Implications of Cognitive Differences for Schooling Within Diverse Societies

Linda S. Gottfredson

School of Education
University of Delaware
Newark, DE 19716 USA
(302) 831-1650
Fax (302) 831-6058
gottfred@udel.edu

Submitted: May 8, 2003

Running head: Implications of Cognitive Differences

Implications of Cognitive Differences for Schooling Within Diverse Societies

School psychologists face the same professional conundrum as do personnel selection psychologists. Specifically, they must deal with a reality—average group differences in cognitive ability—that American law and politics insist do not exist. They are on the front line of the clash between realities and hopes for racial parity in achievement. This chapter describes that clash. It begins by reviewing the scientific research on cognitive abilities, their uneven distribution across individuals and groups, and the group differences in academic placement and achievement that can be expected from current group differences in the distribution of cognitive ability. It concludes by outlining the challenges these differences pose for schools in diverse societies.

CHAOTIC STATE OF KNOWLEDGE THREE DECADES AGO

Racial-ethnic differences in cognitive ability and achievement became a focal point of research during the 1960s and 1970s after the U. S. Supreme Court banned *de jure* racial segregation in the public schools and as the Civil Rights Movement gained momentum. A look back at those early studies reveals how chaotic our understanding was then of the meaning, measurement, and organization of cognitive abilities and of their distribution across population subgroups.

Dreger's (1973) major review of the intellectual differences between blacks and whites typifies the sense of complexity and volatility in abilities that came to guide thinking at that time. I distill four sets of presumptions and conclusions from his review to illustrate this. As I show later, the outpouring of empirical and theoretical work in the decades since has produced a surprisingly simple picture of individual and group differences in cognitive ability. The first two of the following beliefs turned out to be false, and research on the second two has given somewhat unexpected answers.

<u>Prior Belief 1.</u> There are many independent cognitive abilities, and the "gross IQ" is just the average of the distinct intellectual functions that an IQ battery's various subtests measure. Thus the IQ score is not a particularly informative or tractable measure of intellectual functioning. There are big "complexities involved in even the simplest of separate intellectual functions…to say nothing of the almost overwhelmingly complex aspects involved in the 'IQ'" (Dreger, 1973, p. 190, see also p. 202).

<u>Prior Belief 2.</u> Ability test scores are sensitive to a wide variety of factors, perhaps including race of tester, test format and administration (e.g., verbal instructions, formidable appearance), pretraining, imprecise designation of race, non-standard English, and perceptual problems owing to darker retina. The true intellectual capacities of lower class and black children are probably significantly underestimated. Interventions to raise IQ have not yet met expectations, but perhaps because analysts have underestimated the social handicaps to be overcome.

<u>Prior Belief 3.</u> Samples of white children usually outscore samples of black children on "gross measures of intelligence," but racial differences are "less striking" and more complex on tests of specific abilities. There appear to be many interactions between race, sex, social class, and specific ability measured. However, gaps in research and differences among samples in their representativeness, SES, age, locale, and tests make it difficult to pin down just how much the two races differ, if at all, in the various forms of intellectual functioning.

<u>Prior Belief 4.</u> The structure of intellect can be illuminated by factor analysis, but it is not yet clear what that structure is or how it might differ by race. With regard to *g* in particular (the general intelligence factor), research is mixed about whether it even exists and, if it does, whether it constitutes a major intellectual distinction between blacks and whites. Sophisticated analysis indicates that "things are not that simple" (Dreger, 1973, p. 209).

The foregoing beliefs in educational psychology mirrored conventional opinion in personnel psychology at the time: for instance, that there are many different kinds of independent abilities; different combinations of abilities are required to perform well in different jobs and even different locales for the same job; and mental tests are not equally valid for different population subgroups. Researchers in personnel psychology later dubbed this view the "specificity hypothesis," where it held sway until advances in meta-analysis in the 1980s led to research disproving it (Schmidt, 1988). If the watch-word for abilities in the 1970s was "specificity," we will soon see why it is "generalizability" today.

COGNITIVE ABILITY, AS UNDERSTOOD TODAY

Today's understanding of cognitive ability emerges from research on three questions: How many major cognitive abilities are there? How are they organized? And, how might this organization or "structure" of abilities differ from one age or demographic group to another?

Number and Organization of Abilities

There are, of course, many different cognitive abilities, but they turn out not to be independent. Indeed, one of the first discoveries in the study of mental ability, a century ago (Spearman, 1904, 1927), was that all mental tests correlate positively with each other, although some (say, ones involving language) tend to correlate more highly among themselves than with others (say, ones requiring the visualization of objects in three-dimensional space). Today, psychometricians distinguish abilities primarily according to their *generality*, that is, the range of tasks for which they are useful. Carroll's (1993) massive reanalysis of hundreds of prior factor analyses crystallized the new consensus that human mental abilities can be organized hierarchically according to their specificity-generality. He labeled his synthesis the "three-stratum theory." To avoid confusion, I should emphasize that this model, like psychometrics in general, looks at how individuals *differ* in ability and how those *differences* in ability are patterned and affect important life outcomes.

--- Figure 1 About Here ---

Figure 1 provides a simplified version of Carroll's hierarchical model. It has three strata of ability factors, with Stratum I representing the most specific factors and Stratum III the most general. The many dozens of narrow Stratum I abilities include, for example, reading decoding, closure speed, ideational fluency, and memory for sound patterns. When the Stratum I abilities are factor analyzed, they yield the familiar "group factors" of ability, such as verbal ability, spatial reasoning, short-term memory, processing speed, and the like. These constitute the broad Stratum II ability factors, of which Carroll identifies eight. Gardner (1983) suggests that *g* is but one of seven intelligences he has proposed, but Carroll (1993, p. 641) points out that the four most cognitive of Gardner's yet-unmeasured intelligences (linguistic, logical-mathematical, spatial, and musical) resemble abilities at the Stratum II level of the three-stratum model (respectively, crystallized intelligence, fluid intelligence, visual perception, and parts of auditory perception). Thus, seemingly competing views of intelligence may actually be subsumed by the same hierarchical model.

The Stratum II factors correlate highly 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 this level, despite the wide variety of tests that have been analyzed. This single Stratum III factor is called *g*, short for the general mental ability factor, and it accounts for a third to half of the variance in any broad battery of mental tests. The overall score from any IQ test battery (e.g., the WISC-III's FSIQ or the WJ-III's GIQ) provides a good measure of *g*, but it is an impure measure because it always captures some non-*g* components of ability from Strata I and II, the mix of which can differ from one IQ test battery to another.

g turns out to be the major component—the spine—of all mental tests, whatever they were originally intended to measure. Tests can vary considerably, however, in their "*g* loadedness," that is, in how highly they correlate with the *g* factor and hence how well they measure *g*. This can be seen, for example, in the subtests of the Wechsler series of IQ test batteries (Sattler, 2001, pp. 238, 342, and 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 more general, group factors derived from these subtests (memory, verbal, etc.) generally load closer to .8-.9 on their battery's *g* factor.

When test scores are residualized to remove their large *g* component, little useful variance remains in the more *g*-loaded subtests because that residual seldom has much predictive value. Because *g* captures well what most people think of as intelligence, and also because it provides a purer measure of general intelligence than does the IQ score, many researchers have adopted *g* as their measure, or working definition, of intelligence.

In fact, discovery of the hierarchical structure is transforming the debate over what intelligence means. Instead of continuing the century-old debate over "What should we include in the *definition* of intelligence?" the issue is now "To which measurable empirical phenomena should we affix the *label* intelligence—to *g*, to the entire three-stratum cognitive structure, or (as does Gardner) to that structure plus valued *non*-cognitive traits?" Or, should we perhaps follow the lead of researchers who have dropped the term intelligence altogether because it no longer seems scientifically useful?

If we limit our focus to the narrow Stratum I abilities, the early belief in the multiplicity and complexity of intellectual abilities might seem to have been correct. Abilities at this level are complex mélanges of Stratum II and III ability factors and also of highly specific, narrow skills and experiences. What was not understood three decades ago, however, is that the broader abilities are *less*, not more, psychometrically complex than the narrower ones. Rather than the broad abilities being aggregates of many specific ones, the former actually provide the basic building blocks of the latter. In this sense, the broadest, most *g*-loaded abilities are not just the simplest psychometrically but also the most fundamental. They are also the most genetic and least malleable (Plomin, DeFries, McClearn, & McGuffin, 2001, ch. 10), although we did not yet know that when Dreger (1973) wrote. Highly 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. It is no wonder that, having thrown ability tests of all levels of generality and *g* loadedness into the same research pot three decades ago, we mistook the confusion for complexity and volatility. In contrast, the hierarchical model has reduced redundancy and increased interpretability of test scores by organizing abilities according to their relatedness and generality. This *unified model* of abilities (Deary, 2000) thereby allows us to catalog individual and group differences more systematically according to type and importance.

Valid Measurement of Cognitive Structure

The discovery of the hierarchical structure of cognitive abilities, with *g* at its apex, has also changed how mental test batteries and the theories undergirding them are validated. The early intelligence testers conceptualized intelligence as a very broad capability or aggregation of capabilities, so they designed their test batteries to sample a wide variety of seemingly diverse mental functions. They then attempted to prove that their tests really did measure the hypothesized intelligence by showing that the tests not only distinguished between populations as expected (e.g., retarded and gifted) but also consistently and meaningfully predicted the range of outcomes that a very broad ability should.

In contrast, today's researchers are armed with an empirically-derived theoretical model of cognitive differences, against which they can evaluate the construct validity of any test or test battery. For example, now that general intelligence (*g*) can be separated from the vehicles of its measurement (IQ tests), researchers can quantify *how well* any particular test measures *g*. Confirmatory factor analysis (CFA) has become a common method for assessing how well tests measure the constructs they were intended to, whether different batteries assess the same constructs, and whether the same battery assesses the same constructs in different subgroups (e.g., Bickley, Keith, & Wolfe, 1995).

<u>Comparability of *g* Factor Across Test Batteries, Populations, and Methods of Extraction.</u> The effort to determine whether the structure of abilities is the same and is equally well measured in all populations began with a focus on the single Stratum III factor, *g*. The question was, "Do we really get the same *g* factor when it is derived by different methods and from different populations?" 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). Moreover, 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, in press). 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. Congruence 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, & Dotar, 1995, p. 76). Most recently, Jensen (2003) found coefficients of .94-.99 for black-white comparisons in each of Grades 3-8.

Less research has examined the similarity in *g* scores 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.

The *g* factors derived from different test batteries, populations, and procedures thus appear to converge on a common "true" *g*. This means that *g* is a readily-available common yardstick for validly comparing diverse populations. IQ tests provide good but imperfect estimates of *g*, with a battery's composite IQ scores usually correlating about .95 with its own *g* factor (formula in Jensen, 1998, pp. 103-104) and probably at least .8-.9 with "true" *g* (Jensen, 1998, pp. 90-91).

<u>Comparability of Entire Cognitive Structures Across Ages and Races.</u> But what about the comparability, across race and age, of the other, *specific* factors measured by IQ test batteries? Research has found congruence coefficients of .99+ for individual specific factors in large samples of children and young adults (Jensen, 1998, p. 375). In the last decade, however, researchers have turned to confirmatory factor analysis so that they can assess simultaneously the comparability of the entire set of specific and general factors captured by a battery (Keith, 1997).

For our purposes, 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 ability factors and the same factor loadings on them for two different populations. Only when it does so can we conclude that mean group differences in *level* of performance on a test (to be discussed shortly) reflect real differences in the same underlying abilities.

CFA and MCFA research with the major IQ test batteries has consistently found that each battery does, in fact, measure the same thing in all race and age subgroups examined. Taking age first, IQ test batteries are found to measure the same ability constructs at all ages in their standardization samples: for example, across eight age groups from ages 6-79 on the WJ-R (Bickley, Keith, & Wolfe, 1995); 11 age levels from 6-16 on the WISC-III (Keith & Witta, 1997); and three age group from 5-17 on the CAS (Kranzler & Keith, 1999).

The IQ test batteries studied thus far also measure the same constructs equally well in different races at all ages examined: for both blacks and whites across three age groups between ages 7-12.5 on the K-ABC (Fan, Willson, & Reynolds, 1995; Keith et al., 1995); blacks, whites, and Hispanics across four age groups between ages 2-17 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-1, for blacks, whites, and Hispanics (Tu, Scott, Mason, & Urbano, 1999). The structure of test session behavior is likewise the same for Anglo, black, and Hispanic children in two age groups ages 6-16 in the linking sample for the WISC-III and WIAT (Konold, Glutting, Oakland, & O'Donnell, 1995). Turning to young adults, Caretta and Ree (1995) found "near identity" in the *g* factors and the five residualized group factors extracted from the Air Force Officer Qualifying Test (AFOQT) for whites, blacks, Hispanics, Native-Americans, and Asian-Americans. In summary, none of the test batteries has revealed "construct bias" because they have all measured the same ability constructs in different races at different ages.

The foregoing CFA studies provide yet more evidence that the major mental tests are valid for the English-speaking populations for which they are intended in the United States. Not being distorted by test bias, 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).

GROUP DIFFERENCES IN THE DISTRIBUTION OF COGNITIVE ABILITY

I will focus primarily on racial-ethnic differences in the single Stratum III ability, *g*, both because it is better researched than the specific abilities and because it is far more important than the Strata I and II abilities for understanding racial differences in cognitive ability and academic achievement. There are average profile differences between some racial-ethnic groups, but they constitute only a sidebar to the main story.

Group Differences in the Distribution of IQ

Most psychometric research in the United States has compared blacks and whites. Where available, I cite comparable data for Hispanic-, Asian-, and Native-American groups. Besides there being fewer data for the latter three populations, the ability data for Hispanics and Asians are made more ambiguous by higher proportions of these groups being immigrants or non-native speakers of English.

All racial-ethnic groups appear to span the entire IQ range. Virtually all individuals fall within a range of eight *SD*s (IQs 40-160), and 95% (at least in Western societies) are found within the so-called normal range of IQ, which is IQs 70-130. Every racial-ethnic group's IQ distribution is approximately normal, but has a small excess at the lower tail owing to various genetic anomalies and environmental insults.

Population subgroups 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 blacks and 14.67 for whites in the WAIS-R standardization sample (Reynolds, Chastain, Kaufman, & McLean, 1987, p. 330). These 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 among racial groups, for most purposes, is that their IQ bell curves 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, respectively, for blacks, whites, and East Asians (Rushton & Jensen, in press). In sub-Saharan Africa, black Africans, Coloreds (mixed-race individuals), and whites average about 70, 85, and 100, respectively (Rushton & Jensen, in press). (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. The various subgroups under each major racial-ethnic category, including white/European, black, Hispanic, Asian, and indigenous, likewise differ among themselves in average IQ, depending on ancestral origin. Racial-ethnic differences in the distribution of IQ are thus the rule, not the exception, and the pattern is surprisingly uniform worldwide (Rushton & Jensen, in press).

Group Differences in the Distribution of *g*

But do racial differences in IQ really represent differences in an underlying *g*? Several types of research converge in indicating that they do. 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 will differ most, on the average, on mental tests that measure *g* the best (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, in press). The relation between tests' *g* loadings and the size of black-white differences on them can be illustrated with the WISC-R sample in Table 1. As noted before, *g* loading tends to be high for Comprehension tests and low for Coding, and, as Table 1 shows, the standardized black-white differences on the two kinds of test are correspondingly large (.90) and small (.30).

Moreover, the *g* factor alone accounts for more than four times as much of the total between-group variance in test battery 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 the Stratum III ability factor, *g*. Studies of reaction time, which refers to quickness (in milliseconds) in apprehending simple perceptual stimuli (e.g., a light or sound), likewise show that racial differences on reaction time tasks track the tasks' *g* loadings. Accordingly, groups with the highest average *g* also have the fastest reaction times: Asian children tend to be the fastest, black children the slowest, and whites intermediate on choice reaction time tests (Jensen, 1998, pp. 389-402).

Like other aspects of growth in childhood, 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 age-mates, not some absolute level of ability). If group differences in IQ really represent differences in the development of *g*, then between-group differences in IQ should mimic within-race developmental differences in *g*. Jensen (1998) has pursued this line of inquiry, first noting that black elementary school children tend to perform less like white children of the same chronological age than like whites of the same *mental* age, who tend to be chronologically two years younger. The similarities of older black children with chronologically younger whites go beyond merely obtaining the same total test scores to exhibiting the same psychometric particulars, such as the sophistication of errors they make (Jensen, 1998, pp. 365-366). These observations suggest that blacks follow the same developmental path as whites, but just more slowly, and therefore eventually level off at a lower average mental age.

In a more recent test of the development hypothesis, Jensen (2003) contrasted 4^{th} with 6^{th} graders and 5^{th} with 7^{th} graders in a large California school district in 1970, all of whom were given a battery of 18 tests. Black students tended to score as far behind same-age whites as children of either race scored behind race-mates two years their senior. Specifically, after averaging across the 18 tests (whose *g* loadings ranged from .05 to .80), blacks scored .70 *SD* below whites and younger children scored .69 *SD* below race-mates two years older. Both black-white differences and within-race age differences in performance on the 18 tests correlated well with those tests' *g* loadings, confirming the Spearman hypothesis for age as well as race. In fact, racial differences in performance across the 18 tests (holding chronological age constant) were more highly correlated to *g* loading (.80) than were performance differences between older and younger children of the same race (.59). Racial differences imitated age differences to a considerable extent, as predicted, but the former seem even more yoked than the latter to *g*. These results provide more evidence that between-race differences in IQ are, in fact, differences in *g*.

Although IQ tests provide good estimates of *g,* the non-*g* Stratum I and II contaminants in IQ scores mean that racial differences in IQ usually underestimate racial differences in *g*. To illustrate: the standardized mean white-black gap on the major IQ batteries is usually 1.0-1.1

*SD*s, which represents an average black-white difference of 15-17 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; Jensen, 2003). The white-black gap in *g* is thus close to 20 points in the IQ metric.

Stability/Change in Group Differences in IQ/*g* Across Age and Time

But might not the gaps in IQ or *g* increase or decrease during the school years? It is important to know to what extent cognitive gaps exist prior to school entry and how they might change during the school years. And might racial gaps have grown or shrunk over the decades?

Table 1 lists the standardized mean IQ differences (d_{IQ}) or *effect sizes* found in large studies throughout the 20th century. It organizes the studies roughly by date, but places at the end the four datasets including only adults (ages 16+). Several cautions are in order before reviewing the d_{IQ}. First, although I have included only studies that are broadly representative of the specified race and age groups, 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 sub-samples 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, 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, we cannot presume that all the IQ tests listed are equally sound psychometrically (e.g., reliability, ceilings/floors on scores) or equally *g* loaded. This means that we cannot assume 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- rather than individually-administered IQ tests but group tests are the only feasible way to test large samples.

--- Table 1 About Here ---

Finally, the method chosen for calculating effect sizes (*d*) influences their apparent magnitude. 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 Ns and *SD*s for the particular minority groups in the sample. The Anglo-white *SD* would provide the most universally comparable denominator for American samples, but it often is not available. I have therefore opted to use the *SD* for the entire (multi-racial) 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* I use being perhaps 10% smaller than those calculated with an N-weighted *SD* (which explains why some *d* in Table 1 are smaller than those published elsewhere for the same IQ test).

Group Differences in IQ (d_{IQ}) Across Time. Have there been changes in IQ gaps over time? Regardless of the decade of data collection or reporting, virtually all d_{IQ} for blacks in Table 1 fall within the range of 1.00 $\pm$ .2. Hispanic effect sizes are almost always .70 $\pm$ .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 1, Puerto Ricans tend to score at least .15 *SD* below Mexican-Americans. The IQ effect sizes for Native-Americans (.50 $\pm$ .4) and Asians (.15 $\pm$ .4) vary more, owing both to sampling error and to these groups' non-verbal scores being considerably better than their verbal scores. (Some studies in Table 1 provided IQ scores only separately by verbal and non-verbal IQ.) Nonetheless, it appears that Native-Americans probably score close to Mexican-Americans and Asians 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, namely, that group differences in *g* remained the same throughout the last century (cf. Gordon, 1980b).

The secular increases in 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 highly malleable. Whether, and to what extent, the secular increases in IQ represent increases in *g* itself is still a hotly debated issue. As just seen, however, this debate is not relevant to our purposes because the racial-ethnic 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 one *SD* is "seemingly impervious to time."

Group Differences in IQ (d_{IQ}) Across Ages. The standardized differences in IQ from preschool through high school in Table 1 illustrate just about every possible age trend one can imagine: rising—or falling—at higher ages/grades; and peaking—or troughing—in the intermediate ages/grades. Rather than signaling volatility in developmental trends, the instability in estimates probably represents some combination of differential representativeness, sampling error, and differences in the tests' *g* loadings. I therefore see no evidence yet to reject the hypothesis that racial gaps in *g* are the same at all age/grade levels in childhood. Regarding adulthood, the cross-age consistency of effect sizes on the WAIS-R (Table 1) and KAIT (Kaufman, McLean, & Kaufman, 1995) provides evidence that the average white-black gap in general cognitive ability is, in fact, stable from at least adolescence into late adulthood. The sparser data for Hispanics in Table 1 point in that direction too, but I know of no data for Native-Americans and Asians in the United States. Of course, group averages conceal the many small and occasionally moderate shifts in individual IQ during childhood (Moffitt, Caspi, Harkness, & Silva, 1993), but this within-group shifting in rank relative to age-peers does not reposition groups along the IQ continuum.

Group Differences in IQ (d_{IQ}) Within Social Classes. A final set of IQ comparisons is useful for interpreting racial disparities in achievement in particular schools or school districts. The question is, "How similar in IQ are children of different races when they come from the same socioeconomic background?" The answer is, "More similar in the lower classes but less similar in the higher classes." Jensen and Reynolds (1982) found that controlling for SES reduced the average black-white IQ difference from 15 to 12 IQ points. Table 2 shows that the IQ effect size for all blacks and whites in the WISC-R standardization sample is 1.03, but that effect sizes increase from .12 for children from families at the lowest socioeconomic level to 1.20 at the highest. More advantaged family background therefore appears to be accompanied by bigger, not smaller, group differences in children's cognitive ability. Moreover, there is little overlap in 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).

--- Table 2 About Here ---

Turning from SES origins to SES destinations, the bottom panel of Table 2 shows average WAIS-R IQs for adults who completed different levels of education. Once again, there is a trend for larger effect sizes at higher SES levels and for the most advantaged blacks to score no better than much less advantaged whites. For example, college-educated blacks have no higher average IQs (95.9) than do whites with only 9-11 years of education (98.0).

IMPLICATIONS OF COGNITIVE DIFFERENCES FOR ACADEMIC ACHIEVEMENT

Group differences in cognitive ability emerged in bold relief after abilities were organized according to their generality and relatedness and once data were obtained from larger, more representative samples of the American population. As just described, standardized differences in IQ/*g* differ systematically from one racial-ethnic group to another but have remained stable across age and time, as best we can discern. What, however, do the group disparities in cognitive ability portend for academic achievement, for example, in different areas of the school curriculum?

The Vexing "Achievement Gap"

The "achievement gap" is the new shorthand for the enduring, nationwide racial disparities that continue to pervade all forms of academic achievement, regardless of all attempts to eradicate them. 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, p. 123 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 gap is huge, as Ogbu describes in dispiriting detail (2003, 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 vs. 2.87 for all whites; for high school graduates, the GPAs averaged, respectively, 2.22 and 3.34 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 one standard deviation below whites on both the SAT Verbal (485 vs. 600) and SAT Math (471 vs. 598), and about 74% of black graduates vs. 90% of white graduates went to college. Ogbu (2003) suggests academic disengagement as a significant cause, but 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. It did, however, reveal big self-reported racial differences in pay-off for effort invested, such as degree of understanding the material read for school, frequency of understanding the teacher's lesson, and grade-point average (Ferguson, 2002).

There is vigorous debate among social scientists over the potential causes and cures of the large and enduring achievement gaps (Lee, 2002), including those gaps just described "at the top" of the social and academic ladder (Borman, Stringfield, & Rachuba, 2000; College Board, 1999b). All sides now seem to agree, however, that large skills gaps exist in all localities, at all grade levels, in all years, and in all forms of academic achievement. 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 of schools.

American schools are nonetheless widely expected to close all achievement gaps between racial-ethnic groups—and quickly. For instance, in its recently enacted No Child Left Behind Act, the federal government requires that every school steadily improve performance in all its race and SES subgroups so that all groups meet state proficiency standards by 2014 (Banchero & Little, 2002). The requirement is explicitly aimed at eradicating all achievement gaps by race and SES. Schools that fail to produce the federally-mandated change will be sanctioned accordingly. The new law thus resembles federal law and practice in employment, which is that "disparate impact" (racial disparities in rates of passing employment tests or being hired) are prima facie evidence of racial discrimination, even if unintended.

The authors of the foregoing works of law and scholarship all reject the notion that group differences in some general cognitive ability are responsible in any meaningful way for the achievement gaps in school or work. They do, however, now accept what most social scientists did not three decades ago, namely, that group differences in test scores reflect real differences in important skills. But, again like most social scientists, they continue to reject the possibility that the ubiquity and persistence of such differences, in the face of quite varied and changing social circumstances, signals any deeper, more general phenomenon at work. Disillusioned with socioeconomic explanations, they look deeper into the "will do" factors in human behavior (academic disengagement, stereotype threat, etc.), but avoid the "can do" factors, especially the demonstrably most important one of all. They retain the belief of the 1970s that abilities are highly specific, not general, in both origin and utility. As noted earlier, however, personnel psychologists have disproved the so-called specificity hypothesis in employment and training settings. They have also have described in detail the achievement gaps to expect in these settings owing to group disparities in general cognitive ability (Sackett, Schmidt, Ellingson, & Kabin, 2001). I outline below the achievement gaps that can be predicted in the educational realm.

Predictions for Groups Based on *g*'s Importance for Individuals, Regardless of Group

If *g* had no influence on how well *individuals* performed in school, then we would not expect average *group* differences in *g* to matter either. If, however, we see that higher-*g* individuals perform better academically than do lower-*g* members of the same race, we then have to expect racial groups with more favorable IQ distributions to have higher rates of academic success, all else equal. It is therefore helpful to disregard race and ethnicity for the moment and look at *g*'s role in producing achievement differences among individuals of the same race.

Few questions have been so extensively examined in the social sciences as the power of cognitive tests to predict individuals' achievements in life. General cognitive ability turns out to be the single most useful predictor of performance in school, training, and jobs. Moreover, specific cognitive abilities add little or nothing to the prediction of performance in school or work beyond that contributed by *g* (e.g., Crouse, 1979; Ree, Earles, & Teachout, 1994; Schmidt

& Hunter, 1998; Thorndike, 1986). For instance, highly *g*-loaded tests of 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 the tests are (Crouse, 1979). Reschly (1997) argues likewise that only the general factor of IQ batteries has demonstrated treatment validity. The reason is that all mental tests' predictive value rests primarily on their dominant *g* component. The *g* factor, in turn, has pervasive predictive validity because it so strongly influences the generic higher-order thinking skills, such as efficient learning, reasoning, and problem-solving, that can be applied to content of any sort (Gottfredson, 1997, 2002). *g* 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).

Table 3 shows that the median correlations between IQ and school achievement hover around .6 for standardized tests in the three Rs. As shown in the lower half of the table, the typical IQ-achievement correlation goes up to .8, however, when different forms of achievement are aggregated into a composite (Jensen, 1998, pp. 323-324). This approaches the median correlation observed among the major individually-administered IQ test batteries (.85) and exceeds that between those batteries' composite IQs and their own more "academic" subtests (.76). The higher correlation of IQ with composite measures of achievement (.8) than with individual tests of achievement in the 3 Rs (about .6) results from the *g* components of the individual tests cumulating when tests are aggregated but their specific components tending to cancel each other out.

g's predictive validity does not differ by the manifest content of knowledge being assessed, but according to the complexity of the information to be processed: the more complex the information, the bigger the edge higher *g* provides a person in processing that information well. *g* therefore predicts performance better in more complex jobs (Hunter, 1986). It likewise predicts learning proficiency better when the material to be learned is more hierarchical (i.e., keeps building on prior learning), is meaningful, and requires insight rather than depending heavily on rote memorization (e.g., mathematics and vocabulary vs. arithmetic and spelling; Jensen, 1998, pp. 320-328). For instance, compare the correlations of IQ with reading (.61), writing (.56), and math (.63) vs. with spelling (.42) in Table 3 (column 2).

--- Table 3 About Here ---

The *g* factor does not predict all life outcomes equally well because, as just suggested, differences in *g* matter less in some endeavors than others. This is demonstrated with the other forms of achievement listed in the lower half of Table 3. 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, ch. 8; Lynn, 2002). These correlations are comparable to those discussed earlier between IQ and standardized achievement in particular school subjects. Moving down the list in Table 3, other life outcomes correlate less well (college grades, income in midlife) or very little with IQ (deportment). Explaining IQ's gradient of correlations from high to low is beyond the scope of this chapter, but one statistical and three substantive factors will suffice to illustrate what is involved. Populations that are more restricted in range in IQ—for instance, college students vs. elementary students—necessarily yield lower correlations even when the trait in question retains the same absolute importance. Turning to the first substantive issue, correlations with IQ are typically .1-.2 lower when performance is rated subjectively rather than objectively. Not only are teachers/trainers/employers' ratings of performance (e.g., grades) 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. Next, other traits and circumstances of individuals do, in fact, play a substantial role in many cumulative life achievements. For instance, differences in ambition and conscientiousness may have no discernible effect on IQ test performance, but they certainly do affect the grade point averages that students accumulate over their educational careers, just as conscientiousness has been found to contribute to performance in all jobs (Schmidt & Hunter, 1998). Finally, outcomes such as being placed in a class for the gifted, admitted to college, and being hired for a job are influenced by the institution's definition and measurement of "merit," both of which are often hotly debated. In other words, higher *g* seems to enhance one's competitiveness in many life endeavors—but more so in some than others, even among education-related outcomes.

Equally important for our purposes here, tests of cognitive ability predict standardized academic achievement, job performance, and status attainment equally well for all American racial-ethnic groups studied sufficiently to make a determination (e.g., Jencks et al., 1979; Jensen, 1980; Schmidt, 1988). Nor do there seem to be race-specific (e.g., black-only) influences on academic development in childhood (Jensen, 1998, pp. 557-559), such as culturally distinct learning styles (Frisby, 1993a, b). The same factors that account for achievement differences between siblings account for the differences between races (Rowe, 1994; Rowe, Vazsonyi, & Flannery, 1994, 1995). We can therefore use the evidence for *g*'s effect on individual differences in achievement to help predict the impact of average group differences in *g*.

Specifically, we might predict that standardized achievement gaps (d_{ach}) will be larger for forms of academic achievement that are more tightly linked to *g* at the individual level and larger for the groups with larger gaps in average IQ (d_{IQ}). Given the durability of *g*'s real-world effects and the stability of individual and group differences in *g*, we might also expect any *g*-based achievement gaps to be fairly stubborn.

Group Differences in Standardized Academic Achievement (d_{ach})

To what extent do group differences in academic achievement in elementary and secondary school actually follow the pattern predicted by individual differences in *g*? I will first examine whether group differences in achievement are stable over time and age, and then look at whether they are patterned as predicted across races and type of outcome.

Group Differences in Standardized Achievement (d_{ach}) Across Time. 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 ages 9, 13, and 17 with the same tests for over three decades. Table 4 shows standardized mean differences in reading, mathematics, and science achievement for blacks and Hispanics, relative to whites. If we assume that the content and psychometric properties of the three achievement tests have remained the same from year to year, then the achievement gaps for blacks and Hispanics were smaller in the two most recent decades than in the 1970s. Looking down the columns in Table 4 in order to compare the median of the effect sizes (d_{ach}) listed for the 1970s to the median of the effect sizes listed for the last two decades (the bracketed years in Table 4), it appears 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). This narrowing of achievement gaps occurred without any concomitant narrowing of IQ gaps. The already larger d_{ach} in science narrowed less for blacks (15%) and not at all (or grew slightly) for Hispanics. There was no discernible trend in NAEP performance during the last two decades. Other analysts have concluded that the earlier narrowing stopped before or during the 1980s and started to widen again during the 1990s (e.g., Grissmer, Flanagan, & Williamson, 1998; Hedges & Newell, 1998; Sadowski, 2001).

--- Tables 4 and 5 About Here ---

Group Differences in Standardized Achievement (d_{ach}) by Age/Grade. Looking across the columns in Table 4, the NAEP results reveal no clear trends across ages 9, 13, and 17 for the two minority groups studied (blacks and Hispanics). Table 5, which collates data from the 1961-1967 Prospects and NELS longitudinal studies of blacks and whites, tells the same story—no discernible change in d_{ach} for blacks over the twelve years of elementary and secondary school for achievement in reading, math, or vocabulary (cf. Phillips, Crouse, & Ralph, 1998). (The Prospects/NELS effect sizes are unaccountably smaller than those in the other large studies at that time; see Phillips et al., 1998, Table 7A-1.)

Group Differences in Standardized Achievement (d_{ach}) by Academic Subject. Standardized group differences in achievement—and changes in them—differ, however, by academic subject. Table 4 reveals the same pattern for blacks and Hispanics: the NAEP d_{ach} tend to be smallest in reading achievement, somewhat larger in math, and considerably larger in science. As Table 1 shows, this is also the pattern for black 12th graders in the 1965 Coleman data (where a test of general information replaces science), but the between-subject differences were not as pronounced for the Hispanics, Native Americans, or Asians in that study. And, as already noted, racial gaps narrowed more in NAEP reading than in math or science after the 1970s, for both blacks and Hispanics. Thus, group differences in achievement tended to be smallest and narrow the most in recent decades in the subject (reading) and population (blacks) most intensely targeted for intervention.

Similarity in Ordering of Groups in Effect Sizes for IQ (d_{IQ}) vs. Standardized Achievement (d_{ach}). The question here is, "Do the groups with the largest gaps in IQ also have the largest gaps in standardized achievement?" The 1966 Coleman Report is good for this purpose because it provides both IQ and achievement test scores for over 645,000 students in more than 3,000 schools. Despite its age, it also provides the most extensive data yet available for Native-Americans and Asians, in addition to blacks and Hispanics (with Mexican-Americans and Puerto Ricans tabulated separately). Table 1 shows effect sizes in both IQ and achievement for 12th graders. No matter which kind of IQ test (verbal, non-verbal) or achievement test (reading, math, general information) is considered, the order of d_{ach} across groups is the same: in descending size, black, Puerto Rican, Mexican-American, Native-American, and Asian. The more piecemeal data in the other tables are generally consistent with this ordering, too. Group differences in achievement track group differences in IQ.

Group Differences in Concordance of Effect Sizes for IQ (d_{IQ}) and Standardized Achievement (d_{ach}): Direct Comparisons. Because achievement gaps in some subjects decreased in earlier decades, we should next ask whether the achievement gaps for the various groups have become less concordant, or more concordant, with their enduring IQ gaps. The 1966 Coleman Report provides both IQ and achievement data for a large representative sample of American 12th graders in 1965. Shown in Table 1, the two sets of effect sizes suggest that there was rough concordance at that time between the average d_{IQ} for verbal and nonverbal IQ, on the one hand, and average d_{ach} for reading and math achievement, on the other: the respective mean d_{IQ} and d_{ach} were 1.11 vs. .98 (blacks), .82 vs. .78 (Hispanics), .66 vs. .67 (Native Americans), and .14 vs. .18 (Asians). The d_{ach} for the third test, general information, tend to be larger than those for reading and math and, at least for Hispanics, somewhat larger than the group's IQ gap. Many IQ batteries contain a general information subtest, so this test may have been the least curriculum related and more like an IQ subtest than the reading and math achievement tests.

--- Table 6 About Here ---

Table 6 allows the direct comparison of d_{IQ} and d_{ach} in a second study: a large sample of whites, blacks, and Hispanics in Grades 1-8 in one California school district in 1970. Its results are consistent with the data reviewed earlier in that the d_{IQ} for blacks and Hispanics are stable over the school years, the d_{IQ} are comparable in magnitude to the near-contemporaneous Coleman results, and the d_{ach} are stable from at least Grade 3 on. It differs in one regard, however: its d_{ach} for both blacks and Hispanics are considerably smaller in Grades 1 and 2 than in later grades. If this is a real difference, it raises the possibility that d_{ach} emerge and quickly reach concordance with IQ differences right after first exposure to formal schooling, an exposure that does tend, in fact, to produce a leap in academic development (Morrison, Smith, & Dow-Ehrensberger, 1995). We might also wonder whether the Grade 1 and 2 achievement tests were less *g* loaded than the others. If we recalculate the average d_{ach} without those two grades, then the effect sizes for IQ and achievement seem more concordant—1.09 vs. .74 (blacks) and .64 vs. .62 (Mexican-Americans)—though still less so for blacks than in the near-contemporaneous Coleman study for the nation at large.

Finally, Table 6 provides information about social class differences between the races: the white-black mean difference in family SES was much smaller than the groups' mean difference in IQ (.60 vs. 1.09), whereas the opposite was true for Mexican-Americans (1.26 for SES vs. .64 for IQ).

Group Differences in Concordance of Effect Sizes for IQ (d_{IQ}) and Standardized Achievement (d_{ach}): Indirect Comparisons. I am not aware of comparable datasets for examining degree of gap-concordance in later decades, and it is difficult to validly compare 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. The following analysis should therefore be considered only a first attempt to gain purchase on the issue. It will take a moment to explain but is conceptually simple. The question is, “Are the achievement gaps that we *observe* for a group commensurate with what we would *predict* from its IQ gap?”

If we posit that group mean differences in IQ are responsible in whole or part for the “achievement gap,” we can estimate a range within which observed d_{ach} should fall for a given d_{IQ}. The *maximum* possible achievement gap owing to IQ simply mirrors the IQ gap itself: so, if Group A’s observed d_{IQ} is .85, then the largest d_{ach} we would expect for that group in math or reading is also .85.

What is the *minimum* achievement gap we should expect a given IQ gap to create, all else equal? We might expect achievement gaps to be smaller when *g* accounts for less of the variability among individuals in school achievement. For example, the average correlation between IQ and spelling is only .42 but for math it is .63 (Column 2 in Table 3), so we might expect black-white achievement gaps to be smaller in spelling than math. By this reasoning, minimum expected d_{ach} in spelling and math could obtained by multiplying a group’s d_{IQ} (say, 1.0) by IQ’s correlations with spelling (.42) and math (.63)—the predicted minima thus being .42 and .63. This procedure will underpredict the observed achievement gap, however, unless corrections are made for two artifacts: first, IQ is an imperfect measure of *g* (thus understating an outcome’s correlation with *g*) and, second, measurement unreliability mutes true differences. The IQ correlations in column 2 of Table 3 were used to estimate each outcome’s correlation with *g* (column 3). Those correlations were, in turn, disattentuated for unreliability (see Footnote c in Table 3 for details). It is these doubly corrected correlations in column 4 that are multiplied by the d_{IQ} in the headings for columns 5-7 to get the minimum expected d_{ach} listed in those columns. (They are 4% larger than the uncorrected minima.) Minimum expected d_{ach} are provided for standardized tests of achievement in the 3R’s, plus spelling, based on their correlations with one of five major IQ test batteries. For purposes of comparison, analogous minimum expected d_{ach} are calculated when the "achievement outcome" is a composite IQ score on another of the four major IQ batteries or a score on one of the battery's own most "academic-like" (verbal) subtests.

Maximum and minimum predicted d_{ach} are provided for groups having IQ effect sizes of 1.20, .90, or .30 standard deviations from the white mean; the former two are the d_{IQ} documented for black and Hispanic 18-23-year-olds in the largest recent national study, namely, the 1980 ASVAB standardization sample (see the entry in Table 1 for Armed Forces Aptitude Batteries). The *maximum* predicted d_{ach} are, as noted before, simply the d_{IQ} for those groups. The *mimimum* predicted d_{ach} for the 3 R's range from .76-.84 for blacks (where d_{IQ}=1.20) and .57-.63 for Hispanics (where d_{IQ}=.90). They would be somewhat lower for Native-Americans if we assume that the group's d_{IQ} is somewhat less than .90. A d_{IQ} of |.30| might represent Asians, with +.30 for verbal ability and -.30 for non-verbal ability, the predicted minimum d_{ach} thus being |.19-.21|. The question, then, is whether the observed d_{ach} fall within these predicted ranges: roughly, .8-1.2 for blacks, .6-.9 for Hispanics, and |.2-.3| for Asians. If they fall clearly below the predicted minima, then other factors are *neutralizing* some of *g*'s usual influence; if the d_{ach} are clearly larger than the predicted maxima, then something else is *adding* to the achievement gaps besides *g*. For example, if the black d_{ach} in math were much larger than the maximum predicted—say, 2.0 rather than 1.2—then we would have to assume that something in *addition* to *g* must be depressing math achievement in that group relative to whites (perhaps discrimination or relatively worse instruction). On the other hand, if blacks lagged whites in math by only .2 *SD*, on the average, then we have to suspect they have some compensating advantage that is working to neutralize, in part, the impact of lower average IQ (e.g., higher average motivation, more suitable instruction).

I will begin by comparing the d_{ach} predicted for blacks and Hispanics with those observed for them on the NAEP. The NAEP data are good for this purpose because they are extensive as well as psychometrically sound. Table 4 provides the achievement gaps observed in the last three decades for blacks and Hispanics, relative to whites, at ages 9, 13, and 17 in reading, math, and science. A median d_{ach} is provided for the bracketed years for all ages. Table 4 shows that the median observed d_{ach} 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 and .90), and only a third of the way from the maxima toward the minima expected for achievement tests in the 3 Rs (.76-.84 for blacks, .57-.63 for Hispanics). The d_{ach} were close to the minima expected for performance on another IQ test/subtest (1.04-1.15 for blacks, .78-.86 for Hispanics).

After 1980, the d_{ach} for blacks in NAEP reading achievement (.79 median) fell to the minimum expected (.77 language, .82 reading) but not quite that far for Hispanics (.66, where the minimum expected is .62 in reading and .58 in language). The d_{ach} in math moved one-half (Hispanics) to three-quarters (blacks) the way toward the expected minima, but the d_{ach} in science remained near the maximum predicted for both racial-ethnic groups. The d_{ach} in the 1966 Coleman study (Table 1) were near the maxima predicted for all races. In the 1961-1967 Prospects and NELS studies, which are combined in Table 5, the d_{ach} for blacks were at or somewhat below the minimum expected (.68-.77). The direct comparisons of IQ and achievement gaps in the Coleman (Table 1) and California studies (Table 6) also support the conclusion that all groups' past and present achievement gaps are broadly consistent with their respective IQ gaps. There is a tendency, however, for blacks more than others to be nearer their minimum predicted d_{ach}.

Although standardized achievement gaps thereby seem tethered to IQ gaps, the former seem somewhat more elastic than the IQ gaps. Not only have the d_{ach} narrowed over time in some subjects, but they have narrowed most in the subjects most intensely targeted by educational reforms: reading more than math, and math more than science or general information. That the observed achievement gaps did not exceed the maximum predicted gaps suggests that *g* may be the primary or only cause of group disparities in standardized academic achievement. That none of the achievement gaps fell materially below the *g*-predicted minima may signal a natural lower bound for feasible reductions in d_{ach}, absent any reductions in d_{IQ}.

We should also investigate the source of any narrowing of achievement gaps before unreservedly celebrating them, because they may not always be effected by unambiguously good causes (e.g., some shifts in educational policy may cause bigger drops in achievement among higher-performing students). We likewise should not automatically disparage the widening of achievement gaps, because they may result from otherwise desirable processes (e.g., improving instruction for all students generally increases both the mean and variance in achievement).

Group Differences in Achievement (d_{ach}) Expressed on a Common Cross-Grade Proficiency Scale. The practical significance of standardized achievement gaps is brought home more concretely by showing group differences in mean proficiency levels in each grade on a common cross-grade metric. 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 their 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 two years behind whites in achievement during the elementary grades. The ETS Growth Study in Table 7 shows that the absolute gaps in achievement are considerably larger than this by the middle and high school years. Black 11th

graders exhibited mean proficiency levels in math and listening that were comparable to those of white 7th graders, and in other skills (interpolating results) they performed at about the level of white 8th graders. The NAEP data in Table 8 show the same general pattern in the latest available (1999) NAEP proficiency scores in reading, math, and science at ages 9, 13, and 17 for whites, blacks, and Hispanics. At age 13, blacks trail whites in reading and math achievement by three grade levels; by age 17 they trail by four years. The gaps are typically one year smaller for Hispanics at both ages. The gaps in mean proficiency level are larger yet in science, as was suggested earlier by the larger d_{ach} in science. In 1999, blacks trailed whites in science achievement by at least five grade levels during both middle and high school; at age 17, Hispanics trailed whites by three grade levels. The NAEP proficiency gaps were somewhat larger in the 1970s (Smith et al., 1995, pp. 54, 58, 60), as would be expected from their larger d_{ach} at that time. These data show how the same standardized gap in ability produces larger absolute shortfalls in school achievement in later grades. No new factors need be introduced to explain why groups with lower average cognitive ability levels seem to fall further behind in the higher grades.

--- Tables 7 and 8 About Here ---

Group Differences in Standardized Achievement (d_{ach}) Within Social Classes. Finally, we might ask whether the pattern of group IQ differences within the same social class is replicated by achievement gaps within social classes. Table 9 reveals the by-now familiar pattern in such studies, first for social class origin and then for social class destination. The top panel shows that black children of college-educated parents perform the same in NAEP reading achievement, on average, as do white children whose parents dropped out of high school; Hispanic children of college-educated parents perform about the same as white children whose parents completed high school but no more. This is comparable to the results for IQ in Table 2, which showed that black children with parents holding jobs in the highest several strata score no better on IQ tests than do white children with parents working in the lowest several strata of occupations.

-- Table 9 About Here --

The middle panel of Table 9 shows race and SES differences in mean levels of functional literacy on the U.S. Department of Education's National Adult Literacy Survey (NALS). NALS literacy levels among college-educated blacks are comparable to the proficiency of whites who obtained no more than a high school diploma. The four-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. Once again, Hispanics score somewhat better than blacks, but still lag comparably-educated whites by several grade equivalents in NALS literacy. This pattern in adult literacy is also consistent with the results for adult IQ in Table 2, which showed blacks with at least some college scoring somewhat lower in IQ (95.9) than whites who had dropped out of high school (98.0).

The bottom panel of Table 9, for college-bound high school students, 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 $20,000-35,000 (SAT total scores of 1006 vs. 1010). Within-class SAT gaps relative to whites were smaller for Hispanics and 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.

<u>Group Differences in Placement, Persistence, and Selection</u>

The achievement gap is often described as highly general because it pervades all educational outcomes, not just tests of standardized academic achievement. There are large group differences, for instance, in being placed in special education or classes for the gifted, dropping out of high school, and completing college. These disparities in placement and advancement are commonly discussed as part of the "pipeline" problem, namely, that groups do not pass in equal proportion through the successive filters governing educational and occupational advance. The proffered explanations for the pipeline problem are as varied as the studies of it are numerous. But what patterns would we forecast from the foregoing gaps in IQ and standardized achievement?

Figure 2 schematizes the question. The figure's upper half summarizes previous analyses of how the odds of various life outcomes shift along the IQ continuum, specifically, for typical training potential, typical occupation level, and (for whites) dropout rates along five successive segments of the IQ continuum (see Gottfredson, 1997, for the derivation of and additional detail on Figure 2). The lower half of Figure 2 provides estimated cumulative percentages of whites, blacks, Hispanics, and Asians at each 5-point step along the continuum of normal IQ (70-130). The estimates for whites and blacks are good, reasonable for the conglomerate "Hispanic" category, but more questionable for Asians. Bell curves are shown only for blacks and whites; the curves for Hispanics and Native-Americans would be centered roughly five points higher on the IQ continuum than the curve for blacks, and the curve for Asians up to five points above that for whites.

The figure illustrates two major implications of average group differences in *g* and in its near-twin, standardized academic achievement. First, the representation of lower-scoring groups (blacks, Hispanics, and Native-Americans), relative to whites, shifts from overrepresentation at the left tail of the IQ distribution to underrepresentation at the right tail. Disparities in representation are larger for groups separated further from whites on the IQ continuum. So, for example, disparities in representation above IQ 100 relative to Anglo-whites are largest for blacks (1:3), smaller for Hispanics (1:2), and yet smaller (and reversed) for Asians (6:5). Second, disparities in representation are most striking at the tails of the IQ distribution. For instance, whereas blacks are overrepresented by less than 3:1 (41% vs. 18%) between IQs 76-90, 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 110-125 to 1:30 above IQ 125. The disproportions for Hispanics (and probably Native-Americans too) 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 reversed for that group: *under*representation below IQ 75 (2:3) and *over*representation at the highest levels (2:1). These gradients of over- and underrepresentation at different levels of IQ establish baselines for what we might expect for achievement if *g* were the primary or only factor causing group differences in rates of passing through certain gates in the pipeline. *g* is not the only influence on advancement, of course, so deviations of observed from *g*-predicted rates of progress can provide clues to any other factors at work.

--- Figure 2 and Table 10 About Here ---

Placement and Ability Grouping. Non-Asian minority students are usually overrepresented in special education classes—blacks, for example, often by three to one. As school psychologists know well, such disparities are often litigated as evidence of racial bias in placement. Figure 2 shows, however, that these racial differences in rates of placement in special education are smaller than IQ gaps alone would predict. IQ 70-75 is often considered the upper threshold for mild mental retardation, and proportionately five times as many blacks and four times as many Hispanics fall below this threshold as do whites. (See also Gordon, 1980a, for an extended analysis.)

Actual performance in the classroom is usually more important than IQ test results in recommending placements, whether in special education, curriculum tracks, or the like. Table 10 therefore provides rates, by race, of various forms of academic progress, or lack thereof, that might affect such placements: specifically, rates of performing below acceptable levels on criterion-referenced tests of academic skill and knowledge, performing at a moderately high level on such tests, and failing to get a high school degree vs. obtaining a high school or college diploma by age 25-29. As Table 10 shows, rates for all non-Asian minorities are elevated on all indicators of failure to make satisfactory progress, and they are depressed on all indicators of high-level achievement. Asian rates are similar to those for whites, sometimes better and sometimes worse. This consistent pattern of differential progress by race reflects the generality of achievement gaps, but the question here is more specific: do the *magnitude* and *patterning* of these differences in progress and placement accord with what would be expected from group differences in *g* alone? That is, are differences across groups in these academic outcomes *commensurate* with their IQ differences (Gordon, 1997)? For example, do the group differences in rates of failing to reach some minimally satisfactory level of performance in reading or math mirror group differences in rates of falling below some specific IQ threshold?

The top panel of Table 10 shows the percentages of white, black, Hispanic, and Asian 4th and 12th graders who scored below the "basic" level expected of their grade on recent NAEP achievement tests in reading, math, and science. "Basic" is defined as "partial mastery of prerequisite knowledge and skills that are fundamental for proficient work at each grade" (Donahue, Voelkl, Campbell, & Mazzeo, 1999, p. 9). Table 10 shows that from 21-27% of the whites, 64-68% of blacks, and 58-60% of Hispanics *failed* to reach the "basic" level in Grade 4 core subjects. Looking at Figure 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. Figure 2 also shows that blacks and Hispanics are overrepresented below IQ 90 by a factor of almost three to one for blacks and over two to one for Hispanics. Thus, even in the unlikely event that 4^{th} graders were placed into special education randomly from among children below the 25^{th} 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 placement, these data reveal a huge disparity in preparedness for academic progress. By this measure, black 4^{th} graders are twice as likely to lack as to possess the minimum prerequisites for developing proficiency at their grade level.

In the 12^{th} 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). Turning to an indicator of good progress, the percentages of whites, blacks, and Hispanics who meet or exceed the "proficient" criterion line up near the 50^{th}–63^{rd} percentile in IQ for reading (that is, above IQ 100-105) and with the 75^{th} 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 at or *below* a particular IQ level in Figure 2.) Rates for Asians once again oscillate around those for whites, tending to be more favorable in math and science by Grade 12, but always less favorable in reading. The rates of poor and good NAEP performance thus show commensurability with the groups' IQ differences. The ratios of students performing poorly to those achieving at high levels rest primarily on where the threshold for proficiency in a subject lines up on the IQ continuum: in the 12^{th} grade, it ranges from the 50^{th} for reading to the 75^{th} for science. The ratios of low to high performance are therefore much more favorable in reading than science and in the groups with the higher average IQs: for instance, 1:3 and 5:3 for whites, respectively, in reading and science; 4:3 and 10:1 for Hispanics; and 2:1 and 26:1 for blacks.

Classes for the gifted and talented are likewise perennially under fire for underrepresenting non-Asian minorities. Selection into gifted classes has often required, at least in years past, scoring well on an intelligence test (e.g., above IQ 130). The predictable result is that blacks and Hispanics are rarely selected. Underrepresentation would still be large for blacks (1:3) and Hispanics (1:2) even if the entrance criterion were lowered to IQ 100. Achievement tests produce the same pattern of underrepresentation, as just illustrated by the results for NAEP proficiency. For instance, if entry to gifted classes were set at the NAEP "proficient" level in math and science (about the 75^{th} percentile in ability), blacks would be underrepresented by at least 1:5 and Hispanics by 1:3 in both Grades 4 and 12.

Children within the greater normal range of intelligence (IQs 70-130) 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, grouping and tracking have been equally contentious (Kulik, 2003). Figure 2 reveals the dilemma as it relates to race. If we assume that college-preparatory classes in high school recruit students primarily of IQ 90 and above (i.e., above the 25^{th} percentile of the population), then the eligible pools among blacks and Hispanics are, respectively, one-half and two-thirds as large as for those for whites (41% and 53%, respectively, vs. 78% for whites; and 86% for Asians). Recall that this would include all students performing merely at or above the "basic" level (partial mastery of a subject's prerequisites). If honors and AP classes draw students primarily from IQ 110 and above (above the 75^{th} percentile for the general population), then the pools of black and Hispanic eligibles drop in relative size to one-seventh for blacks (4% blacks vs. 28% whites) and to one-third for Hispanics (10% Hispanics vs. 28% whites; and 40% for Asians) relative to whites. Other factors affect track placement, but they would have to favor blacks and Hispanics heavily over whites and Asians to begin to equalize group representation across tracks when assignment is sensitive to *g* or school achievement.

Persistence. Figure 2 illustrates the strong relation between IQ and proportion of young white adults who drop out of high school: 55% of those below IQ 75 (5th percentile of general population), 35% for IQ 76-90 (5th-25th percentile), but essentially zero beyond IQ 110 (75th percentile). We saw in Table 10 that rates of whites, blacks, and Hispanics performing below the NAEP basic level in Grades 4 and 12 were commensurate with their percentages falling below certain IQ levels. The bottom panel of that table reveals, however, that the rate of 25-29-year-olds who are high school dropouts is better for blacks relative to whites, and far worse for Hispanics—12.4% (blacks), 37.6% (Hispanics), and 7.0% (Anglo-whites)—than would be predicted on the basis of their disparities in IQ or achievement. This incommensurability between dropout rates and IQ/achievement illustrates the long-known fact that blacks tend to go further in school, and Hispanics less far, than whites of the same ability level (hence the black-white IQ differences in Table 2 at given levels of education).

The racial disparities in rates of college graduation likewise bear little relation to the groups' gaps in IQ and standardized achievement, with Hispanics again graduating at far lower rates than blacks despite their higher average IQ and achievement levels: 18.0% (blacks), 8.9% (Hispanics), and 35.9% (Anglo-whites). If we look at the ratio of high school dropouts to college graduates in the three populations, the disparities are large: 1.5:l (blacks), 4:1 (Hispanics), and 1:5 (Anglo-whites). The lack of IQ-commensurability for persistence rates signals that there are important additional, non-cognitive factors that enhance persistence among blacks and depress it among Hispanics, relative to whites of the same ability level. It is relevant in this regard to note that much higher proportions of Hispanic (19.3%) and Asian students (27.0%) than Anglo-whites (1.4%) and blacks (3.1%) are foreign born (U. S. Census Bureau, 2002, p. 138).

Selection for College and Occupational Training. A look at post-secondary education and training illustrates why the difficult challenges that group differences in cognitive ability pose for schools follow graduates in their travels through the institutions of adult life. Admission to college or graduate and professional school generally involves taking a standardized test such as the SAT, ACT, GRE, GMAT, or MCAT. Although these tests are taken primarily by the more able members of any group, they all yield large effect sizes for blacks (.82-1.11), somewhat smaller ones for Hispanics (.50-1.00), and, for Asians, scores more comparable to whites (-.46 to .25; Camara & Schmidt, 1999, p. 2, using data from 1997-1998). A recent meta-analysis of SAT, ACT, and GRE scores since 1970 (Roth, Bevier, Bobko, Switzer, & Tyler, 2001, pp. 317, 319) likewise documents large d_{ach} at the gateway to both undergraduate and graduate education: for SAT and ACT total scores, d_{ach} were .97 and 1.02 for blacks and .77 and .56 for Hispanics; for GRE total, they were 1.34 for blacks and .72 for Hispanics. (These *d* cannot be directly compared to the others reported here, because college students are restricted in range on IQ.)

Whenever groups pass through an educational gate with unequal skills, those initial disparities in skill result in differential achievement further down the pipeline. The reason is that differences in *g* remain a strong predictor of actual performance (see Table 3). For instance, Camara and Schmidt (1999, p. 5) reported college freshman GPAs of 2.66 for whites and 2.14, 2.37, and 2.80, respectively, for blacks, Hispanics, and Asians. NCAA Division I six-year graduation rates were 56%, 38%, 45%, and 63%, respectively, for the four racial-ethnic groups (Camara & Schmidt, 1999, p. 7). These differences in academic success track the groups' initial differences in skill level, with blacks faring worst and Asians best. Analogous disparities in career progress (job performance ratings and promotion rates) are observed in employment settings when groups enter jobs with unequal ability levels, on the average.

Selection for training and employment in high-volume entry-level jobs also often requires passing a standardized test of some sort. Military selection and placement always does. Great difficulties in training and utilizing low-aptitude men during World War II led the U.S. Congress to ban the induction of persons below the 10th percentile in AFQT ability (about IQ 80); none of the military services has inducted recruits below the 16th percentile (IQ 85) for many years; and some of the services enlist only from considerably higher percentiles. “Trainability” is not the only selection criterion, but it is an essential one. Largely because of group disparities in cognitive ability, much lower proportions of blacks and Hispanics are eligible for enlistment. Among males aged 18-23, the percentages eligible to enlist in the Army, Navy, Marine Corps, and Air Force are estimated to be, respectively, 84, 82, 82, and 71% for whites; 41, 41, 34, and 21% for blacks; and 53, 51, 48, and 38% for Hispanics (Eitelberg, 1988, pp. 122-125). The percentages eligible for the most selective military occupations are even more discrepant across the groups: 28-36% for white males, 3-5% for blacks, and 7-13% for Hispanics. Looking at Figure 2, these eligibility percentages are commensurate with the percentages of these groups above IQ 105-110. (Relatively fewer females are eligible, especially for the Marines, owing to fewer women meeting the physical standards.)

The military’s official minimum cognitive ability threshold is comparable to the lower boundary of IQs routinely observed among *applicants* to the lowest-level civilian jobs: for instance, three-quarters of applicants for packer, material handler, janitor, assembler, and food service worker score above the equivalent of IQ 80 on the Wonderlic Personnel Test (Gottfredson, 1997, pp. 88-89). Again mirroring the civilian sector, the military’s most selective occupations require IQs comparable to the 25th percentile of *applicants* to the highest-level civilian jobs (IQ>110), such as systems analyst, chemist, engineer, and attorney. Thus, although the military is sometimes viewed as the employer of last resort, by official policy it draws recruits for its lowest- and highest-level jobs from the same segments of the IQ distribution (respectively, above IQs 80 and 110) as do civilian employers.

Incumbents tend, however, to have higher IQs than do applicants to a job. IQ 115 is near the *minimum* (and IQ 125-130 is the average) found among persons employed as physicians, lawyers, professors, engineers, high-level executives, and other such professionals. High *g* cannot be dismissed as an arbitrary and thus dispensable barrier to such employment, because *g* level is an increasingly important predictor of performance in successively higher level jobs. The ratios in Figure 2 show that the proportions of blacks and Hispanics above IQ 115 are respectively one-ninth and one-third as large as for whites (2% and 6% vs. 18% for whites; and 27% for East Asians). As is the case with educational credentials, actual levels of employment are more nearly equal in these occupations than the group IQ gaps would forecast (Gottfredson, 1986). The disparities in representation are nonetheless still quite large and have spawned much litigation (Sharf, 1988). The pattern of successively worse underrepresentation of lower-scoring minority groups further up the educational and occupational ladders is disturbing and causes much social turmoil. But it is also entirely predictable from the groups' disparities in *g*.

CHALLENGE FOR DIVERSE SOCIETIES

The "achievement gap" originates mostly in group disparities in general cognitive ability. Research has not yet established why the disparities exist, but their past resistance to change warns us that they will likely be with us for some time to come. What does this mean for schools?

The Democratic Dilemma

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

All racially diverse societies face this conflict, as Klitgaard (1985, 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. In fact, the underlying dilemma is shared by *all* societies, even racially homogeneous ones, because all human populations exhibit wide dispersion in *g*. Differences in *g* reflect differences in the ability to profit from instruction, learn on one's own, figure things out, and the like. These generic higher-order thinking skills are useful in all aspects of life, but perhaps especially in schools. When everyone is given equal opportunities to learn, higher-*g* individuals will learn more effectively and efficiently with the same level of effort. Equalizing learning opportunities in diverse societies therefore does not eradicate achievement gaps, it assures them. Conversely, equalizing outcomes across groups requires favoring members of lower-scoring groups with more opportunities and assistance than members of higher-scoring groups.

The Schools' Unspoken Challenge

American schools are expected to provide an equal education to a population whose members (and some subgroups) differ greatly in their ability to profit from common instruction. To even acknowledge this challenge—namely, that the democratic dilemma is created by the differences among us—is to risk being blamed for creating or desiring social inequality. That teachers (or citizens) even *notice* average group differences in ability and achievement sometimes earns them the epithet "prejudiced" or worse (including from social scientists; see Gordon, 1997, on national surveys of "racist" beliefs). Like personnel selection specialists, school psychologists have often felt compelled by their employers to suppress politically inexpedient evidence about ability differences. 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.

Schools clearly can affect whether their student bodies learn a little or a lot, on the average. Currently, however, they have no means of narrowing differences in *g* either within or between demographic groups, and they therefore have only limited ability to narrow gaps in academic achievement. Their choices rest primarily in *where* in the educational process they allow achievement gaps to become visible. As I illustrate next, they can shuffle them around, disguise them, or deny them, but they cannot eliminate them (cf. a similar analysis by Jensen, 1991).

To begin, consider that many schools differentiate their curricula in order to better target instruction to students' intellectual needs. Lacking the means to individualize instruction, schools deliver differentiated curricula to students grouped by ability or achievement—for instance, via classes for exceptional children, within-class or whole-class grouping for reading and math in elementary school, and different curriculum tracks in high school. When mixed-race schools deliver instruction differentiated in this manner, there result large racial disproportions in assignment to instructional groups. Such disproportions have become politically unacceptable, and grouping systems are often assailed as elitist, discriminatory tools for sustaining social inequality. They are said to "resegregate the races within desegregated schools."

One response by schools has been to maintain racial proportionality or near-proportionality in assignment to differentiated curricula by selecting equal proportions from all races. Racially proportionate assignment to gifted classes is sometimes achieved, for example, by race-norming test scores, which means selecting the top-scoring X% of all races (e.g., Richert, 2003). Many employers likewise used race-norming to assure equal pass rates on their employment tests, whatever the racial gaps in actual skill level, until Congress outlawed the practice in 1991. But this practice does nothing to neutralize the effects of racial gaps in cognitive ability, which will often average one *SD* for blacks and almost that much for Hispanics. The practice therefore guarantees that, once on the job or in the classroom, the beneficiaries of lower entry standards will fail at high rates and constitute the majority of poor performers if performance standards are maintained. The appearance of achievement gaps is thus merely postponed, but perhaps at greater personal cost to the ill-placed individuals. The reemergent achievement gaps are, once again, politically unacceptable.

A second response by schools has been simply to eliminate curricular differentiation and to assign all students to racially-mixed ability-heterogeneous classrooms. Advocates refer to this as desegregating the resegregated school. It is simple enough administratively to integrate a school racially and socioeconomically by abolishing ability groups and tracks, but, once again, this action does nothing to neutralize the differences in *g* that inevitably create achievement gaps. It only moves the achievement gap elsewhere. The problem is that de-grouping and de-tracking place into the same classroom students who differ by many grade equivalents in achievement level and in their ability to profit from further instruction, regardless of race. Recall that blacks and Hispanics tend to lag two grade equivalents behind whites and Asians by the end of elementary school and by four grade levels toward the end of high school. When the different racial groups in a school manifest achievement gaps of this size, as usually seems the case, then assignment to ability-heterogeneous classes—like race-norming—will guarantee highly noticeable racial differences in academic performance *within* classrooms, especially in hierarchical subjects such as math. Most of the failing students will be black and Hispanic. As before, this result is politically unacceptable.

Gifted programs are prized as well as despised, so even de-tracked schools may retain their gifted programs by "democratizing" access to them. They do this by adopting broader, less intellectual definitions of gifted and talented and more open methods of referral (including self and parent nominations), in effect making selection more random with regard to *g*. As both educational and personnel psychologists have documented, near randomization with regard to *g* or achievement is the only way, when race-norming is not available, to approach racial parity in higher-level assignments. To avoid high failure rates and the emergence of achievement gaps, the tasks that students perform must themselves be stripped of most *g* demands. This is not a viable option in mid- to high-level jobs, of course, where workers must reliably carry out an organization's work, but it appears to be an option pursued by schools today. This is evidenced by the intentionally reduced rigor of Advanced Placement courses and many programs for the gifted (Bleske-Rechek, Lubinski, & Benbow, in press; Colangelo & Davis, 2003). Denuding these programs of their intellectual challenge vitiates the educational purpose of such classes, of course, which is to challenge and nurture the talents of highly gifted students who are not served well by the regular curriculum.

Returning to the regular curriculum, advocates of de-tracking are not blind to the challenges of ability differences, so they offer new technologies for educating differently able children within the same classroom. Some, 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, "succeed" primarily by deflecting high-achieving students from additional learning into helping low-achieving classmates improve (Robinson, 2003). Cooperative learning essentially assigns brighter students responsibility for remediating the achievement gaps in their classroom, which many educators of the gifted label exploitation.

Such methods for redistributing achievement gains within classrooms may show up as success in district statistics, but the private costs are not lost on the parents of students whose achievement has been capped in the name of educational improvement. 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. Some efforts to close achievement gaps within schools thereby have the effect of widening the gaps *between* schools.

Schools presently have no feasible, ethical way to prevent the emergence of distressingly large achievement gaps, and many of their attempts at doing so produce unwelcome side-effects. To close the achievement gaps between groups without first closing their IQ gaps would require all black and Hispanic students to perform as well, on the average, as white and Asian classmates who, on the average, are about one *SD* higher in IQ, where one *SD* represents fully one-quarter of the range of normal IQ. For example, blacks of somewhat-below-average IQ (IQs 75-90) would have to perform as well as whites of average ability (IQs 90-110). In like manner, blacks who are somewhat above average in IQ (IQs 110-125) would have to match the performance levels of gifted whites and Asians (IQ>125). Flagellating one group or another for lack of will or commitment will have 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.

The Human Challenge

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 of the same social class is no larger than the mean difference between biological siblings growing up in the same household—about 12 IQ points. The daily challenges faced by persons in the "high risk" or "up-hill battle" ranges of IQ (see Figure 2) are difficult, whatever their race. There is no race-specific cure for low achievement. Likewise, the socioeducational integration and fair treatment of individuals of disparate ability levels is always a challenge, whatever the race, even within the same family.

All children can learn, and perhaps many can learn far more than they do now in American schools. There is certainly no call to "give up" on anyone. What all students require is ability-appropriate instruction, not identical instruction or race-specific curricula. Lower-*g* students, like low-*g* military recruits (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 (Snow & Lohman, 1984). No other aptitude-treatment-interaction (ATI) has ever been verified (Reschly, 1997; Snow, 1996). As noted earlier, we know of no way to equalize rates of learning. *Amount* learned is more manipulable, but amount learned can be equalized only by stalling the progress of brighter students while helping the less able to catch-up. The more we improve instruction for students of all ability levels in all races, the greater the variance in amount learned will become.

It is possible to raise *relative* achievement selectively among lower-scoring groups, but it comes at the cost of neglecting or avoiding such improvement among higher-scoring ones. Some might view this as essential to social justice and racial harmony, but it is only one among other choices. None will satisfy everyone, because they all involve difficult tradeoffs between equal opportunity and equal results, between individual excellence and group parity. Human dispersion in *g* means that schools and societies will always face hard choices. Unfortunately, school psychologists might justifiably feel caught in the middle.

References

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.

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

Belluck, P. (1999, July 4). "Reason is sought for lag by blacks in school effort." *New York Times*.

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

Borman, G. D., Stringfield, S., & Rachuba, L. (2000, February). *Advancing minority high achievement: National trends and promising programs and practices. A report prepared for the National Task Force on Minority High Achievement*. New York: The College Board.

Brown, R. T., Reynolds, C. R., & Whitaker, J. S. (1999). Bias in mental testing since *Bias in Mental Testing*. *School Psychology Quarterly, 14*(3), 208-238.

Camara, W. J., & Schmidt, A. E. (1999). *Group differences in standardized testing and social stratification*. College Board Report No. 99-5. New York: College Entrance Examination Board.

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. http://nces.ed.gov/nationsreportcard//pdf/main1994/96045.pdf

Canivez, G. L., & Watkins, M. W. (1999). Long-term stability of the Wechsler Intelligence Scale for Children—Third Edition among demographic subgroups: Gender, race/ethnicity, and age. *Journal of Psychoeducational Assessment, 17*, 300-313.

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 and G. A. Davis (Eds.), *Handbook of gifted education (3rd ed.)* (pp. 3-10). Boston: Allyn and 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 (1999a). *Counselor's handbook for the SAT program, 1999-2000*. Princeton, NJ: The College Board and ETS.

College Board (1999b). *Reaching the top: A report of the national task force on minority high achievement*. New York: College Board Publications.

Crouse, J. (1979). The effects of academic ability. *Who gets ahead? The determinants of economic success in America* (pp. 85-121). New York: Basic Books.

Deary, I. J. (2000). *Looking down on human intelligence: From psychometrics to the brain*. Oxford: 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*. U. S. Department of Education, Office of Educational Research and Improvement.

Dreger, R. M. (1973). Intellectual functioning. In K. S. Miller & R. M. Dreger (Eds.), *Comparative studies of blacks and whites in the United States.* New York: Seminar Press.

Eitelberg, M. J. (1981). *Subpopulation differences in performance on tests of mental ability: Historical review and annotated bibliography*. Technical memorandum 81-3. Alexandria, VA: Office of the Secretary of Defense, Directorate for Accession Policy.

Eitelberg, M. J. (1988). *Manpower for military occupations*. Alexandria, VA: Office of the Assistant Secretary of Defense, Force Management and Personnel.

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

Frisby, C. L. (1993a). "Afrocentric" explanations for school failure: Symptoms of denial, frustration, and despair. *School Psychology Review, 22*(3), 568-577.

Frisby, C. L. (1993b). 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. *School Psychology Quarterly, 14*(3). (Special issue)

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

Gordon, R. A. (1980a). Examining labelling theory: The case of mental retardation. In W. R. Gove (Eds.), *The labelling of deviance: Evaluating a perspective (2nd ed.)* (pp. 111-174). Beverly Hills: Sage.

Gordon, R. A. (1980b). Implications of valid (and stubborn) IQ differences: An unstatesmanlike view. *Behavioral and Brain Sciences, 3*(3), 343-344.

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

Gottfredson, L. S. (1986). Societal consequences of the g factor in employment. *Journal of Vocational Behavior, 29*, 379-410.

Gottfredson, L. S. (1994). The science and politics of race-norming. *American Psychologist, 49*(11), 955-963.

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). The science and politics of intelligence in gifted education. In N. Colangelo & G. A. Davis (Eds.), *Handbook of gifted education (3rd ed.)* (pp. 24-40). Boston: Allyn & Bacon.

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.

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.

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

Hilton, T. L., Beaton, A. E., & Bower, C. P. (1971). Stability and instability in academic growth—a compilation of longitudinal data. Princeton, NJ: Educational Testing Service.

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

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

Jencks, C., & Phillips, M. (1998). The black-white test score gap: An introduction. In C. Jencks & M. Phillips (Eds.), *The black-white test score gap* (pp. 1-51). Washington, DC: Brookings Institution Press.

Jensen, A. R. (1974). Equality for minorities. In H. J. Walberg (Ed.), *Evaluating educational performance* (p. 175-222). Berkeley, CA: McCutchan.

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

Jensen, A. R. (1981). *Straight talk about mental tests*. New York: Free Press.

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.

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.

Keith, T. Z, & Witta, E. L. (1997). Hierarchical and cross-age confirmatory factor analysis of the WISC-III: What does it measure? *School Psychology Quarterly, 12*(2), 89-107.

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. (1985). *Choosing elites*. New York: Basic Books.

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.

Kranzler, J. H., & Keith, T. Z. (1999). Independent confirmatory factor analysis of the Cognitive Assessment System (CAS): What does the CAS measure? *School Psychology Review, 28*(1), 117-144.

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* (p. 36-52). (Technical Memorandum 82-2) Alexandria, VA: Office of the Secretary of Defense, Directorate of Accession Policy.

Lee, F. R. (2002, November 30). "Why are black students lagging?" *New York Times*.

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

Lynn, R. (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.

Morrison, F. J., Smith, L., & Dow-Ehrensberger, M. (1995). Education and cognitive development: A natural experiment. *Development Psychology, 31*, 789-799.

National Center for Education Statistics (2000, August). Results over time—NAEP 1999 long-term trend summary data tables. http://nces.ed.gov/nationsreportcard/tables/Ltt1999/ (accessed January 2003).

Neisser, U., Boodoo, G., Bouchard, T. J., Boykin, A. W., Brody, N., Ceci, S. J., Halpern, D. F., Loehlin, J. C., Perloff, R., Sternberg, R. J., & Urbina, S. (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*. <http://www.inmotionmagazine.com/er/pnrp1.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: Volume 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 III, J. F., & 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., 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.

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.

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.

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 Social 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 and G. A. Davis (Eds.), *Handbook of gifted education (3rd ed.)* (pp. 146-158). Boston: Allyn and Bacon.

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

Rowe, D. C. (1994). No more than skin deep. *American Psychologist, 49*, 215-216.

Rowe, D. C., Vazsonyi, A. T., & Flannery, D. J. (1994). No more than skin deep: Ethnic and racial similarity in developmental process. *Psychological Review, 101*, 396-413.

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. (in press). Thirty years of research on Black-White differences in cognitive ability. *Psychology, Public Policy, and Law*.

Sackett, P. R., Schmitt, N., Ellingson, J. E., & Kabin, M. B. (2001). High-stakes testing in employment, credentialing, and higher education - Prospects in a post-affirmative-action world. *American Psychologist, 56*(4), 302-318.

Sadowski, M. (2001, May/June). Closing the gap one school at a time. *Harvard Education Letter*. http://www.edletter.org/past/issues/2001-mj/gap.shtml

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

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.

Sharf, J. C. (1988). Litigating personnel measurement policy. *Journal of Vocational Behavior, 33*(3), 235-271.

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

Smith, T. M., Perie, M., Alsalam, N., Mahoney, R. P., Bae, Y., & Young, B. A. (1995). *The condition of education, 1995*. Washington, DC: U. S. Department of Education, Office of Educational Research and Development.

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

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

Sticht, T. (Ed.) (1975). *Reading for working: A functional literacy anthology*. Alexandria, VA: Human Resources Research Organization.

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.

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 http://www.census.gov/population/socdemo/education/tableA-2.pdf (Accessed April 10, 2003).

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

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.

Vernon, P. E. (1982). *The abilities and achievements of Orientals in North America*. New York: Academic Press.

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

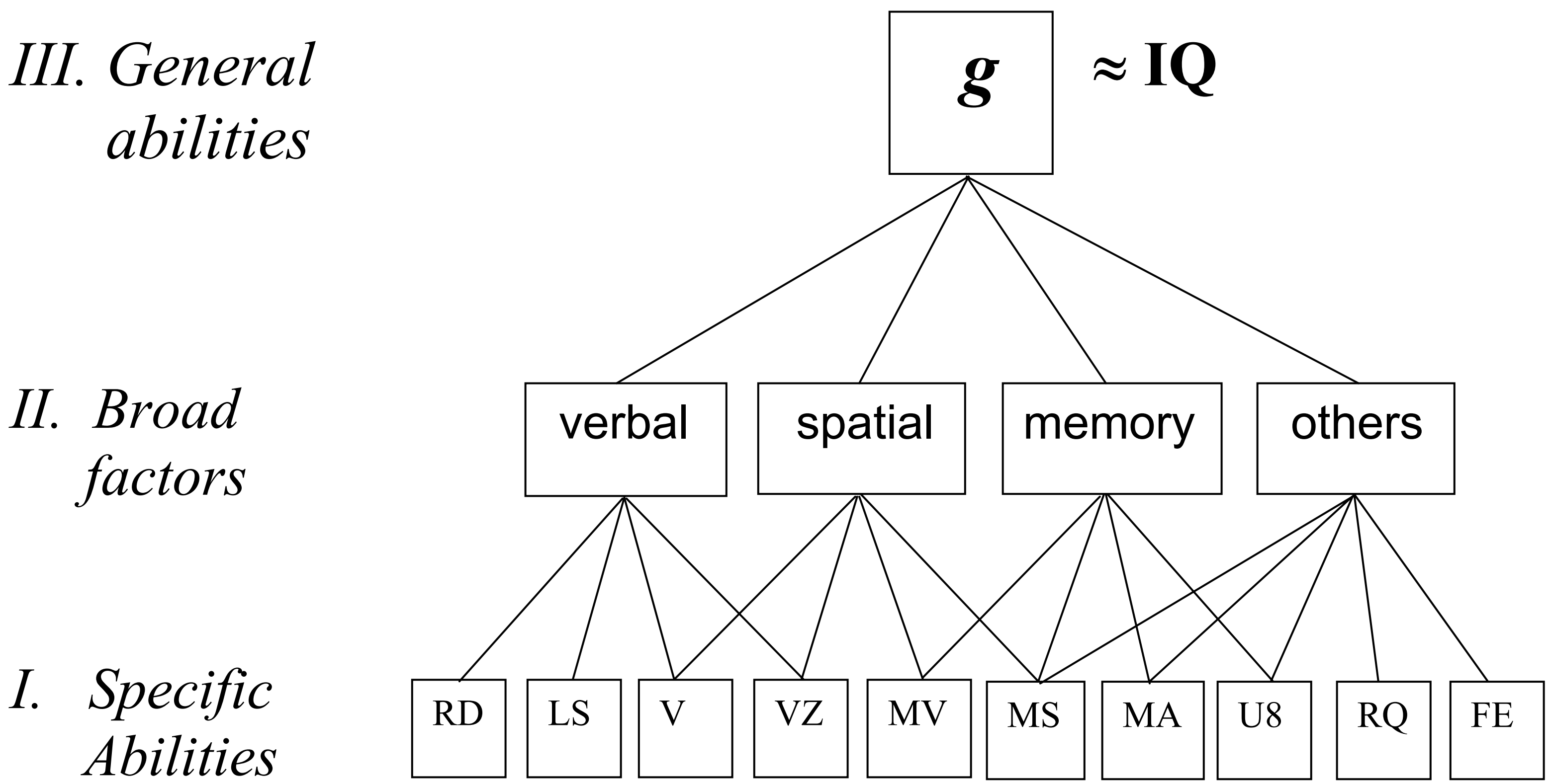
Figure Captions and Notes

Figure 1. Hierarchical model of human abilities.

This simplified rendition of the hierarchical model draws from Carroll's (1993, chap. 15) three-stratum summary of the evidence. Verbal, spatial, and memory represent three of his eight Stratum II factors, respectively, crystallized intelligence (2C), broad visual perception (2V), and general memory and learning (2Y). The Stratum I abilities sampled here are reading decoding (RD), listening ability (LS), verbal (printed) language comprehension (V), visualization (VZ), visual memory (MV), memory span (MS), associative memory (MA), maintaining and judging rhythm (U8), quantitative reasoning (RQ), and expressional fluency (FE). See Carroll (1993, p. 626) for the five other Stratum II factors in his summary model, as well as for the other Stratum I factors that are correlated with the Stratum II factors shown in Figure 1. Reprinted from Gottfredson (2003) with permission of Allyn & Bacon.

Figure 2: The Distribution of People and Life Chances Along the IQ Continuum

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. Percentiles for IQ scores were estimated by use of cumulative normal probability tables. Minority/white ratios were calculated before percentiles were rounded. Adapted from "Why *g* Matters: The Complexity of Everyday Life," by L. S. Gottfredson, 1997, *Intelligence, 24*, Figure 3, p. 117. Copyright 1997 by Elsevier Science. Reprinted with permission.

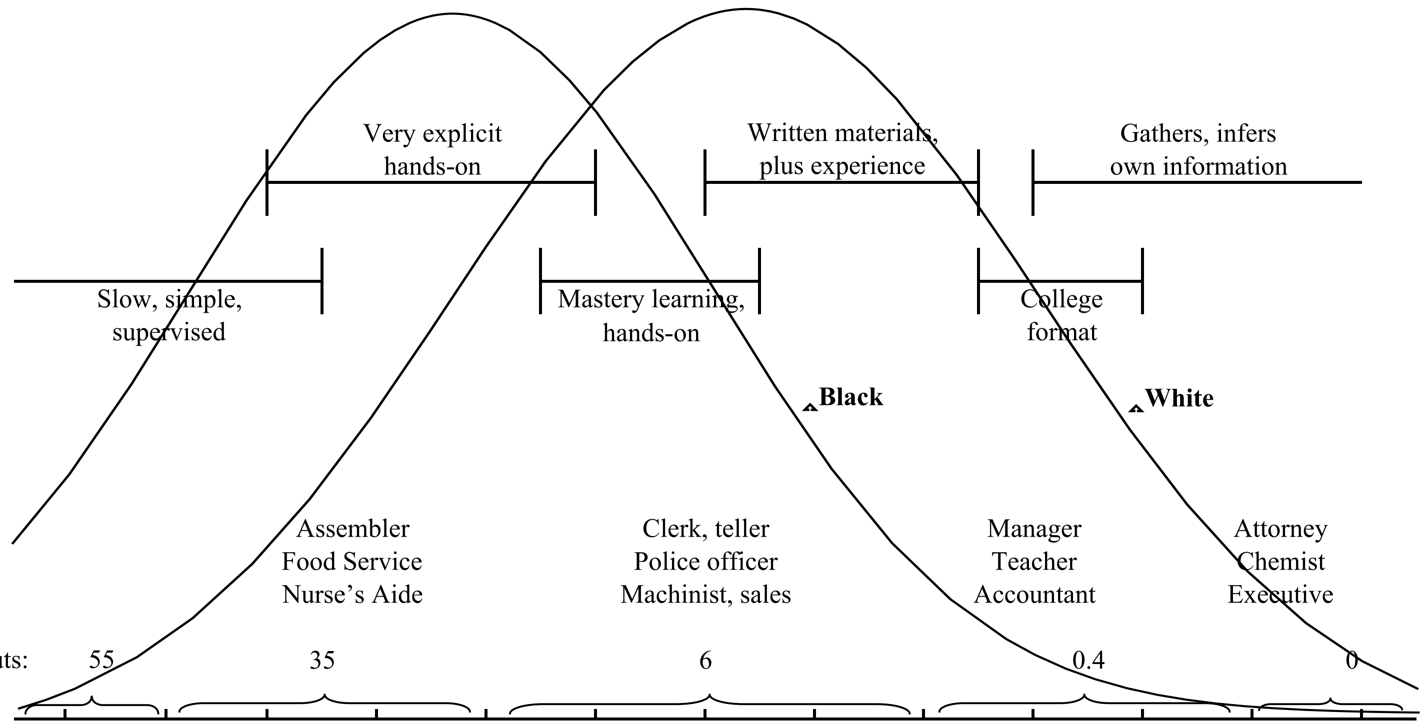

III. General abilities
g
≈ IQ
II. Broad factors
verbal
spatial
memory
others
I. Specific Abilities
RD
LS
V
VZ
MV
MS
MA
U8
RQ
FE

Life chances:	"High Risk"	"Up-Hill Battle"	"Keeping Up"	"Out Ahead"	"Yours to Lose"
% pop.:	5%	20%	50%	20%	5%

Training potential:

Career potential:

% HS dropouts: (whites)



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 %[c]													
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**

Table 1

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

Year[b]/Ages	White		Black			Hispanic			Native-American			Asian		
	Mean	SD	Mean	SD	d_{IQ}	Mean	SD	d_{IQ}	Mean	SD	ES_{IQ}	Mean	SD	d_{IQ}
380 Studies in Shuey (1966)[c]														
1922-1944														
2-6					**.60**									
School children														
Individually-admin. tests					**.93**									
Group tests: verbal					**.87**									
Group tests: non-verbal					**1.07**									
High school					**.73**									
All samples					**.87**									
1945-1966														
2-6					**1.07**									
School children														
Individually-admin. tests					**.93**									
Group tests: verbal					**1.07**									
Group tests: non-verbal					**.87**									
High school					**1.27**									
All samples					**1.07**									
Coleman et al. (1966) Equality of Educational Opportunity Report[d]														
1965 data														
1st grade IQ														
Verbal	53.2		45.4		**.78**	44.9/46.5[e]		**.83/.67**	47.8		**.54**	51.6		**.16**
Non-verbal	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**
Non-verbal	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***
Gen. 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[f]														
1972 Data														
12th grade	208	31	169	28	**1.16**									
WISC-R Standardization Sample[g]														
1974														
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**									
Perfor 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 completn	10.4	2.9	8.1	3.0	**.76**									
Picture arrangmt	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[h]														
1975														
5-11	103.2	13.4	87.8	13.1	**1.03**									
Stanford-Binet IV Standardization Sample[i]														
1986														
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**
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**
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**
DAS Standardization Sample[j]														
1986 data														
2.6-3.5	103.0	14.1	91.5	13.4	**.77**	96.0	14.4	**.47**						
3.6-5.11	103.6	13.7	86.7	13.3	**1.23**	93.6	12.0	**.67**						
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[k]														
1991														
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)														
3-4[l]	52		40		**1.20**									
5-6[m]	98.9	15.2	81.9	14.6	**1.13**									
Mothers' AFQT[m]	99.7	12.4	85.7	10.8	**.93**									
Armed Forces Aptitude Batteries														
1917-1918 (WWI) data: Recruits[n]					**1.16**									
1944-45 (WWII) data: AGCT for recruits[o]					**1.25**									
1980 data: AFQT standardization sample (NLSY)[p]														
18-23	522	86.9	401	94.5	**1.21**	429	105.8	**.93**						
WAIS-R Standardization Sample[q]														
1981														
16-19	100.8	14.1	86.9	14.5	**.93**									
20-34	101.8	15.1	87.0	11.6	**.99**									
35-54	101.4	14.8	86.6	13.2	**.99**									
55-74	101.4	14.6	87.0	13.0	**.96**									
Total	101.4	14.7	86.9	13.0	**.97**									

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

[a]Effect size is calculated here as the group mean difference (e.g., W-B) 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.

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

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

U. S. Employment Service (1970, Table 17-12). Mean=100, SD=20. "Hispanic" refers to Mexican-American.

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

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

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

[g]WISC-R=Wechsler Intelligence Scale for Children-Revised. Source: Jensen & Reynolds (1982, p. 425). For IQ scales, mean = 100 and SD = 15; for subtests, 10 and 3.

[h]Source: Jensen & 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).

[i]Source: Thorndike, Hagen, & Sattler (1986, pp. 34-36). Mean=100, SD=16.

[j]DAS=Differential Ability Scales. Source: Lynn (1996, p. 272).

[k]Source: Sattler (2001, p. 232). Mean=100, SD=15.

[l]PPVT-R=Peabody Picture Vocabulary Test-Revised. Source: Jencks & Phillips (1998, p. 2). Mean=50, SD=10; effect sizes based on differences in medians, not means.

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

[n]Source: Loehlin et al. (1975, pp. 143, 408-409). Based on a variety of tests (Army Alpha, etc.) put on one scale.

[o]Source: Jensen (1998, p. 376).

[p]AFQT=Armed Forces Qualifying Test. Source: Laurence, Eitelberg, & Waters (1982, p. 43). Mean=500, SD=100. Same sample as used in National Longitudinal Survey of Youth (NLSY; e.g., Herrnstein & Murray, 1994). "White" includes all racial-ethnic groups other than blacks and Hispanics.

[q]WAIS-R=Wechsler Adult Intelligence Scale-Revised. Source: Reynolds, Chastain, Kaufman, & McLean (1987, p. 330). Mean=100, SD=15.

[r]GATB=General Aptitude Test Battery. Source: U.S. Employment Service (1970, Table 17-12). Mean=100, SD=20; "Hispanic" refers to Mexican-American.

[s]Source of data: Wonderlic (1999, p. 34). Mean=22; SDs=8.02 for 1970, 7.70 for 1983, and 7.10 for 1992.

Table 2

Standardized IQ Gap (d_{IQ}) Between Whites and Blacks, by Socioeconomic Status (SES), on WISC-R (Ages 5-12) and WAIS-R (Ages 16-74)

SES Level	Full-Scale IQ					
	Whites		Blacks		White-Black	
	Mean	SD	Mean	SD	IQ pts.	d_{IQ}[a]
WISC-R, Ages 5-12 (98 California school districts, 1970)[b]						
Parents' occupational level						
0	85.0	16.3	83.1	13.4	1.9	**.12**
1	93.5	12.9	87.5	11.4	6.0	**.40**
2	97.9	12.4	88.5	11.7	9.4	**.63**
3	103.0	14.4	87.7	12.4	15.3	**1.02**
4	101.5	9.6	88.9	14.2	12.5	**.84**
5	103.5	12.7	90.7	12.8	12.8	**.85**
6	106.8	11.8	92.2	11.0	14.7	**.98**
7	105.1	11.4	94.3	12.5	10.7	**.71**
8	111.2	14.2	94.7	20.6	16.5	**1.10**
9	106.6	13.6	88.6	11.9	18.0	**1.20**
Total	103.2	13.8	87.8	13.1	15.4	**1.03**
WAIS-R, Ages 16-74 (1981 standardization sample)[c]						
Years of education completed						
0-8	88.6	13.7	80.2	12.0	8.4	**.56**
9-11	98.0	13.8	86.3	12.9	11.7	**.78**
12	101.1	12.2	90.7	12.1	10.4	**.69**
13-16	111.6	12.0	95.9	9.5	15.8	**1.05**
Total	101.4	14.7	86.9	13.0	14.5	**.97**

[a] d_{IQ}=standardized difference between the means, specifically, (W-B)/SD where SD=15.

[b] WISC-R=Wechsler Intelligence Scale for Children-Revised. Source: Jensen & Figueroa (1975, pp. 885, 887).

[c] WAIS-R=Wechsler Adult Intelligence Scale. Source: Reynolds, Chastain, Kaufman, & McLean (1987, pp. 330-338).

Table 3

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

<table>
<tr><th rowspan="3">Outcome That Is Being Correlated with IQ</th><th colspan="3">Outcome's correlation with:</th><th rowspan="2">(3) is corrected for unreli-ability[c]</th><th colspan="3">Minimum exp. d_{ach} if d_{IQ} equals:[c]</th></tr>
<tr><th colspan="2">IQ[a]</th><th>g[b]</th><th>1.20</th><th>.90</th><th>|.30|[d]</th></tr>
<tr><th>Range
(1)</th><th>Median
(2)</th><th>(est.)
(3)</th><th>(4)</th><th>Black
(5)</th><th>Hisp.
(6)</th><th>Asian
(7)</th></tr>
<tr><td>Composite IQ from one of 5 major IQ test batteries</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Correlated with:</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Each other's composite IQ[e]</td><td>.61-.93</td><td>.85</td><td>.92</td><td>.96</td><td>1.15</td><td>.86</td><td>-.29</td></tr>
<tr><td>Scores on own most "academic" subtests[f]</td><td>.68-.83</td><td>.76</td><td>.83</td><td>.87</td><td>1.04</td><td>.78</td><td>|.26|</td></tr>
<tr><td>Standardized academic achievement in specific subjects</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Correlated with composite IQ from 5 major IQ batteries[g]</td><td></td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Math/Arithmetic</td><td>.32-.81</td><td>.63</td><td>.67</td><td>.70</td><td>.84</td><td>.63</td><td>-.21</td></tr>
<tr><td>Reading</td><td>.30-.76</td><td>.61</td><td>.66</td><td>.69</td><td>.82</td><td>.62</td><td>.21</td></tr>
<tr><td>Language (3 R's)</td><td>.48-.68</td><td>.57</td><td>.62</td><td>.64</td><td>.77</td><td>.58</td><td>.21</td></tr>
<tr><td>Writing</td><td>.47-.68</td><td>.56</td><td>.61</td><td>.63</td><td>.76</td><td>.57</td><td>.19</td></tr>
<tr><td>Spelling</td><td>.15-.63</td><td>.42</td><td>.46</td><td>.48</td><td>.57</td><td>.43</td><td>.14</td></tr>
<tr><td>Correlated with IQs from diverse IQ tests[h]</td><td colspan="3">Typical correlations</td><td></td><td></td><td></td><td></td></tr>
<tr><td>Composite score for various kinds of achievement</td><td>.8</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Hierarchical subjects (e.g., math)</td><td>↓</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Less hierarchical (e.g., social studies)</td><td>↓</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Less academic (e.g., art, typing)</td><td>↓ (smaller)</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Grades in elementary or secondary school, job training[i]</td><td>.5-.7</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Years of education, occupational status (midlife)[j]</td><td>.5-.6</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Grades in college[i]</td><td>.4-.5</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Income (midlife)[j]</td><td>.3-.4</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
<tr><td>Law-abidingness, deportment[k]</td><td>.2</td><td></td><td></td><td></td><td></td><td></td><td></td></tr>
</table>

[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 (2001, pp. 225, 338, 379, 380, 461,and 512 for IQ tests and 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.

[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 DAS, SB-IV, WPPSI, WISC-III, and WAIS-R.

[f]Data taken from Sattler (2001, pp. 238, 342, 389, 466, and 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, and 514-515), and the achievement tests are the WIAT, WRAT-R, WRAT-III, and WJ-R. Data published mostly in the 1990's and late 1980's.

[h]Correlation for achievement composite is from Jensen (1980, pp. 323-326). Jensen (1980, pp. 316-329) discusses, but does not put numbers to, the fact that correlations with IQ are highest for the most hierarchical school subjects and lowest for less academic subjects that require special cognitive or non-cognitive abilities.

[i]Jensen (1981, pp. 30-31) for school grades; Hunter (1986) and Sticht (1975) for performance in training.

[j]Jencks et al. (1979, ch. 4); see also review by Gottfredson (2002).

[k]See Gordon (1997) on racial differences in rates of crime and delinquency.

Table 4

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

	Observed d_{ach}[a]											
	Reading				Math				Science			
Age:	9	13	17	Median[b]	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	.87	.68	.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	n.a.	n.a.	n.a.									
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						

[a]Effect sizes calculated with SDs 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 3.

Table 5

Achievement Gaps (d_{ach}) of Blacks Relative to Whites in Reading, Math, and Vocabulary, in Grades 1-12 in the Prospects and NELS National Studies, 1988-1992

Grade	Study	d_{ach}[a]		
		Reading	Math	Vocabulary
1(fall)[b]	Prospects[c]	.74	.87	.70
2	Prospects[d]	.80	.75	.77
3	Prospects[c]	.67	.67	.77
5	Prospects[d]	.72	.51	.83
7	Prospects[c]	.56	.62	.77
8	NELS[c]	.70	.78	--
9	Prospects[d]	.61	.71	.78
10	NELS[c]	.66	.77	--
12	NELS[c]	.69	.80	--
Average		.68	.72	.77

[a]Achievement gaps calculated by subtracting black mean from white mean and then dividing standard deviation of total sample at that grade level, including all races.
[b]Except for Grade 1, all tests were given in the spring.
[c]Source: Phillips, Crouse, & Ralph (1998, pp. 258-260, cross-sectional samples)
[d]Source: Phillips, Crouse, & Ralph (1998, p .270, longitudinal samples)

Table 6

Gaps in IQ (d_{IQ}), Standardized Achievement (d_{ach}), and SES (d_{SES}) for Blacks and Hispanics, Relative to Whites, in Grades 1-8 in a Large California School District, 1970

Grade	d_{IQ} [a]		d_{ach} [b]		d_{SES} [c]	
	Black	Mexican	Black	Mexican	Black	Mexican
1	1.07	.53	.25	.34	--	--
2	1.03	.70	.57	.37	--	--
3	.98	.53	.83	.68	.58	1.13
4	.95	.48	.69	.59	.38	1.18
5	1.05	.62	.75	.54	.70	1.18
6	1.23	.67	.84	.69	.47	1.36
7	1.13	.72	.71	.57	.71	1.36
8	1.18	.79	.64	.62	.77	1.34
Mean: G1-8	1.08	.63	.66	.55	--	--
G3-8	1.09	.64	.74	.62	.60	1.26

Source: Jensen (1974, p. 200).

[a] d_{IQ} is for Lorge-Thorndike non-verbal IQ and is based on white, not total, SD.

[b] d_{ach} is for Stanford Achievement Tests and is based on white, not total, SD.

[c] d_{SES} is for the Home Index of family socioeconomic level and is based on white SD.

Table 7

Estimated Mean Achievement Levels of Blacks and Whites on STEP and SCAT Tests in the ETS Growth Study When Scores Put on a Common Proficiency Scale for Grades 5-11, 1961-1967 Longitudinal Data

	Black Means (B)		White Means (W)			
Grade:	9	11	5	7	9	11
	Mean Proficiency Level					
STEP Tests[a]						
Math	257	261	248	261	272	278
Science	263	271	257	267	275	283
Social Studies	260	266	253	263	274	280
Reading	271	280	261	273	284	296
Listening	275	280	270	280	287	293
Writing	265	276	259	268	281	290
SCAT Tests[b]						
Verbal	264	272	251	265	278	284
Quantitative	277	280	258	276	292	297
			Average Achievement Gap (d_{ach}):			
8 STEP and SCAT tests listed above			.88	.87	.91	.90
8 highly specific information tests[c]			.90	.81	.88	.91

Source: Hilton, Beaton, & Bower (1971) in Humphreys (1988, pp. 238-239).
[a]STEP = Scholastic Tests of Educational Progress.
[b]SCAT = School and College Aptitude Test.
[c]Industrial arts, home arts, physics, biology, music-arts, history-literature, recreation, government.

Table 8

Mean Proficiency Levels of Whites, Blacks, and Hispanics on 1999 NAEP Tests of Reading, Mathematics, and Science at Ages 8, 13, and 17 (Cross-Age Age Proficiency Scales)[a]

Subject	Mean NAEP Score		
	9	13	17
Reading			
Whites	221	267	295
Blacks	186	238	264
Hispanics	193	244	271
Math			
Whites	239	283	315
Blacks	211	251	283
Hispanics	213	259	293
Science			
Whites	240	266	306
Blacks	199	227	254
Hispanics	206	227	276

[a]Source: National Center for Education Statistics (2000).

Table 9

Mean Proficiency Levels of Whites, Blacks, Hispanics, and Asians in NAEP Reading for Grade 12 in 1994[a;] NALS Prose and Quantitative Literacy for Ages 16 and Older in 1992[b]; and SAT Verbal+Math Composite in 1999,[c] by Socioeconomic Status Level

	1994 NAEP Reading			
Grade 12:	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 Prose Literacy				1992 NALS Quantitative Literacy			
Ages 16+:	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-yr degree	313	276	291	--	313	267	286	--
4-year degree	328	288	282	271	329	280	286	286
Grad degree	341	298	312	--	338	285	312	314

College-bound	1999 SAT (Verbal + Math)			
seniors only:	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

[a]Source: Campbell, Donahue, Reese, & Phillips (1996, p. 37).
[b]Source: Kirsch, Jungeblut, Jenkins, & 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]Source: Data provided by The College Board (9/3/99).

Table 10

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)[a]					
	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)[b]					
	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]			% College-Bound Seniors with SAT Scores >600, 1999[f]	
	Persons Aged 25-29, 2002				
	<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

[a]Source of reading scores: Donahue, Voelkl, Campbell, & Mazzeo (1999).
[b]Source of math scores: Reese, Miller, Mazzeo, & Dossey (1997).
[c]Source of science scores: O'Sullivan, Lauko, Grigg, Qian, & Zhang (2003).
[d]Data for Asians is for 1996, not 2000.
[e]Source: U.S. Bureau of the Census (2002, Table A-2).
[f]Source: College Board (1999a, 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 (1999a) publication.