed for that group: *under*representation below IQ 75 (2:3) and *over*representation at the highest levels (2:1).

GROUP DIFFERENCES IN PLACEMENT IN SPECIAL EDUCATION ARE SMALLER THAN EITHER IQ GAPS OR RATES OF NAEP BASIC PROFICIENCY WOULD PREDICT

Black students have long been diagnosed as mentally retarded at a higher rate than Whites, and such disparities have often been litigated as evidence of racial bias in placement. The ratio of Black placements for mental retardation has fallen from over 3:1 to just over 2:1 relative to Whites in the last 30 years (Donovan & Cross, 2002, p. 46). Hispanics are now placed at a lower rate than Whites, Native Americans at a slightly higher rate, and Asians at about half the rate of Whites. Figure 18.2 shows that these racial disproportions in placement for mental retardation are considerably smaller than either IQ or achievement gaps alone would predict for Blacks and Hispanics, but only somewhat so for Asians. An IQ of 70 to 75 is often considered the upper threshold for mild mental retardation, and proportionately five times as many Blacks, four times as many Hispanics (and presumably Native Americans), and two thirds as many Asians fall below IQ 75 as do whites.

Actual performance in the classroom is usually more important than IQ test results in recommending educational placements. The top panel of Table 18.5 therefore shows the percentages of White, Black, Hispanic, and Asian 4th and 12th graders who scored *below* the basic level expected of their grades on recent NAEP tests in reading, math, and science—that is, as not having even "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade" (Donahue et al., 1999, p. 9). Table 18.5 shows that from 21 to 27% of Whites, 64 to 68% of Blacks, and 58 to 60% of Hispanics in the 4th grade failed to reach the basic level in core subjects.

Looking at Figure 18.2, these percentages conform roughly to the cumulative percentages of the three groups scoring below IQ 90, which is the 25th percentile of the general population. In the 12th grade, the criterion for basic proficiency is again set commensurate with about IQ 90 in math, but closer to IQ 85 in reading and IQ 95 in science (judging from the proportions of each group failing to meet the basic criterion). Figure 18.2 also shows that Blacks are overrepresented below IQ 90 by a factor of almost 3:1 and Hispanics by over 2:1. Thus, even in the unlikely event that 4th and 12th graders were both placed into special education randomly from among children below the 25th percentile in either ability or achievement, racial disproportions in special education placement would be at least as large as those actually observed. But regardless of actual rates of placement, these data reveal a huge racial disparity in preparedness for academic progress. By this measure, Black 4th graders are twice as likely to lack as to possess the minimum prerequisites for developing proficiency at their grade levels, and the ratio is not much better for Hispanics. Whites and Asians, in contrast, are more than twice as likely as not to *have* basic proficiency.

GROUP DIFFERENCES IN PLACEMENT IN GIFTED EDUCATION ARE SMALLER THAN EITHER IQ GAPS OR RATES OF PROFICIENT ACHIEVEMENT ON THE NAEP WOULD PREDICT

Classes for the gifted and talented are likewise perennially under fire for underrepresenting non-Asian minorities (Boothe & Stanley, 2004; Ford, 2003). Selection into gifted classes has often required, at least in years past, scoring well on an intelligence

Table 18.5

Percentages of Students Reaching Particular NAEP or SAT Proficiency Levels or Persisting Until Graduation, by Racial/Ethnic Group and Grade Level

	% Low in NAEP Proficiency (Below "Basic" Level)					
	Grade 4			Grade 12		
	Reading[a] 1998	Math[b] 1996	Science[c] 2000	Reading 1998	Math 1996	Science 2000
White	**27**	**24**	**21**	**17**	**21**	**38**
Black	**64**	**68**	**66**	**43**	**62**	**78**
Hispanic	**60**	**59**	**58**	**36**	**50**	**70**
Native American	**53**	**48**	**43**	**35**	—	**56**
Asian	**31**	**27**	**34**[d]	**25**	**19**	**41**
	% High in NAEP Proficiency (At or Above "Proficient" Level)					
	Grade 4			Grade 12		
	Reading 1998	Math 1996	Science 2000	Reading 1998	Math 1996	Science 2000
White	39	28	38	47	20	23
Black	10	5	7	18	4	3
Hispanic	13	8	11	26	6	7
Native American	14	8	10	27	—	9
Asian	37	26	29[d]	38	33	26
	% Persisting to High School and College Graduation,[e] Persons Aged 25–29, 2002				% College-Bound Seniors With SAT Scores >600, 1999[f]	
	<12 yrs	≥12	≥16		Verbal	Math
White	**7.0**	93.0	35.9		25	26
Black	**12.4**	87.6	18.0		6	5
Hispanic	**37.6**	62.4	8.9		9[g]	9
Native American	—	—	—		15	14
Asian	—	—	—		23	41

Note. NAEP = National Assesment of Educational Progress; SAT = Scholastic Achievement Test. Dashes indicate sample size too small to permit reliable estimate.

[a] Source of reading scores: Donahue, Voelkl, Campbell, and Mazzeo (1999).

[b] Source of math scores: Reese, Miller, Mazzeo, and Dossey (1997).

[c] Source of science scores: O'Sullivan, Lauko, Grigg, Qian, and Zhang (2003).

[d] Data for Asians are for 1996, not 2000.

[e] U.S. Census Bureau (2002, table A-2).

[f] College Board (1999, p. 34).

[g] Data listed for Hispanics are for "Mexicans/Mexican Americans," but the data are highly similar for all Hispanic groups in the College Board (1999) publication.

test (often above IQ 130, about the 98th percentile). Blacks and Hispanics are rarely selected under this criterion because they are relatively rare at that IQ level. Figure 18.2 indicates ratios above IQ 125 (about the 95th percentile) of 1:30 and 1:5, respectively, for the two groups. Even if the entrance criterion were lowered to IQ 100, underrepresentation would still be large for Blacks (1:3) and Hispanics (1:2). This is actually closer the rates currently found. As reported to the Office of Civil Rights, placement into gifted education has increased from under 1% to over 6% of students in the last 30 years, but the rate relative to Whites has remained a bit under 1:2 for both Blacks and Hispanics (Donovan & Cross, 2002, p. 54). The relative rate for Asians has fallen, however, from 2:1 to almost 1:1. Whereas lower-scoring minorities are represented in gifted education in far greater proportion than their IQ gaps would predict, Asians are now represented at half their expected rate.

Achievement tests would produce the same pattern of underrepresentation for Blacks and Hispanics as would IQ, judging from group differences in performing at or above the "proficient" level on the NAEP. As seen in Table 18.5, the percentages of Whites, Blacks, and Hispanics who meet or exceed the proficient criterion in reading line up near the 50th to 63rd percentile in IQ (that is, above IQ 100–105) and with the 75th for math and science (above IQ 110). (Note that the percentages of students at or *above* proficiency must be subtracted from 100% before comparing them to the cumulative percentages *below* a particular IQ level in Figure 18.2.) If entry to gifted classes were set at the NAEP proficient level in math and science (about the 75th percentile in ability), Blacks would be underrepresented by at least 1:5 and Hispanics by 1:3 in both Grades 4 and 12. Asians would once again be overrepresented if placement rested on performance in math and science.

Thus, placement decisions that rest mostly on student achievement will yield striking racial imbalances. Conversely, when placement differentials are much smaller than expected on the basis of achievement, as they are today, then nonachievement factors will necessarily have strongly influenced placements.

BLACKS ARE MUCH MORE LIKELY AND HISPANICS MUCH LESS LIKELY TO GRADUATE HIGH SCHOOL AND COLLEGE THAN IQ AND ACHIEVEMENT GAPS WOULD PREDICT

As just described, the proportions of White, Black, and Hispanic 4th and 12th graders who perform below the NAEP basic level are commensurate with the proportions of those groups scoring below about IQ 90. The bottom panel of Table 18.5 reveals, however, that the rate at which 25- to 29-year-olds have dropped out of high school is far better for Blacks and far worse for Hispanics than expected on the basis of these groups' IQ and achievement gaps relative to Whites. While dropout rates in 2002 were 12.4%, 37.6%, and 7.0% for Blacks, Hispanics, and Whites (Table 18.5), the percentages of 12th graders scoring below the basic level in NAEP reading in 1998 were, respectively, 43%, 36%, and 17% (Table 18.5). The percentages below IQ 85 are 44%, 34%, and 13% (Figure 18.2).

Hispanics likewise graduate from college at far lower rates than Blacks despite their higher average IQ and achievement levels: 18.0% (Blacks), 8.9% (Hispanics), and 35.9% (Whites). This incommensurability between dropout rates, on the one hand, and IQ and achievement, on the other, illustrates the long-known fact that Blacks tend to go further in school, and Hispanics less far, than Whites of the same ability levels. It thereby signals that there are important noncognitive factors enhancing persistence among Blacks and depressing it among Hispanics relative to Whites of the same ability levels.

WHAT CHALLENGES DO IQ GAPS CREATE FOR SCHOOLS?

Research has not yet established conclusively why the group disparities in *g* exist, but their resistance to change warns us that they will likely be with us for some time to come. What implications do they have for schools, especially when political considerations press schools to deny their existence?

Within-Group Differences in *g* Create a Dilemma for Democracies and Between-Group Differences Intensify It

The public schools are located on the front lines of a clash between two guiding aims of democratic nations: equal opportunity and equal outcomes. In educational circles, this clash is often reframed as the conflict between excellence and equity, where excellence requires helping all *individuals* achieve to their potential (equal opportunity) and equity requires helping all *groups* achieve to the same level regardless of potential (equal outcomes).

All racially diverse societies face this conflict, as Klitgaard (1986) illustrates in his analysis of selection practices in China, Malaysia, the United States, the Philippines, Indonesia, and other nations with large subgroups who differ in mean cognitive ability. Many multiethnic countries are now struggling with this conflict, from the Netherlands (te Nijenhuis, de Jong, Evers, & van der Flier, 2004) to Brazil (Lloyd, 2004) to South Africa (Skuy et al., 2002). The underlying dilemma is shared, however, even by racially homogeneous societies, because all human populations exhibit wide dispersion in *g*. Equalizing learning opportunities never eradicates inequality of achievement, but always assures it. Conversely, equalizing outcomes across groups requires providing members of lower scoring groups more opportunities or assistance than members of higher scoring groups. That is, there is a trade-off between equal opportunity and equal outcomes. American schools are nonetheless expected to achieve both.

Expecting Schools To Do the Impossible Yields Inevitable Failure and Destructive Cycles of Blame

Interventions that raise average achievement levels typically increase variance in achievement as well. It is therefore unreasonable to expect schools to equalize achievement at a high level, as No Child Left Behind requires, as long as current IQ gaps remain. Equalizing achievement at any level would require all Black and Hispanic students routinely to perform as well, on the average, as White and Asian classmates who average about 1 *SD* higher in IQ, where 1 *SD* represents fully one quarter of the range of normal IQ. For example, Blacks somewhat below average in IQ (IQs 76–90) would have to perform as well as Whites of average ability (IQs 91–110). Figure 18.2 hints at how much more difficult learning is for individuals in the former than the latter IQ range. In like manner, Blacks who are somewhat above average in IQ (IQs 111–125) would have to match the performance levels of gifted Whites and Asians (IQ > 125). This, too, is probably impossible, even if the nation were willing to provide systematically better instruction to Blacks and Hispanics than to Whites and Asians.

But even to acknowledge the existence of the democratic dilemma is to risk being accused of desiring or creating social inequality. That teachers or citizens even notice average group differences in ability can earn them the epithet *prejudiced* or worse (including from social scientists; see Gordon, 1997, on national surveys of racist beliefs).

When no one can speak the facts, when "straight talk" is prohibited (Frisby, 1999), unrealistic promises proliferate unchecked. Frustrated expectations devolve into blame. Test critics blame the tests, test companies blame the schools, educators blame already angry parents, and the races accuse each other of moral weakness. But flagellating one group or another for lack of will or commitment has no constructive effect. As long as group disparities in cognitive ability remain, it is unreasonable to expect—and unwise to demand—that schools produce parity in achievement.

SCHOOLS THAT ATTEMPT TO ERADICATE ACHIEVEMENT GAPS BY DEMOCRATIZING EDUCATION PRODUCE A HIGHLY PREDICTABLE CASCADE OF DESTRUCTIVE SIDE EFFECTS

Because schools cannot eradicate achievement gaps, political imperatives force them to disguise, deny, or shuffle them around. Their policy choices reside primarily in where in the educational process and in what form the gaps will become visible, as they inevitably do (cf. Jensen, 1991). This process can be observed in current efforts to supposedly democratize education.

First, using racial preferences to equalize selection into higher level curricula leads later to noticeable group differences in achievement. When groups pass through an educational gate with unequal skills, big achievement gaps emerge further down the pipeline. For example, some gifted programs use race-norming to attain racial proportionality in gifted enrollment (e.g., Richert, 2003), and many colleges use racial preferences in admissions. As they all discover, however, the favored groups do no better than their lower academic skills would predict. Gifted programs and elite universities that use preferences end up with Black and Hispanic student bodies whose ability distributions hardly overlap those of Whites and Asians (e.g., Herrnstein & Murray, 1994, p. 455). This, in turn, guarantees that failure will become color coded, leaving the institution vulnerable to charges of discriminating against its minority members, especially if it has claimed that all groups were equally qualified when selected. Many colleges try to mute such side effects by providing special race-specific curricula and support programs, but such programs create their own problems (Bernstein, 1994; Bunzel, 1992).

Second, broadening access to higher level programs or curricula creates pressures to make them easier. As opposition to racial preferences has mounted, institutions have turned to what they describe as more holistic selection standards. These involve using more varied and more subjective indicators to net a greater percentage of non-Asian minority individuals. To the extent they succeed in doing so, it is primarily by making selection more random with respect to *g*. For example, many gifted programs now use broader, less intellectual definitions of giftedness and talent and more open methods of referral (including self- and parent nominations; Colangelo & Davis, 2003; Richert, 2003).

The destructive side effect, of course, is that when students of all races are selected under lower standards, all are less able to cope with advanced curricula. Failure rates increase unless the curriculum is watered down, so it is no surprise that many gifted programs are replacing acceleration with mere enrichment. Figure 18.2 reveals, furthermore, that race-neutral selection systems have to be almost random with regard to *g* (or achievement) to produce anything near racial parity in selection (cf. Schmitt, Rogers, Chan, Sheppard, & Jennings, 1997). The more open admissions are made, the more fully curricula have to be stripped of cognitive demands in order to forestall high failure rates and the emergence of achievement gaps. A recent example is the College

Boards reducing the rigor of Advanced Placement (AP) courses to bring them within reach of a broader cross-section of students (Bleske-Rechek, Lubinski, & Benbow, 2004). Denuding these programs of their intellectual challenge vitiates their educational purpose, of course, which is to challenge and nurture the talents of highly able students who are not served well by the regular curriculum.

Third, degrouping and detracking students only highlights the IQ gaps that such practices are meant to hide. Children differ greatly in their readiness for instruction, and therefore have often been grouped for instruction by ability or achievement level. Although less often litigated than placement in special education, ability grouping and curriculum tracking have been equally contentious (Kulik, 2003). Figure 18.2 helps to illustrate why—they produce big group disparities in placement. For instance, if college-preparatory classes in high school were to recruit students from above IQ 90 (i.e., the 25th percentile of the population), only 41% and 53% of Blacks and Hispanics but 78% and 86% of Whites and Asians would be eligible. Even under this low standard, then, the pools of eligible Blacks and Hispanics would be, respectively, no more than one half and two thirds as large as for those for Whites. Recall that this standard would include all students performing merely at or above the basic level (partial mastery of a subject's prerequisites), which is hardly auspicious for college work. Turning to honors and AP classes, if they were to draw students primarily from IQ 110 and above (the 75th percentile), the pools of eligible Blacks and Hispanics would be only one seventh as large for Blacks (4% of Blacks) and one third as large for Hispanics (10% of Hispanics) as for Whites (28%). Other factors affect curriculum placement, but achievement (and hence *g*) are obviously central. Noticeable racial disproportions in placement have become politically unacceptable, however, and grouping systems are routinely assailed as elitist, discriminatory tools for sustaining social inequality (e.g., Ford, 2003). They are said to resegregate the races within desegregated schools.

Many schools are therefore attempting to degroup and detrack their students (Good & Brophy, 2003, chap. 7; Kulik, 2003). It is simple enough administratively to abolish ability grouping, but this places into the same classroom students who differ by many grade equivalents in their ability to profit from further instructions, regardless of race. For instance, students in heterogeneous junior high school classrooms generally span 8 to 10 grade levels in achievement (Good & Brophy, p. 329). Recall that the average Black or Hispanic student tends to lag several grade equivalents behind the average White or Asian by age 13 and by at least another year by Grade 12. These gaps guarantee unmistakable racial differences in academic performance in the typical ability-heterogeneous classroom, especially in hierarchical subjects such as math. Not surprisingly, research shows that degrouping hurts the self-esteem of less able students and raises it among more able ones (Kulik, p. 272). These side effects generate the suspicion that schools are hostile environments for non-Asian minorities.

Fourth, instructional strategies for ability-heterogeneous classrooms work largely by holding back faster learners. Advocates of degrouping are not blind to its difficulties, so they offer new instructional strategies for educating unequally able children side-by-side in the same classroom (Good & Brophy, 2003, chap. 8). Some strategies like mastery learning, have been claimed to equalize learning rates among students, but that claim has been shown to be false (Snow, 1996). Others, like cooperative learning and peer tutoring, narrow gaps primarily by diverting high-achieving students from additional learning into helping low-achieving classmates improve (Robinson, 2003). Successfully instructing the full range of abilities in the same classrooms, which now often contain mainstreamed special education students, too, requires truly heroic efforts by teachers (see a case example in Good & Brophy, pp. 315–316). Not only must teachers minister to quite different learning needs, but they must also do so in a manner that protects low-achieving students from being stigmatized by classmates observing their lower levels of achievement. On the basis of site visits, experts have concluded that "untracking brings no guarantee of high-quality instruction for everyone but may instead lead all to a common level of educational mediocrity" (Kulik, 2003, p. 279).

Fifth, degraded and homogenized instructional programs drive middle-class families out of schools, leaving behind high concentrations of low-achieving students and peer climates hostile to learning. Parents are generally sympathetic with schools' allocating proportionately more resources to slow learners, but they are far less tolerant of schools' reducing opportunities for their children to receive appropriately differentiated curricula. Many parents who are able to do so will enroll their children in private schools or move out of the school district. The schools from which middle-class parents flee eventually end up with high concentrations of low-achieving students from dysfunctional families and with peer climates that are increasingly hostile to learning. Bankston and Caldas (2002) describe this sort of unintended devastation after the courts required three Louisiana districts to balance the public schools racially.

HOW SHOULD WE DEAL WITH GROUP DIFFERENCES IN COGNITIVE ABILITY?

Racial disparities in cognitive ability are, empirically, just the summation of individual differences in ability. In fact, the average Black-White IQ difference among children from the same social class is no larger than the mean difference (about 12 IQ points) between biological siblings who grow up in the same household (Jensen, 1998). The daily challenges faced by persons in the high-risk or uphill-battle ranges of IQ (see Figure 18.2) are difficult, regardless of race. There are no race-specific cures for low achievement and no technical solutions to the democratic dilemma. The socioeducational integration and fair treatment of individuals of disparate ability levels is always a challenge, even within the same family. Most families, however, neither expect nor demand that all siblings perform to the same level, and most believe it inappropriate for parents to treat children who have different needs in an identical manner. It is likewise unwise for a nation to insist that all its subgroups perform to the same average level.

The democratic dilemma creates difficult choices, but there is certainly no call to give up on anyone because all students could probably learn more than they do now (for several promising programs, see Thernstrom & Thernstrom, 2003). What all children need is ability-appropriate instruction, not identical instruction, race-specific curricula, or patronizing pretense. Lower-*g* students, like low-*g* workers (Sticht, Armstrong, Hickey, & Caylor, 1987, p. 94), require more complete and more concrete instruction in smaller increments with more scaffolding, whereas higher-*g* students profit more from abstract, self-directed, incomplete instruction that allows them to assemble new knowledge and reassemble old knowledge in idiosyncratic ways (Reschly, 1997; Snow & Lohman, 1984).

We have no way to equalize the rate at which students learn. *Amount* learned is more manipulable, but it can be equalized only by stalling the progress of brighter students while helping the less able to catch up. Some might view this as essential to social justice and racial harmony, but there are other choices. None will satisfy everyone, because they all involve some trade-off between equal opportunity and equal results, between individual excellence and group parity. Human dispersion in *g* will always pose tough choices, but understanding that fact can help us choose more wisely.

REFERENCES

Armour-Thomas, E., & GoPaul-McNicol, S. (1998). *Assessing intelligence: Applying a bio-cultural model.* Thousand Oaks, CA: Sage.

Banchero, S., & Little, D. (2002, November 13). "Minorities score poorly even at high scoring schools." *Chicago Tribune*, p. 1.

Bankston, C. L., III, & Caldas, S. (2002). *A troubled dream: The promise and failure of school desegregation in Louisiana.* Nashville, TN: Vanderbilt University Press.

Bernstein, R. (1994). *Dictatorship of virtue: Multiculturalism and the battle for America's future.* New York: Knopf.

Bickley, P. G., Keith, T. Z., & Wolfle, L. M. (1995). The three-stratum theory of cognitive abilities: Test of the structure of intelligence across the life span. *Intelligence, 20,* 309–328.

Bleske-Rechek, A., Lubinski, D., & Benbow, C. P. (2004). Meeting the educational needs of special populations: Advanced placement's role in developing exceptional human capital. *Psychological Science, 15,* 217–224.

Boothe, D., & Stanley, J. C. (2004). *In the eyes of the beholder: Critical issues for diversity in gifted education.* Waco, TX: Prufrock Press.

Brody, N. (1992). *Intelligence* (2nd ed.). San Diego, CA: Academic Press.

Brody, N. (2003). Construct validation of the Sternberg Triarchic Abilities Test: Comment and reanalysis. *Intelligence, 31,* 319–329.

Bunzel, J. H. (1992). *Race relations on campus: Stanford students speak.* Stanford, CA: Stanford Alumni Association.

Campbell, J. R., Donahue, P. L., Reese, C. M., & Phillips, G. W. (1996). *NAEP 1994 reading report card: Findings from the National Assessment of Educational Progress.* Washington, DC: National Center for Education Statistics. Retrieved January 5, 2005, from http://nces.ed.gov/nationsreportcard//pdf/main1994/96045.pdf

Carretta, T. R., & Ree, M. J. (1995). Near identity of cognitive structures in sex and ethnic groups. *Personality and Individual Differences, 19,* 149–155.

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

Colangelo, N., & Davis, G. A. (2003). Introduction and overview. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 3–10). Boston: Allyn & Bacon.

Coleman, J. S., Campbell, E. Q., Hobson, C J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity.* Washington, DC: U.S. Department of Health, Education, and Welfare, Office of Education.

College Board. (1999). *Counselor's handbook for the SAT program, 1999–2000.* Princeton, NJ: Author.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain.* Oxford, England: Oxford University Press.

Donahue, P. L., Voelkl, K. E., Campbell, J. R., & Mazzeo, J. (1999). *NAEP 1998 reading report card for the nation and the states.* Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement.

Donovan, M. S., & Cross, C. T. (Eds.). (2002). *Minority students in special and gifted education.* Washington, DC: National Academy Press.

Dreger, R. M. (1973). Intellectual functioning. In K. S. Miller & R. M. Dreger (Eds.), *Com-*

parative studies of Blacks and Whites in the United States (pp. 185–229). New York: Seminar Press.

Eitelberg, M. J. (1981). *Subpopulation differences in performance on tests of mental ability: Historical review and annotated bibliography* (Tech. Mem. No. 81-3). Alexandria, VA: Office of the Assistant Secretary of Defense, Directorate for Accession Policy.

Fan, X., Willson, V. L., & Reynolds, C. R. (1995). Assessing the similarity of the factor structure of the K-ABC for African-American and White children. *Journal of Psychoeducational Assessment, 13*, 120–131.

Ferguson, R. F. (2002). Addressing racial disparities in high-achieving suburban schools. *NCREL Policy Issues*, Issue 13. http://www.ncrel.org/policy/pubs/html/pivol13/dec2002b.htm

Ford, D. Y. (2003). Equity and excellence: Culturally diverse students in gifted education. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 506–520). Boston: Allyn & Bacon.

Frisby, C. L. (1993). One giant step backward: Myths of Black cultural learning styles. *School Psychology Review, 22*(3), 535–557.

Frisby, C. L. (Ed.). (1999). Straight talk about cognitive assessment and diversity [Special issue]. *School Psychology Quarterly, 14*(3).

Gardner, H. (1983) *Frames of mind: The theory of multiple intelligences*. New York: Basic Books.

Good, T. L., & Brophy, J. E. (2003). *Looking in classrooms* (9th ed.). Boston: Allyn & Bacon.

Gordon, R. A. (1997). Everyday life as an intelligence test: Effects of intelligence and intelligence context. *Intelligence, 24*(1), 203–320.

Gottfredson, L. S. (1997). Why *g* matters: The complexity of everyday life. *Intelligence, 24*(1), 79–132.

Gottfredson, L. S. (2002). *g*: Highly general and highly practical. In R. J. Sternberg & E. L. Grigorenko (Eds.), *The general factor of intelligence: How general is it?* (pp. 331–380). Mahwah, NJ: Erlbaum.

Gottfredson, L. S. (2003). Dissecting practical intelligence theory: Its claims and evidence. *Intelligence, 31*(4), 343–397.

Gottfredson, L. S. (2004). Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology, 86*(1), 174–199.

Gray, J. R., & Thompson, P. M. (2004). Neurobiology of intelligence: Science and ethics. *Nature Reviews: Neuroscience, 5*, 471–482.

Grissmer, D., Flanagan, A., & Williamson, S. (1998). Why did the Black-White score gap narrow in the 1970s and 1980s? In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 182–226). Washington, DC: Brookings Institution Press.

Gustafsson, J.-E. (1988). Hierarchical models of individual differences in cognitive abilities. In R. J. Sternberg (Ed.), *Advances in the psychology of human intelligence* (Vol. 4, pp. 35–71). Hillsdale, NJ: Erlbaum.

Hedges, L. V., & Nowell, A. (1998). Black-White test score convergence since 1965. In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 149–181). Washington, DC: Brookings Institution Press.

Helms, J. E. (1992). Why is there no study of cultural equivalence in standardized cognitive ability testing? *American Psychologist, 47*, 1083–1101.

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

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29*(3), 340–362.

Jencks, C., Bartlett, S., Corcoran, M., Crouse, J., Eaglesfield, D., Jackson, G., et al. (1979). *Who gets ahead? The determinants of economic success in America.* New York: Basic Books.

Jencks, C., & Phillips, M. (Eds.). (1998). *The Black-White test score gap.* Washington, DC: Brookings Institution Press.

Jensen, A. R. (1980). *Bias in mental testing.* New York: Free Press.

Jensen, A. R. (1984). Test validity: *g* vs. the specificity doctrine. *Journal of Social and Biological Structures, 7,* 93–118.

Jensen, A. R. (1991). Spearman's *g* and the problem of educational equality. *Oxford Review of Education, 17*(2), 169–187.

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

Jensen, A. R. (2003). Do age-group differences on mental tests imitate racial differences? *Intelligence, 31,* 107–121.

Jensen, A. R., & Figueroa, R. A. (1975). *Journal of Educational Psychology, 67*(6), 882–893.

Jensen, A. R., & Reynolds, C. R. (1982). Race, social class, and ability patterns on the WISC-R. *Personality and Individual Differences, 3,* 423–438.

Jensen, A. R., & Weng, L.-J. (1994). What is a good *g*? *Intelligence, 18,* 231–258.

Kaufman, A. S., & Lichtenberger, E. O. (2001). *Assessing adolescent and adult intelligence* (2nd ed.). Boston: Allyn & Bacon.

Kaufman, A. S., McLean, J. E., & Kaufman, J. C. (1995). The fluid and crystallized abilities of White, Black, and Hispanic adolescents and adults, both with and without an education covariate. *Journal of Clinical Psychology, 51*(5), 636–647.

Keith, T. Z. (1997). Using confirmatory factor analysis to aid in understanding the constructs measured by intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 373–402). New York: Guilford Press.

Keith, T. Z., Fugate, M. H., De Graff, M., Diamond, C. M., Shadrach, E. A., & Stevens, M. L. (1995). Using multi-sample confirmatory factor analysis to test for construct bias: An example using the K-ABC. *Journal of Psychoeducational Assessment, 13,* 347–364.

Keith, T. Z., Quirk, K. J., Schartzer, C., & Elliott, C. D. (1999). Construct bias in the Differential Ability Scales? Confirmatory and hierarchical factor structure across three ethnic groups. *Journal of Psychoeducational Assessment, 17,* 249–268.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993). *Adult literacy in America: A first look at the results of the National Adult Literacy Survey.* Washington, DC: National Center for Education Statistics.

Klitgaard, R. (1986). *Elitism and meritocracy in developing countries: Selection policies for higher education.* Baltimore: Johns Hopkins University Press.

Konold, T. R., Glutting, J. J., Oakland, R., & O'Donnell, L. (1995). Congruence of test-behavior dimensions among child groups that vary in gender, race-ethnicity, and SES. *Journal of psychoeducational assessment, 13*(2), 111–119.

Kulik, J. A. (2003). Grouping and tracking. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 268–281). Boston: Allyn & Bacon.

Kush, J. C., Watkins, M. W., Ward, T. J., Ward, S. B., Canivez, G. L., & Worrell, F. C. (2001). Construct validity of the WISC-III for White and Black students from the WISC-III standardization sample and for Black students referred for psychological evaluation. *School Psychology Review, 30*(1), 70–88.

Laurence, J. H., Eitelberg, M. J., & Waters, B. K. (1982). Subpopulation analyses of 1980 youth population aptitudes. In Office of Secretary of Defense (Ed.), *The profile of American youth study: Results and implications* (Tech. Mem. No. 82-2, pp. 36–52). Alexandria, VA: Office of the Secretary of Defense, Directorate of Accession Policy.

Lloyd, M. (2004, February 13). In Brazil, a new debate over color. *Chronicle of Higher Education, 50*(23). http://chronicle.com/prm/weekly/v50/i23/23a03802.htm

Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence.* San Francisco: Freeman.

Lynn, R. (1996). Racial and ethnic differences in intelligence in the United States on the Differential Ability Scale. *Personality and Individual Differences, 20,* 271–273.

Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations.* Westport, CT: Praeger.

Moffitt, T. E., Caspi, A., Harkness, A. R., & Silva, P. A., (1993). The natural history of change in intellectual performance: Who changes? How much? Is it meaningful? *Journal of Child Psychology and Psychiatry, 34*(4), 455–506.

National Center for Education Statistics (2000, August). Results over time—NAEP 1999 long-term trend summary data tables. Retrieved January 5, 2005, from http://nces.ed.gov/nationsreportcard/tables/Ltt1999/

Neisser, U. (Ed.). (1998). *The rising curve: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., et al. (1996). Intelligence: Knowns and unknowns. *American Psychologist, 51,* 77–101.

Noguera, P. A. (2001). Racial politics and the elusive quest for excellence and equity in education. *In Motion Magazine.* Retrieved January 5, 2005, from http://www.inmotionmagazine.com/er/pnrpl.html

Ogbu, J. U. (2003). *Black American students in an affluent suburb.* Mahwah, NJ: Erlbaum.

Osborne, R. T. (1982). National Longitudinal Study of the High School Class of 1972: An analysis of sub-population differences. In R. T. Osborne & F. C. J. McGurk (Eds.), *The testing of Negro intelligence* (Vol. 2, pp. 249–276). Athens, GA: Foundation for Human Understanding.

O'Sullivan, C. Y., Lauko, M. A., Grigg, W. S., Qian, J., & Zhang, J. (2003). *The nation's report card: Science 2000.* Washington, DC: National Center for Education Statistics.

Peoples, C. E., Fagan, J. F., III, & Drotar, D. (1995). The influence of race on 3-year-old children's performance on the Stanford-Binet: Fourth edition. *Intelligence, 21,* 69–82.

Phillips, M., Brooks-Dunn, J., Duncan, G. J., Klebanov, P., & Crane, J. (1998). *Family background, parenting practices, and the Black-White test score gap.* Washington, DC: Brookings Institution Press.

Phillips, M., Crouse, J., & Ralph, J. (1998). Does the Black-White test score gap widen after children enter school? In C. Jencks & M. Phillips (Eds.), *The Black-White test score gap* (pp. 229–272). Washington, DC: Brookings Institution Press.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral genetics* (4th ed.). New York: Worth.

Ree, M. J., & Earles, J. A. (1991). The stability of convergent estimates of *g*. *Intelligence, 15,* 271–278.

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.

Reed, T. E. (1971). The population variance of the proportion of genetic admixture in human intergroup hybrids. *Proceedings of the National Academy of Science U.S.A, 68,* 3168–3169.

Reese, C. M., Miller, K. E., Mazzeo, J., & Dossey, J. A. (1997). *NAEP 1996 mathematics report card for the nation and states.* Washington, DC: National Center for Education Statistics.

Reschly, D. J. (1997). Diagnostic and treatment utility of intelligence tests. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests, and issues* (pp. 437–456). New York: Guilford Press.

Reynolds, C. R., Chastain R. L., Kaufman, A. S., & McLean, J. E. (1987). Demographic characteristics and IQ among adults: Analysis of the WAIS-R standardization sample as a function of the stratification variables. *Journal of School Psychology, 25,* 323–342.

Reynolds, C. R., Lowe, P. A., & Saenz, A. L. (1999). The problem of bias in psychological assessment. In C. R. Reynolds & T. B. Gutkin (Eds.), *The handbook of school psychology* (3rd ed., pp. 549–595). New York: Wiley.

Richert, E. S. (2003). Excellence with justice in identification and programming. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 146–158). Boston: Allyn & Bacon.

Robinson, A. (2003). Cooperative learning and high ability students. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education* (3rd ed., pp. 268–281). Boston: Allyn & Bacon.

Rowe, D. C. (1997). A place at the policy table? Behavior genetics and estimates of family environmental effects on IQ. *Intelligence, 24*(1), 133–158.

Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1995). Ethnic and racial similarity in developmental process: A study of academic achievement. *Psychological Science, 6*(1), 33–38.

Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on Black-White differences in cognitive ability. *Psychology, Public Policy, and Law.*

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,* 573–605.

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.

Sattler, J. M. (2001). *Assessment of children: Cognitive applications* (4th ed.). San Diego, CA: Author.

Schmidt, F. L. (1988). The problem of group differences in ability test scores in employment selection. *Journal of Vocational Behavior, 3*(3), 272–292.

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.

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.

Shuey, A. M. (1966). *The testing of Negro intelligence* (2nd ed.). New York: Social Science Press.

Skuy, M., Gewer, A., Osrin, Y., Khunou, D., Fridjhon, P., & Rushton, J. P. (2002). Effects of mediated learning experience on Raven's Matrices scores of African and non-African university students in South Africa. *Intelligence, 30,* 221–232.

Snow, R. E. (1996). Aptitude development and education. *Psychology, Public Policy, and Law, 2*(3/4), 536–560.

Snow, R. E., & Lohman, D. F. (1984). Toward a theory of cognitive aptitude for learning from instruction. *Journal of Educational Psychology, 76,* 347–376.

Spearman, C. (1927). *The abilities of man: Their nature and measurement.* New York: Macmillan.

Steinberg, L. (1996). *Beyond the classroom: Why school reform has failed and what parents need to do.* New York: Simon & Schuster.

Steinberg, L. (2002) *Adolescence* (6th ed.). Boston: McGraw-Hill.

Sternberg, R. J. (1997). *Successful intelligence.* New York: Plume.

Sternberg, R. J., & Grigorenko, E. L. (2001). Ability testing across cultures. In L. A. Suzuki,

J. G. Ponterotto, & A. J. Meller (Eds.), *Handbook of multicultural assessment: Clinical, psychological, and educational applications* (2nd ed., pp. 335–358). San Francisco: Jossey-Bass.

Sticht, T., Armstrong, W. B., Hickey, D. T., & Caylor, J. S. (1987). *Cast-off youth: Policy and training methods from military experience.* New York: Praeger.

te Nijenhuis, J., de Jong, M.-J., Evers, A., & van der Flier, H. (2004). Are cognitive differences between immigrant and majority groups diminishing? *European Journal of Personality, 18,* 405–434.

Thernstrom, A., & Thernstrom, S. (2003). *No excuses: Closing the racial gap in learning.* New York: Simon & Schuster.

Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior, 29*(3), 332–339.

Thorndike, R. L., Hagen, E. P., & Sattler, J. M. (1986). *Technical manual for the Stanford-Binet Intelligence Scale: Fourth edition.* Chicago: Riverside.

Tu, S., Scott, M. S., Mason, C. A., & Urbano, R. C. (1999). Assessing cognitive similarities among young children of different ethnic/cultural backgrounds. *Journal of Psychoeducational Assessment, 17,* 113–126.

U.S. Census Bureau. (2002). *Statistical abstract of the United States: 2002.* Washington, DC: Author.

U.S. Department of Labor. (1970). *Manual for the USTES General Aptitude Test Battery. Section III: Development.* Washington, DC: Superintendent of Documents, U.S. Government Printing Office.

Valencia, R. R., & Suzuki, L. A. (2001). *Intelligence testing and minority students: Foundations, performance factors, and assessment issues.* Thousand Oaks, CA: Sage.

Wonderlic Personnel Test, Inc. (1999). *Wonderlic Personnel Test and Scholastic Level Exam: User's manual.* Libertyville, IL: Author.