Are Americans Becoming More or Less Alike?

Trends in Race, Class, and Ability Differences in Intelligence

Wendy M. Williams and Stephen J. Ceci
Cornell University

American students' test scores have been slowly but steadily declining for the past half century. Some recent explanations for this decline have focused on dysgenic trends resulting from low-IQ parents outbreeding high-IQ parents. In this article, the authors examined the evidence for dysgenic trends by considering race-, class-, and ability-related changes in intelligence test scores over time. They concluded that (a) racial differences in intelligence decreased from 1973 to 1988 and have remained fairly constant since, (b) intelligence differences between the upper and lower thirds of social class groups have been decreasing slightly since 1932, and (c) Preliminary Scholastic Assessment Test-score differences between the top and bottom quartiles have been relatively stable since 1961. Thus, the authors found no evidence supporting the dysgenic hypothesis. Rather, the combined evidence points to a growing convergence across racial, socioeconomic, and ability-related segments of American society.

It's no longer news: American students' test scores have been slowly but steadily declining to the point that the children of America's major trading partners have overtaken American children. Consider several signs of the decline. Despite some recent gains, U.S. children's achievement test scores in math and science have not kept pace with the scores of European and Asian children. Math scores for 13-year-olds fell from 4th place out of 17 nations in 1971 to 14th place by 1986; similar declines can be charted for 9- and 17-year-olds in virtually all areas except reading (Bronfenbrenner, McClelland, Wethington, Moen, & Ceci, 1996). Other indications of decline include the following: Within a 20-year period, Scholastic Assessment Test (SAT) scores slid by a full standard deviation, from 980 in 1963 to 890 in 1981 (Berliner & Biddle, 1995, Table 2.1). Even the number of students in the highest scoring group (those scoring higher than 700 on the Verbal subtest) declined, despite the greatly expanded number of students who took the test (Hayes, Wolfer, & Wolfe, 1996). It is little wonder that U.S. business leaders complain that the workforce of the future may not be able to cope with the increasingly technological and inherently complex nature of work itself (Hunt, 1995).

In view of the rather precipitous decline in cognitive test performance of American students, one can begin to understand the appeal of a major argument advanced in books such as *The Bell Curve: Intelligence and Class Structure in American Life* (Herrnstein & Murray, 1994). This argument attributes changes in test scores to the net result of a combination of beneficial and baleful forces in society. On the positive side, the forces include large increases in early childhood educational spending, adoption, and better nutrition; on the negative side, the forces include dysgenic pressures resulting from low-IQ parents outbreeding high-IQ parents. At times, the positive pressures are seen as offsetting the negative ones, but the negative pressures (dysgenic trends) are seen as dampening possible gains by forcing Americans to struggle against a "head wind": "Mounting evidence indicates that demographic trends are exerting downward pressure on cognitive ability in the U.S. and that these pressures are strong enough to have social consequences" (Herrnstein & Murray, 1994, p. 341).

In this article, we address a set of interrelated issues having to do with perceptions, both real and imagined, that the intellectual ability of various racial, cultural, and economic groups is diverging. This is not a new argument, of course. Variants of it have been around since before Terman normed the Binet test during the earlier part of this century and reported sizable ethnic differences in IQ (see Gould, 1981). Indeed, Galton (1892) made a similar claim long before the inception of modern

Editor's note. William C. Howell served as action editor for this *Current Issues* section.

Author's note. Correspondence concerning this article should be addressed to Wendy M. Williams or Stephen J. Ceci, Department of Human Development, Cornell University, MVR Hall, Ithaca, NY 14853. Electronic mail may be sent to wmw5@cornell.edu or sjc9@cornell.edu.

 0003-066X/97/$2.00
Vol. 52, No. 11, 1226–1235

methods to measure intelligence. But in its modern instantiation, the dysgenesis–divergence argument takes on the power of sophisticated statistical analyses, validation data from industry and the armed forces, and biotechnological advances in brain imaging and genetic mapping of neurologically relevant sites.

For example, Rushton and Ankney (1996) reviewed evidence for the relationship among IQ, race, and brain size, including recent magnetic resonance imaging volumetric estimations of brain size showing significant racial differences in brain volume. Others have reported that the largest racial differences are found on tests possessing the highest heritabilities (Jensen, 1985). Still others have demonstrated that IQ is correlated with central nerve conductance velocity and oscillation (e.g., Reed & Jensen, 1992, 1993). It is understandable that some commentators have explained changes in specific test scores—especially when they fall more in one group than in others—in terms of sophisticated biological data showing group differences in cranial capacity, heritability, and nerve conductance.

But have test scores been declining this century, in general? And, if they have, is there a growing gap between the scores of the top and bottom segments of American society? In this article, we consider three separate sources of data that bear on these questions.

The Genetic Meritocracy Claim

A spate of books and articles have proffered various arguments for the existence of an ever widening test-score gap between the haves and the have-nots. This gap is said to be creating pressures for social chaos and economic decline (e.g., Eysenck, 1982; Gottfredson, in press; Itzkoff, 1989; Lynn, 1991, in press; Rushton, 1995; Seligman, 1992; Waller, 1971; for rebuttals, see Ceci, 1996; Durlauf, Arrow, & Bowles, in press; Fischer, Houts, Chodrow, & Duster, 1996; Fraser, 1995; Jacoby & Glauberman, 1995). The most comprehensive of these new books is, of course, *The Bell Curve* (Herrnstein & Murray, 1994), an 840-page tome full of statistical formulas and charts that has sold more than a million copies and that has seized the media's attention.

In *The Bell Curve,* Herrnstein and Murray (1994) argued that a genetically induced bifurcation of intelligence may be taking place as a result of the tendency for the offspring of high-IQ, disproportionately White professionals to attend elite colleges and universities, where they meet and marry the offspring of other high-IQ professionals. According to Herrnstein and Murray, the offspring of these high-IQ pairings are statistically more likely to possess higher IQs than are the offspring of low-IQ pairings. This claim is consistent with the statistical evidence recently reviewed by the American Psychological Association's Task Force on Intelligence (Neisser et al., 1996).[1] This dysgenic pressure is claimed to disproportionately affect the intelligence test scores of Blacks, Latinos, and socioeconomically disadvantaged individuals:

> Blacks and Latinos are experiencing even more severe dysgenic pressures than Whites, which could lead to further divergence in future generations. . . . Putting the pieces together, something worth worrying about is happening to the cognitive capital of the country (Herrnstein & Murray, 1994, p. 341). . . . The effect is dysgenic when a low-IQ group has babies at a younger age than a high-IQ group. . . . In the United States women of lower intelligence have babies younger than women of higher intelligence (p. 351). . . . The higher fertility rates of women with low IQs have a larger impact on the Black population than on the White. The discrepancies are so dramatically large that the probability of further divergence seems substantial. (pp. 353–354)

Importantly, Herrnstein and Murray (1994) wrote about dysgenic *pressures* rather than dysgenic *effects.* They did so because compensating and countervailing pressures in one direction may be canceled by pressures in the opposite direction. For example, Herrnstein and Murray argued that although the tendency for lower IQ persons to have more offspring than higher IQ persons (particularly during times of economic scarcity) may have potentially deleterious consequences for the cognitive capital of America, positive pressures in the environment, such as early childhood and educational interventions, could offset this dysgenic effect and could even produce a net increase in test scores. However, Herrnstein and Murray went on to assert that "whatever good things we can accomplish with changes in the environment would be that much more effective if they did not have to fight a demographic headwind" (p. 342).

In this article, we asked whether an intellectual dysgenesis has been taking place and, if it has, whether racial, socioeconomic, and ability-related gaps in intelligence are widening. In the first part of this article, we reviewed the descriptive demographic data. Next, we examined three forms of possible divergence. In the final part of this article, we attempted to determine if the alleged widening of the cognitive gap will continue into the next century, and, if so, we considered the policy implications of such a trend. Throughout this discussion, we focused on test-score declines as the most visible (and testable) form of dysgenesis. However, we acknowledge the possibility that downward genetic pressures could be obscured or offset by positive societal interventions, with the result being that test scores will not decline even while the gene pool becomes poorer. But it is the observable decline of American students' test scores that has animated recent alarm about dysgenesis, and it is at this level of discourse that this article is focused.

Descriptive Data

Throughout this century, Whites have outscored Blacks and Hispanics on IQ tests as well as standardized achieve-

[1] And because low-IQ parents tend to start childbearing at an earlier age than do their high-IQ counterparts, they end up having greater numbers of children across generations (this will be true even if the actual number of offspring remains the same in both groups within a given generation).

ment tests. The gap most commonly reported is approximately one standard deviation. On the most widely used individual IQ test, the Wechsler series, one standard deviation translates into a 15-point gap between Blacks and Whites, with Hispanics falling midway between these groups.[2] Providing evidence for this claim, Lynn (1996) reported that a representative sample of 2,260 children between the ages of 6 and 17 years revealed that the average IQs of Whites and Blacks were 103 and 89, respectively; Peoples, Fagan, and Drotar (1995) reported a similar gap of one standard deviation between 3-year-old White children's IQs (M = 100) and Black children's IQs (M = 85) on the current edition of the Stanford–Binet Intelligence Scale.

Racial and ethnic gaps on IQ and achievement tests have existed throughout this century. For example, IQ differences between Blacks and Whites were evident on the first Stanford–Binet IQ test normed in 1932, and earlier signs of a racial gap of approximately one standard deviation were apparent on the tests administered to recruits during World War I (Loehlin, Lindsay, & Spuhler, 1975). Differences between the scores of rich and poor samples also have been observed since the first tests were administered. Although none of these facts are in dispute among researchers, their meaning and putative causes are.

Below, we address the following four questions: (a) What are the achievement test score trends for various racial, socioeconomic, and ability-based groups in America? (b) What do these trends reveal about IQ changes among members of these groups? (c) How can these changes be explained? and (d) What do these data portend for public policy in America's future? Because of the close interrelationship between the first two questions, we considered them in tandem.

Achievement Versus IQ

Readers who are unfamiliar with the psychometric tradition may wonder why we focused on achievement test scores instead of intelligence test scores. After all, the dysgenesis argument is about changes in scores on tests of intellectual aptitude rather than on tests of school achievement. However, putting aside one's theoretical orientation (i.e., whether achievement and aptitude are conceptually different, albeit causally related, or whether they are held to be one and the same entity), the empirical reality is that trends in achievement test scores closely mimic IQ trends (see review in Neisser et al., 1996). Notwithstanding any theoretical distinction one may wish to draw between intelligence and achievement, the empirical reality is that a reliable measure of one is highly correlated with a reliable measure of the other.

For example, in the recent American Psychological Association Task Force on Intelligence, Neisser et al. (1996) noted that a wide range of "content-oriented achievement tests" correlate highly with IQ as well as with all widely used aptitude tests (e.g., the American College Test [ACT], the SAT, the Graduate Record Examination, and the Medical College Admission Test).[3] This explains why intelligence researchers frequently use measures derived from batteries of achievement test scores as a proxy for IQ scores even though at times the content may appear to be quite dissimilar (e.g., some types of IQ tests contain visual matrices, mazes, and puzzles, whereas some achievement test batteries contain only verbal content, such as political editorials or arithmetical word problems). Ceci, Rosenblum, and Kumpf (in press) described this situation as follows:

> If you doubt the above assertion [that trends in IQ test scores mimic achievement test trends], you can do the following experiment: Select a random stratified sample of children and administer to them two batteries of tests, an achievement battery (e.g., mathematics, reading comprehension, scientific reasoning), and the most widely used IQ battery (e.g., the ten subtests from the Wechsler Adult Intelligence Scale—III [WAIS–III]). Next, distill from the achievement-test-battery intercorrelations a single summative score that captures the covariance among the math, reading, and scientific reasoning scores. (This is traditionally accomplished by taking the first principal component from it.[4]) This summary score of the sample's achievement test scores will be very closely related to the IQ scores for the same sample. Often, the correlations with IQ approach the internal reliability of the IQ itself, i.e., the achievement test summative score correlates with IQ about as highly as IQ correlates with itself (Ceci, 1996). It is rare to find a measure of achievement that does not correlate highly with an IQ score as noted above.

[2] Although the Black–White gap in IQ scores is agreed to be on the order of one standard deviation, there is less consensus on the Asian–White gap in IQ scores, which, summing all studies, is on the order of Asian Americans scoring approximately 3 points higher than Whites. However, there is some debate among researchers on the magnitude of this gap. For example, Vernon (1982) reported a mean IQ of 106 among Asian Americans, and Lynn (1996) reported a mean IQ of 107 among Asian Americans; conversely, Neisser et al. (1996) reported a mean IQ of only 98 for Asian Americans, on the basis of Flynn's 1991 estimate. In addition, Herrnstein and Murray's (1994) own analyses of the National Longitudinal Assessment of Youth data revealed a mean IQ of 106 for Asian Americans. Taken together, a prudent position seems to be that IQ scores of Asian Americans are slightly higher than those of Whites.

[3] Note that the claim here is not that tests of intelligence and achievement do not differ in principle but merely that it is empirically difficult to disentangle them. Virtually any battery of achievement tests given to a group of individuals (e.g., math, science, verbal reasoning), when factor analyzed, will yield a sizable general factor that correlates highly with those same individuals' IQ scores (also see Cooley & Lohnes's 1976 demonstration that commingled items from a popular mental ability test were difficult to tell apart from items taken from a widely used achievement test). It is often the case that specific achievement test batteries and IQ tests yield factor structures that differ in the magnitude of the first or general factor or in the nature and size of lower order factors (e.g., speed, various so-called crystallized factors, memory). But at the apex of the factor structure in all diverse test batteries is a general factor that permeates most or all subtests that compose the battery, regardless of whether it happens to be an achievement test battery or an IQ test. This general factor is variously termed *fluid intelligence, general intelligence, g, abstract reasoning,* or the *first principal component.* Because it appears to be involved in successful performance on all types of achievement tests as well as IQ tests, it is held responsible for the empirical difficulty of disentangling them.

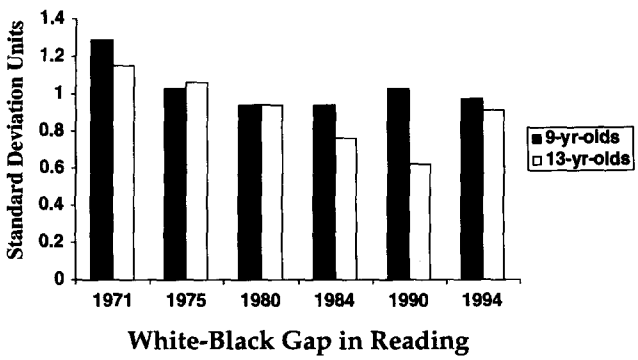
[4] This is operationalized as the maximum (linear) variance that can be accounted for, independent of any type of factor rotation, in the matrix of correlations among the various test scores.

To some, the close relationship between scores on tests that are avowedly designed to measure school achievement (e.g., math, reading, and scientific reasoning) and scores on tests designed to measure so-called intelligence is attributed to the role that the latter is assumed to play in the former. To others, however, the high correlation between achievement and intelligence confirms the folly of creating hard distinctions where none may exist (Cooley & Lohnes, 1976). Whatever one's view on this topic is (we are agnostic), the high degree of correlation between IQ and achievement scores permits the use of one as a proxy for the other. Hence, in this article, we, at times, relied on trends in achievement test scores for hints about correlated trends in IQ scores. At other times, however, we drew directly on trends in IQ scores.

Trends in Racial Differences on Achievement Test Scores

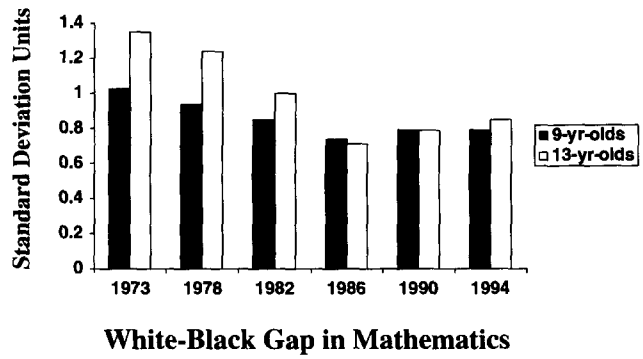
As we already mentioned, one standard deviation (15–16 IQ points) separates the IQ scores of American Blacks and Whites (e.g., Lynn, 1996; Peoples et al., 1995), and one fifth of a standard deviation (3–4 IQ points) separates the IQ scores of East Asians and Whites, the former scoring higher than the latter (Herrnstein & Murray, 1994). On batteries of achievement tests, a similar racial–ethnic gap has been shown to exist. However, recent analyses have shown that the size of this gap narrowed significantly by the late 1980s (Bronfenbrenner et al., 1996; see also Grissmer, Williamson, Kirby, & Berends, 1994; Hauser & Huang, in press). Although a racial gap of one standard deviation has persistently held for as long as records have been kept, in the early 1970s there began to appear signs that a racial convergence was taking place. Over a period of approximately 15 years, the achievement test scores of Black students narrowed the gap with those of White students by between one third and one half, so that by 1986, the racial gap on the National Assessment of Educational Progress (NAEP) had closed by approximately one half (see Figures 1 and 2).

Figure 1
Narrowing of the White–Black Gap in Reading Between 1971 and 1994



Note. yr = year.

Figure 2
Narrowing of the White–Black Gap in Mathematics Between 1973 and 1994



Note. yr = year.

There are multiple achievement tests on the NAEP (math, science, reading, and writing), as well as multiple age groups taking these tests (9-, 13,- and 17-year-olds) and multiple grades that crosscut each age group (e.g., on some trend data, 8th graders include not only 13-year-olds but also somewhat older and younger students). Thus, the magnitude of racial convergence varies somewhat by the specific achievement test, grade, and age group being discussed. Between the early 1970s and the late 1980s (National Center for Education Statistics, 1991), substantial gap-closing was evident for nearly all tests, grades, and age groups (for a detailed breakdown, see Hauser & Huang, in press, Tables 1 and 2).

Unfortunately, however, it appears that the racial convergence in test scores ceased by the late 1980s. The trend to converge did not appear to have continued in the 1994 NAEP trend data that came out in November 1996 (National Center for Education Statistics, 1996). These latest NAEP trend data showed no further systematic changes in racial means, if the discussion is restricted to statistically reliable differences. (After 1988, however, there was a hint of a potential divergence in the scores of Blacks and Whites, although so far this suggestion of divergence has not reached statistically significant levels.)

A more detailed analysis revealed the following: In the latest NAEP trend data (National Center for Education Statistics, 1996), there were 12 categories: 3 age groups (9-, 13-, and 17-year-olds) and 4 subject areas (science, math, reading, and writing). Comparing the gap between Black and White students' scores in these most recent NAEP data with those in the 1990 report revealed that 8 of the 12 trends showed a slight divergence between White and Black students' test scores but that these differences were not statistically reliable. (In addition, the

remaining 4 of the 12 contrasts showed signs of further convergence, but these also were not significant.) We plotted these data in Figures 1 and 2. Clearly, the next release of NAEP data in 1998 will be extremely interesting: Will the hint of a potential reversal (not statistically significant as of yet) apparent in the 1994 data continue and reach conventional levels of statistical significance? Will there be a reversal of the gains made during the 1970s and the 1980s? These questions are ones to which educators and policymakers should be attuned.

Recently, several independent teams of researchers have reported similar findings through the 1990 NAEP data. Grissmer, Williamson, Kirby, and Berends (in press), using a slightly different statistical model than the one we used, reported highly similar findings in which the 0.9–1.2-standard deviation gap that existed on a broad array of achievement test scores in the early 1970s had been narrowed anywhere from 25% to 50% by 1986. Along these same lines, Hauser and Huang (in press) reported that the one-standard deviation gap between IQ scores for Black and White students that existed in 1973 had narrowed by 1988, on average, by 40%. (Hauser and Huang's finding reflects the inclusion of science tests that showed very little change over this period except among 9-year-olds; when only reading and math changes were included, to make their data more comparable with our own as well as with Grissmer et al.'s [in press], the average magnitude of gap-closing among 13- and 17-year-olds was 55%.)

In fairness to Herrnstein and Murray (1994), two points should be mentioned: First, growing gaps in the scores of Blacks and Whites are not necessarily inconsistent with a dysgenesis hypothesis, because countervailing beneficial pressures could be offsetting any negative pressures. However, there is no scientific means of testing the extent to which such claims may be true because they are empirically unfalsifiable. Second, Herrnstein and Murray themselves mentioned three studies showing that the IQ gap in cognitive test performance between Blacks and Whites seems to be converging. Having made this acknowledgment, however, they went on to raise statistical concerns about one of the studies, and they further noted that in their own analysis of the NAEP data, they found a convergence of math and verbal fluency measures of Black and White 17-year-olds but far less than what we, Grissmer et al. (in press), and Hauser and Huang (in press) have reported.[5]

As the table indicates, Black progress in narrowing the test score discrepancy with Whites has been substantial on all three tests and across all of the age groups. The overall average gap of .92 standard deviation in the 1969–1973 tests had shrunk to .64 standard deviation by 1990. The gap narrowed because Black scores rose, not because White scores fell. Altogether, the NAEP provides an encouraging picture (Herrnstein & Murray, 1994, p. 291). . . . The question that remains is whether Black and White test scores will continue to converge. If all that separates Blacks from Whites are environmental differences and if fertility patterns for different socioeconomic groups are comparable, there is no reason why they shouldn't. The process would be very slow, however, . . . reaching equality sometime in the middle of the twenty-first century. . . . If Black fertility is loaded more heavily than White fertility toward low-IQ segments of the population, then at some point convergence may be expected to stop, and the gap could begin to widen again. (p. 293)

Thus, for all nine tests (science, math, and verbal fluency, for each of three age groups), Herrnstein and Murray (1994) agreed that there was racial convergence. However, troubled by the earlier onset of childbearing by Black teenagers, these authors cling to a dysgenesis hypothesis. On the basis of the 1990 NAEP analyses reported above, however (see also Grissmer et al., in press; Hauser & Huang, in press), if Blacks' fertility is loaded downward, it is difficult to explain why all evidence points in the opposite direction from that resulting from a dysgenic trend.

If the trends in IQ test scores were to mimic the achievement test score trends just described, one would expect a similar reduction in the historically stubborn racial intelligence-score gap, indexed by the two most widely used IQ tests (i.e., the Stanford–Binet and the Wechsler series). Specifically, Black students (who showed the greatest gains on the NAEP achievement test scores in the 1970s and the 1980s) might also exhibit a comparable gain in IQ, thus closing the one-standard deviation gap by approximately one half. It is important to note that closing the racial gap in achievement test scores has resulted from gains in test scores by Black students rather than from reductions in White students' scores.

In contrast, because the gains made by Blacks ended by 1988, and actually have shown signs of reversing direction since then, it might be expected that the later cohort's IQ scores will show a commensurate gap-widening. Again, this assumes that trends in achievement test scores mimic trends in IQ scores, regardless of the conceptual distinctions one wishes to draw between achievement and intelligence.

As pointed out by demographers and social scientists (e.g., Grissmer et al., 1994; Hauser & Huang, in press), the NAEP data sets have three unique advantages for those interested in testing the divergence–convergence hypothesis. First, in contrast to the SAT and IQ tests, the NAEP has not changed its item content since its inception in 1969; that is, the very same items are used today that were administered to earlier cohorts, thus enabling a direct comparison of the number answered correctly. Second, the NAEP is administered to a large, representative, national sample of children at each age.

[5] Herrnstein and Murray (1994) arrived at a smaller estimate of the size of the convergence in the NAEP data, arguing that approximately 33% of the racial gap had been closed by the late 1980s. As Hauser and Huang (in press) correctly noted, however, Herrnstein and Murray arrived at their lower estimate of convergence by using the wrong measure of variance in making their calculations. The appropriate age-corrected (hence smaller within-group) standard deviations led to the higher rates of convergence reported by us, Grissmer et al. (in press), and Hauser and Huang.

Third, the NAEP sampling procedure does not exclude certain groups (e.g., those not bound for college), as do tests like the SAT. Hence, having shown that the best available data suggest that racial divergence in intelligence is not increasing but actually decreasing, we turn next to an examination of a different form of divergence, namely, that between the test scores of the offspring of rich and poor parents.

Trends in the Intelligence of Different Socioeconomic Status Groups

Recent analyses by the political scientist James Flynn (in press) of New Zealand have called into question the claim of a widening gap in the IQ scores of rich and poor people. The claim that there is a genetically driven socioeconomic status (SES) divergence leads to the expectation of a growing tendency for good genes for IQ to rise to the top of the occupational scale and for bad genes to fall to the bottom. Specifically, this claim leads to the expectation that the IQ gap between the children of the upper and lower income groups has been diverging over time. Flynn was able to provide a strong test of this claim in his most recent analyses.

Flynn's (in press) data came from the Stanford–Binet normative sample tested in 1932 and the Wechsler Intelligence Scale for Children (WISC) samples tested in 1947–1948, 1972, 1985, and 1989. The relevant comparisons involved the mean IQs of children in American families representing a hierarchy of occupations, ranked from professions at the top to unskilled workers at the bottom. Flynn provided details on the specific U.S. census occupational codes used for each IQ test standardization sample as well as the manipulations required to satisfy various measurement issues.[6]

Flynn (in press) measured the difference between the mean IQs of children whose parents were in the top and bottom thirds of occupational status in each standardization period to see whether these groups diverged in IQ over time. As can be seen in Figure 3, for Whites, the IQ gap between the top and bottom SES groups was higher than 12 points in 1932, falling to 10 points in 1948. The size of this gap has not changed through the 1989 WISC–III sample. (If the recent Stanford–Binet and WISC–III samples are pooled, the gap stands at little more than 9 points.[7])

If occupational status is partially determined by genes for intelligence, it would be expected that over time the correlation between occupational status and intelligence would increase as a result of selective mating. As seen in Figure 3, however, the correlation between IQ and occupational status, if anything, actually decreased over time. In summarizing these analyses, Flynn (in press) concluded, "Concerning the effect of social mobility on stratifying genes for intelligence within White America, the most parsimonious conclusion is this: nothing, nothing, nothing, absolutely nothing has happened."

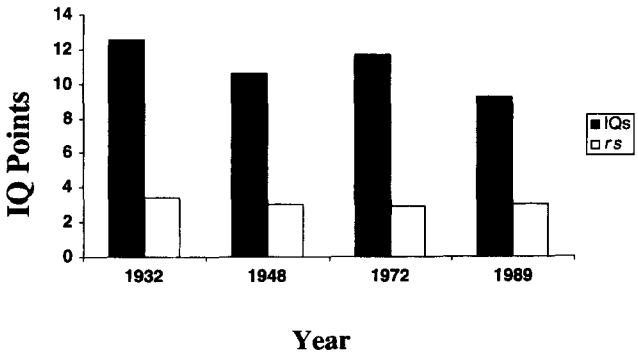
So, on the basis of Flynn's (in press) findings, the claims of social immobility and further widening of the SES gap in intelligence are not supported. Coupled with the racial data reported above, there is reason to doubt the claim of a steady downward pressure on "genes for intelligence" (e.g., Lynn, in press). Although the magnitude of the narrowing of high- and low-SES groups' IQs is not large, there is absolutely no evidence of an opposite trend or divergence.

To these two failures to find evidence in support of claims of racial and SES divergences, we add an examination of one final claim of divergence—specifically, the claim that the gap between the scores of the brightest and dullest students is widening. To adumbrate our findings, we found no compelling support for this claim either.

Trends in Intelligence of Ability-Related Groups

Examining changes in average scores between bright and dull students is yet another way of evaluating the dysgen-

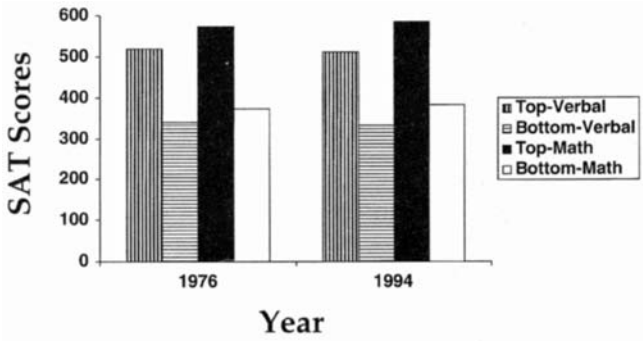
Figure 3
IQ Gap Between Upper and Lower Thirds of Social Class



[6] Rendering occupational categories comparable across time poses several thorny measurement problems, which Flynn (in press) discussed at length, because the relative ranking of prestige for occupations has changed somewhat over this century. For example, managers were ranked in the second from the top category of occupations in the 1932 Stanford–Binet standardization, but by the 1985 Stanford–Binet and 1989 WISC–III standardizations, managers had been merged with the professions in the top category. Flynn also made adjustments for changes in the way IQ test manufacturers classified the occupational status of American homes; earlier standardization samples were based on the status of the occupation of the male head of household, whereas in 1989, this was based on whichever adult had the highest status occupation. Flynn constructed mean IQs for children whose parents came from each of the five occupational categories to assess whether changes had occurred.

[7] Flynn (in press) showed that the gap for all races combined had increased between 1972 and 1989 from 11.64 to 12.85 points. However, this increase was due solely to increased immigration, plus a natural increase in the number of minorities, which together doubled the number of non-Whites in the recent samples. Because non-Whites have a lower mean IQ than Whites and are concentrated in low-status occupations, such a demographic trend automatically increases the all-races SES–IQ gap. This has nothing to do with the divergence thesis of *The Bell Curve* (Herrnstein & Murray, 1994), which predicts a widening class–

Figure 4
Top and Bottom Scoring Scholastic Assessment Test (SAT) Groups for 1976 and 1994



esis–divergence hypothesis. Specifically, if genes for intelligence are leading to increasingly disparate test scores, then it ought to be possible to test this claim by showing that the gap between students at the top and the bottom of the distribution is getting larger. Testing this expectation is problematic because the content of IQ tests (and surrogates like the SAT) changes over time. Even more problematic is the fact that the demographic makeup of the samples taking these tests shifts dramatically over time. One cannot be sure that a widening of the gap between bright and dull test takers over time is due to actual aptitude-related changes in the gene pool over time, as opposed to changes in the proportions of economically disadvantaged students taking these tests (see Berliner & Biddle, 1995; Hayes et al., 1996).

Yet, despite the much larger size of the pool of applicants taking the SAT today as compared with 20 years ago—meaning that less selective, more economically diverse students take the SAT today (Berliner & Biddle, 1995; cf. Hayes et al., 1996)—there is no real evidence of a divergence between the top 10% and the bottom 20% of scorers on the Verbal and Mathematical sections (see Figure 4). Specifically, the gap between the scores of the highest and lowest groups of students has not widened: In 1976, the mean SAT Verbal score for the top 10% of high school seniors was 518, whereas the mean SAT Verbal score for the top 10% in 1994 was 512, a decline of 6 points. For the SAT Mathematical section, the comparable scores for the top 10% in 1976 and 1994 were 574 and 586, respectively, an increase of 12 points. So, no clear downward trend is evident. But for the purposes of testing the dysgenesis claim, it is necessary to go beyond these data and to ask about the relative gap between the top and bottom groups of students. Here are the relevant means: For the bottom 20% of SAT Verbal scores in 1976, the mean score was 339; for the bottom group in 1994, the mean score was 332—a decline of 7 points. In contrast, the means for the SAT Mathematical scores over this same time period were 364 and 363, respectively. Thus, the gap between the SAT scores had not widened significantly over this 18-year period.

Preliminary Scholastic Assessment Test (PSAT) scores have an advantage over SAT scores because the PSAT has been administered to a nationally representative sample of high school juniors since 1961 (Solomon, 1983), thus avoiding the problems of self-selection that plague interpretations based on the SAT.[8] Berliner and Biddle (1995) showed that there has been no decline in the aggregate trend of PSAT scores, but the question that we ask here is whether the PSAT scores of the top and bottom groups of students have diverged over time. To examine this question, we plotted the gap between the highest and lowest quartiles of juniors from 1961 through the most recent year (1995) for which data are available from the Educational Testing Service.

As can be seen in Figure 5, our analysis suggests that the gap between the top and bottom students has not been widening over the past 35 years. In 1961, the gap in Verbal scores between the top 25% and the bottom 25% of students was 15.2; this same gap in 1995 was 14.0. Similarly, for the gap in Mathematical scores, the trend was slightly downward, going from 15.3 to 15.0 over the past 35 years.[9] Both correlations between the size of the Verbal and Mathematical gaps over time were negative and significant ($rs = -.46$ and $-.43$, respectively). This finding confirms what is visibly apparent in Figure 5; namely, the best and worst students have converged on the PSAT over time.

A more direct way to determine whether Americans are diverging is to correlate changes in the size of standard deviations over time, assuming that greater divergence would be reflected in larger standard deviations. However, in the PSAT data, there is evidence that American high school juniors have been converging since 1966 ($r = -.55$ between year of test administration and standard deviation for that year, $p = .002$). (There was no significant correlation between means and standard deviations over this same period [$r = .23$].)

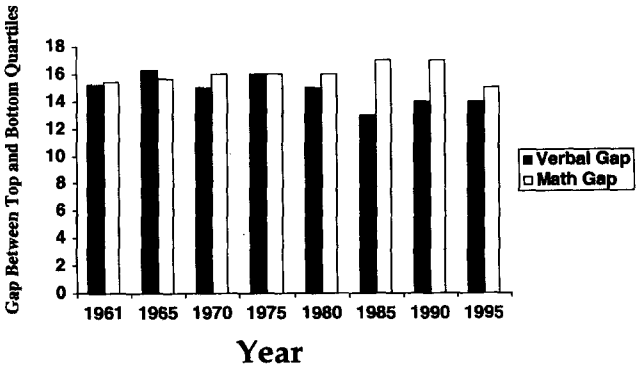
Taken together, these three analyses of alleged divergences (racial, SES, and ability-related) converge on the view that no sizable cognitive dysgenesis seems to have been occurring in America. If there is downward pressure

IQ gap on the basis of dynamics within groups that have been in America throughout this century but that differ in genetic talent.

[8] Another interpretative snarl that confronts anyone attempting to make sense of changes in SAT scores is that there have been dramatic shifts in the proportion of colleges that require it; this has resulted in some of the most select students taking the SAT because they wish to attend elite out-of-state colleges because their in-state colleges require the ACT. To whatever degree the distribution of colleges requiring the SAT and the ACT has changed over time, this will confound the interpretation of such results.

[9] We interpolated the 1961–1965 scores, which resulted in some imprecision. Moreover, the 1961 data were based on only 508,000 high school juniors who took the PSAT, whereas by the mid-1960s, the number had jumped to one million. So, some caution is needed in making comparisons of this sort, even though we suspect that any

Figure 5
Gap Between the Top and Bottom Quartiles on the Preliminary Scholastic Assessment Test



Note. The data for 1961 and 1965 were interpolated.

on the gene pool for intelligence, there is no apparent manifestation that we have been able to detect in these three nationally representative data sets (NAEP, WAIS–III IQ norms, and PSAT).

An even more basic problem with the dysgenesis–divergence hypothesis can be raised: If the presumed downward pressure on the gene pool for intelligence is tied to a growing economic stratification, as Herrnstein and Murray (1994) claimed, how can this claim be reconciled with the empirical reality that society's economic resources are not parceled out in accordance with differences in intelligence? Put bluntly, if income were distributed according to differences in IQ, one would expect a far less asymmetric distribution of income than there is now. "Many more people would earn close to the national mean, and far fewer would earn at either of the extremes" (Ceci & Williams, 1997, p. 1057).

> In a recent econometric analysis, Dickens, Kane, and Schultze (1995) showed that if IQ were equated among all people and only nonintellective variables were allowed to vary (e.g., parental SES and motivation), then the resultant income distribution would resemble the one we now have. Conversely, if all nonintellective differences were equated and income was distributed solely in accordance with differences in IQ scores, then a far more egalitarian income distribution would be observed than the one we now have. (Ceci & Williams, 1997, p. 1057)

Granted, no serious scholar would argue that income is determined solely by IQ; however, Dickens et al.'s analysis demonstrates the hurdles faced by those who maintain that economic variation is due largely to differences in genes for intelligence.

> Another way to think about this is to compare the incomes of those who possess the top 10% of IQs with the incomes of those who possess the top 10% of wages. The incomes of those with the top 10% of IQs in Herrnstein and Murray's (1994) National Longitudinal Survey of Youth sample earned 55% more than average-IQ persons earned. In contrast, the top 10% of wage earners in this same sample earned 200% more than the average person earned! Hence, the proportion of the variation in income that can be explained on the basis of variation in IQ is actually rather small. (Ceci & Williams, 1997, p. 1057)

So, the claim by meritocracy advocates that society's resources are being bifurcated as a result of a widening gene pool for intelligence does not mesh with the empirical reality: Income varies much more because of non-IQ differences than because of IQ differences, leading one team of economists to remark, "If all that mattered was [IQ] scores, U.S. society would clearly be very egalitarian. Eliminating differences due to IQ would have little effect on the overall level of inequality" (Dickens et al., 1995, p. 20).

Conclusion

In this article, we have shown that the claims of divergence in Americans' intelligence are not supported by analyses of national data sets of cognitive scores. Although SAT scores did indeed spiral downward in the 1960s and the 1970s, there are good reasons for this decline that appear to have nothing to do with dysgenic trends. For example, the pool of high school seniors taking the SAT became less selective during this period, and the number of universities requiring it increased. Had the SAT been taken by all high school seniors rather than by a self-selected sample in the early days of its administration, we suspect that the decline would have been greatly diminished in magnitude (Berliner & Biddle, 1995).

Support for this assertion was provided by the finding that during the same period for which SAT scores were declining, PSAT scores were increasing. As we have pointed out, PSAT scores are a population-based measure not susceptible to sampling fluctuations due to self-selection. If we contrast the means for a 20-year period during which the size of the pool of juniors taking the test was at least one million, the mean PSAT Verbal score actually rose from 42.7 in 1966 to 48.7 in 1995, and the mean PSAT Mathematical score rose from 45.0 to 48.9 over this same period. This is the best evidence that the often-reported downward trend in SAT scores probably does not reflect any systemic factors.

In addition to the stasis in PSAT scores, there appears to be no divergence in intelligence test scores between upper and lower SES groupings. If anything, there has been approximately a 25% convergence in the IQ scores of the upper and lower one thirds of SES groups.

The area of most dramatic convergence (as opposed to divergence) was in racial differences in intelligence during the 1970s and the 1980s. What might account for the finding that Black students during this period closed approximately half the gap that had separated them from White students? To answer this question, we begin by

sampling error is in the direction of a less elite sample taking the PSAT over time, thus rendering our argument even stronger. As support for this position, even if we were to choose later data that are not interpolated but based on large samples that are nationally representative, we would arrive at the same conclusion.

considering some nongenetic factors that could influence test scores.

As Ceci et al. (in press) and Williams (in press) have pointed out, potential factors responsible for the increase in Blacks' test scores include substantial increases in educational spending throughout this entire century (see Hanushek & Rivkin, 1997), especially for programs targeted at minorities (e.g., Head Start; Title I; and desegregation, busing, and school lunch programs). In addition, there has been an enormous increase in parental educational attainment by Black and Hispanic parents during the same period of rapid score gains by Black youngsters (Grissmer et al., 1994, in press).

It is well-known that parental educational level is tied to children's educational attainment (Bronfenbrenner et al., 1996). So, if parents are becoming better educated and if Black parents are making disproportionately greater gains in becoming better educated than are White parents, this difference should elevate Black students' scores relative to those of other students. Cook and Evans (1996) and others recently have estimated that approximately a quarter of the racial gap was closed as a result of the disproportionately large gains made by Black parents in their educational levels during the period in question. Other factors contributing to Black students' test gains include a disproportionate reduction in family size for Black families over the same time period and the associated increase in financial resources per child that accompanies a reduction in family size.

One could always argue that dysgenic trends are operating but that they are too recent to be detected in the analyses reported in this article. To whatever extent this is the case, there is no way of knowing. The 1996 NAEP data are the most recent that are available, and they also represent the best data source, given their constancy over the past 25 years. In the most recent NAEP trend data, which were released in the fall of 1996, the racial gap appears to have held firm. In general, the gap does not seem to have increased or decreased reliably since the late 1980s. Thus, one could be heartened that the racial gap is not increasing, or one could be saddened that the closing of the gap has not continued. But regardless of one's reaction to these most recent data, they do not support the suspicion that dysgenesis is taking place but rather that it is too recent to be detected.

Along similar lines, one could argue that dysgenesis is occurring but that it cannot be detected by examining changes in mean scores, because differences between top and bottom groups are meaningful only if the standard deviations are the same. Although this is a problem with some data sources, such as IQ, for which test makers force scores into a distribution with a fixed standard deviation, there is evidence in the data sources used in this article that the standard deviations have remained constant over the same period during which means have converged for social classes, ability groups, and races. Because means and standard deviations are independent, this convergence in means constitutes further evidence of a lack of divergence. For example, depending on the age and the test under consideration, Blacks' gains on the NAEP have taken place in the face of relatively stable standard deviations (roughly 34 points, collapsing across the three ages and the three content tests). The same is true of the PSAT data: The standard deviations have remained almost identical over the period of rising mean scores.

Finally, how can one reconcile the argument and the analyses reported here with the frequently reported dysgenic effects in the scientific literature (e.g., Vining's 1982 analyses that formed the basis of much of Herrnstein and Murray's 1994 argument)? According to Vining and others (see Lynn, in press), lower-class and lower educated persons have tended to have larger families than have higher educated persons, and this trend has been evident for nearly 200 years. This is a complex issue, one that goes beyond the constraints of this article. Interested readers can consult the debate between Preston (in press) and Lynn (in press) for their respective views as well as statistical evidence for and against the argument that low-IQ families have been outbreeding high-IQ ones. There are other relevant views as well (Loehlin, in press; Waldman, in press). Until there is a consensus within the scientific community on this point, there is no alternative to the test-score evidence that has been presented in this article.[10]

In sum, we found no compelling evidence supporting the hypothesis that a dysgenic trend is at work, undermining Americans' intellectual capital. In the battle to control the hearts and the minds of American educational policymakers and opinion shapers in the mass media, it seems imperative that the combatants temper their passions with a close inspection of the data.

[10] Although anecdotal, it seems relevant too that many societies throughout the world have far more rigid social structures than the United States, with marriages strictly confined within social strata. Despite centuries of such social immobility, there does not appear to have been any obvious deleterious consequences.

REFERENCES

Berliner, D. C., & Biddle, B. J. (1995). *The manufactured crisis: Myths, fraud, and the attack on America's public schools.* Reading, MA: Addison-Wesley.

Bronfenbrenner, U., McClelland, P., Wethington, E., Moen, P., & Ceci, S. J. (1996). *The state of Americans: This generation and the next.* New York: Free Press.

Ceci, S. J. (1996). *On intelligence: A bioecological treatise on intellectual development* (Expanded ed.). Cambridge, MA: Harvard University Press.

Ceci, S. J., Rosenblum, T. B., & Kumpf, M. (in press). The shrinking gap between high- and low-scoring groups: Current trends and possible causes. In U. Neisser (Ed.), *Intelligence on the rise: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Ceci, S. J., & Williams, W. M. (1997). Schooling, intelligence, and income. *American Psychologist, 52,* 1051–1058.

Cook, M. D., & Evans, W. N. (1996, July). *Families or schools? Explaining the convergence in White and Black academic performance* (Working Paper). Monterey, CA: Naval Postgraduate School.

Cooley, W. W., & Lohnes, P. R. (1976). *Evaluation research in education.* New York: Irvington Press.

Dickens, W. T., Kane, T. J., & Schultze, C. L. (1995, Summer). Ring true? A closer look at a grim portrait of American society. *The Brookings Review, 13,* 18–23.

Durlauf, S., Arrow, K., & Bowles, S. (Eds.). (in press). *Meritocracy and equality.* Princeton, NJ: Princeton University Press.

Eysenck, H. J. (1982). *A model for intelligence.* New York: Springer-Verlag.

Fischer, C., Houts, M., Chodrow, N., & Duster, T. (1996). *Inequality by design: Cracking the myth of the bell curve.* Princeton, NJ: Princeton University Press.

Flynn, J. R. (1991). *Asian-Americans: Achievement beyond IQ.* Hillsdale, NJ: Erlbaum.

Flynn, J. (in press). IQ trends over time: Intelligence, race, and meritocracy. In S. Durlauf, K. Arrow, & S. Bowles (Eds.), *Meritocracy and equality.* Princeton, NJ: Princeton University Press.

Fraser, S. (1995). *The bell curve wars.* New York: Basic Books.

Galton, F. (1892). *Hereditary genius.* London: Macmillan.

Gottfredson, L. S. (in press). Why *g* matters: The complexity of everyday life. *Intelligence.*

Gould, S. J. (1981). *The mismeasure of man.* New York: Norton.

Grissmer, D. W., Williamson, S., Kirby, S. N., & Berends, M. (1994). *Student achievement and the changing American family.* Washington, DC: RAND Institute on Education and Training.

Grissmer, D. W., Williamson, S., Kirby, S. N., & Berends, M. (in press). Exploring the rapid rise in Black achievement scores in the United States (1970–1990). In U. Neisser (Ed.), *Intelligence on the rise: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Hanushek, E. A., & Rivkin, S. G. (1997). Understanding the twentieth century growth in U.S. school spending. *Journal of Human Resources, 32,* 35–68.

Hauser, R. M., & Huang, M.-H. (in press). Trends in Black–White score differentials. In U. Neisser (Ed.), *Intelligence on the rise: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Hayes, D. P., Wolfer, L. T., & Wolfe, M. F. (1996). Schoolbook simplification and its relation to the decline in SAT-Verbal scores. *American Educational Research Journal, 33,* 1–18.

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

Hunt, E. (1995). *Will we be smart enough?* New York: Sage.

Itzkoff, S. W. (1989). *The making of the civilized mind.* New York: Longmans.

Jacoby, R., & Glauberman, N. (1995). *The bell curve debate.* New York: Random House.

Jensen, A. R. (1985). The nature of Black–White differences on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences, 8,* 193–263.

Loehlin, J. C. (in press). Whither dysgenics? Comments on Lynn and Preston. In U. Neisser (Ed.), *Intelligence on the rise: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

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

Lynn, R. (1991). Race differences in intelligence: A global perspective. *Mankind Quarterly, 31,* 255–296.

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

Lynn, R. (in press). The decline of genotypic intelligence. In U. Neisser (Ed.), *Intelligence on the rise: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

National Center for Education Statistics. (1991). *National Assessment of Educational Progress (NAEP), Trends in academic progress, 1970–1990.* Washington, DC: U.S. Department of Education.

National Center for Education Statistics. (1996). *National Assessment of Educational Progress (NAEP), 1994 long-term trend assessment.* Washington, DC: U.S. Department of Education.

Neisser, U., Boodoo, G., Bouchard, T. J., Jr., 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.

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

Preston, S. H. (in press). Differential fertility by IQ and the IQ distribution of a population. In U. Neisser (Ed.), *Intelligence on the rise: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Reed, T. E., & Jensen, A. R. (1992). Conduction velocity in a brain nerve pathway of normal adults correlates with intelligence level. *Intelligence, 16,* 259–272.

Reed, T. E., & Jensen, A. R. (1993). Choice reaction time and visual pathway conduction velocity both correlate with intelligence but appear not to correlate with each other: Implications for information processing. *Intelligence, 17,* 191–203.

Rushton, J. P. (1995). *Race, evolution, and behavior: A life history perspective.* New Brunswick, NJ: Transaction.

Rushton, J. P., & Ankney, C. D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class, and race. *Psychonomic Bulletin and Review, 3,* 21–36.

Seligman, D. (1992). *A question of intelligence: The IQ debate in America.* New York: Birch Lane Press.

Solomon, R. J. (1983). *Information concerning the mean test scores for the GMAT, GRE, LSAT, PSAT, and SAT for the National Commission on Excellence in Education.* Princeton, NJ: Educational Testing Service.

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

Vining, D. R. (1982). On the possibility of reemergence of a dysgenic trend with respect to intelligence in American fertility differentials. *Intelligence, 6,* 241–264.

Waldman, I. D. (in press). Complexities in inferring dysgenic trends for intelligence. In U. Neisser (Ed.), *Intelligence on the rise: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Waller, J. H. (1971). Achievement and social mobility: Relationships among IQ score, education, and occupation in two generations. *Social Biology, 18,* 252–259.

Williams, W. M. (in press). Are we raising smarter children today? School- and home-related influences on IQ. In U. Neisser (Ed.), *Intelligence on the rise: Long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.