Intelligence: Knowns and Unknowns

ULRIC NEISSER • GWYNETH BOODOO • THOMAS J. BOUCHARD, JR. • A. WADE BOYKIN • NATHAN BRODY • STEPHEN J. CECI • DIANE F. HALPERN • JOHN C. LOEHLIN • ROBERT PERLOFF • ROBERT J. STERNBERG • SUSANA URBINA

Of the many individual differences that distinguish one person from another, none has produced a stormier and more prolonged debate than intelligence. During the entire twentieth century psychologists have sought valid measures of children’s intelligence that would not be influenced by family background or cultural origin. Unfortunately, this effort has met with limited success, and has fueled an ongoing debate.

The following article is part of a report commissioned by the American Psychological Association and written by a team of scholars knowledgeable about the construct and study of intelligence. The group was convened to respond to questions raised in the mid-1990s in the book The Bell Curve. In this book, ethnic differences in IQ were attributed to biological differences in intellectual capacity.

As the report explains, answers to some of these questions are known and indicate that nurture and environment are powerful forces in organizing and directing human intelligence. Answers to other questions remain unknown. These “unknowns” will undoubtedly set the research agenda in the area of intelligence for decades to come. In the meantime, the debate continues on this important, and controversial, area of development.

In the fall of 1994, the publication of Richard Herrnstein and Charles Murray’s book The Bell Curve sparked a new round of debate about the meaning of intelligence test scores and the nature of intelligence. The debate was characterized by strong assertions as well as by strong feelings. Unfortunately, those assertions often revealed serious misunderstandings of what has (and has not) been demonstrated by scientific research in this field. Although a great deal is now known, the issues remain complex and in many cases still unresolved.

Another unfortunate aspect of the debate was that many participants made little effort to distinguish scientific issues from political ones. Research findings were often assessed not so much on their merits or their scientific standing as on their supposed political implications. In such a climate, individuals who wish to make their own judgments find it hard to know what to believe.

Reviewing the intelligence debate at its meeting of November 1994, the Board of Scientific Affairs (BSA) of the American Psychological Association (APA) concluded that there was urgent need for an authoritative report on these issues—ones that all sides could use as a basis for discussion. Acting by unanimous vote, BSA established a Task Force charged with preparing such a report. Ulric Neisser, Professor of Psychology at Emory University and a member of BSA, was appointed Chair.


This is a “Report of a Task Force Established by the American Psychological Association.”

The Task Force appreciates the contributions of many members of the APA Board of Scientific Affairs (BSA) and the APA Board for the Advancement of Psychology in the Public Interest (RAPPI), who made helpful comments on a preliminary draft of this report. We also wish to acknowledge the indispensable logistical support of the APA Science Directorate during the preparation of the report itself. Correspondence concerning the report should be addressed to Ulric Neisser, Department of Psychology, Emory University, Atlanta, GA 30322. Electronic mail may be sent via Internet to neisser@hel.psy.emory.edu.
The APA Board on the Advancement of Psychology in the Public Interest, which was consulted extensively during this process, nominated one member of the Task Force; the Committee on Psychological Tests and Assessment nominated another; a third was nominated by the Council of Representatives. Other members were chosen by an extended consultative process, with the aim of representing a broad range of expertise and opinion.

The Task Force met twice, in January and March of 1995. Between and after these meetings, drafts of the various sections were circulated, revised, and revised yet again. Disputes were resolved by discussion. As a result, the report presented here has the unanimous support of the entire Task Force.

1. CONCEPTS OF INTELLIGENCE

Individuals differ from one another in their ability to understand complex ideas, to adapt effectively to the environment, to learn from experience, to engage in various forms of reasoning, to overcome obstacles by taking thought. Although these individual differences can be substantial, they are never entirely consistent: a given person's intellectual performance will vary on different occasions, in different domains, as judged by different criteria. Concepts of "intelligence" are attempts to clarify and organize this complex set of phenomena. Although considerable clarity has been achieved in some areas, no such conceptualization has yet answered all the important questions and none commands universal assent. Indeed, when two dozen prominent theorists were recently asked to define intelligence, they gave two dozen somewhat different definitions (Stemberg & Detterman, 1986). Such disagreements are not cause for dismay. Scientific research rarely begins with fully agreed definitions, though it may eventually lead to them.

... Several current theorists argue that there are many different "intelligence" (systems of abilities). Only a few of which can be captured by standard psychometric tests. Others emphasize the role of culture, both in establishing different conceptions of intelligence and in influencing the acquisition of intellectual skills. Developmental psychologists, taking yet another direction, often focus more on the processes by which all children come to think intelligently than on measuring individual differences among them. There is also a new interest in the neural and biological bases of intelligence, a field of research that seems certain to expand in the next few years.

In this brief report, we focus on a limited and rather specific set of questions:

[Note from eds.: Because we have not included the text of this article in its entirety, please see Neisser et al. (1996). Intelligence: Knows and Unknowns. Amer-

PART IV MIDDLE CHILDHOOD

The Psychometric Approach

Ever since Alfred Binet's great success in devising tests to distinguish mentally retarded children from those with behavior problems, psychometric instruments have played an important part in European and American life. Tests are used for many purposes, such as selection, diagnosis, and evaluation. Many of the most widely used tests are not intended to measure intelligence itself but some closely related construct: scholastic aptitude, school achievement, specific abilities, etc. Such tests are especially important for selection purposes. For preparatory school, it's the SSAT; for college, the SAT or ACT; for graduate school, the GRE; for medical school, the MCAT; for law school, the LSAT; for business school, the GMAT. Scores on intelligence-related tests matter, and the stakes can be high.
Intelligence Tests. Tests of intelligence itself (in the psychometric sense) come in many forms. Some use only a single type of item or question; examples include the Peabody Picture Vocabulary Test (a measure of children's verbal intelligence) and Raven's Progressive Matrices (a nonverbal, untimed test that requires inductive reasoning about perceptual patterns). Although such instruments are useful for specific purposes, the more familiar measures of general intelligence—such as the Wechsler tests and the Stanford-Binet—include many different types of items, both verbal and nonverbal. Test-takers may be asked to give the meanings of words, to complete a series of pictures, to indicate which of several words does not belong with the others, and the like. Their performance can then be scored to yield several sub-scores as well as an overall score.

By convention, overall intelligence test scores are usually converted to a scale in which the mean is 100 and the standard deviation is 15. (The standard deviation is a measure of the variability of the distribution of scores.) Approximately 95% of the population has scores within two standard deviations of the mean; i.e., between 70 and 130. For historical reasons, the term "IQ" is often used to describe scores on tests of intelligence. It originally referred to an "Intelligence Quotient" that was formed by dividing a so-called mental age by a chronological age, but this procedure is no longer used.

Intercorrelations Among Tests. Individuals rarely perform equally well on all the different kinds of items included in a test of intelligence. One person may do relatively better on verbal than on spatial items, for example, while another may show the opposite pattern. Nevertheless, subtests measuring different abilities tend to be positively correlated: people who score high on one such subtest are likely to be above average on others as well. These complex patterns of correlation can be clarified by factor analysis, but the results of such analyses are often controversial themselves. Some theorists (e.g., Spearman, 1927) have emphasized the importance of a general factor, g, which represents what all the tests have in common; others (e.g., Thurstone, 1938) focus on more specific group factors such as memory, verbal comprehension, or number facility. As we shall see in Section 2, one common view today envisions something like a hierarchy of factors with g at the apex. But there is no full agreement on what g actually means: it has been described as a mere statistical regularity (Thomson, 1939), a kind of mental energy (Spearman, 1927), a generalized abstract reasoning ability (Gustafson, 1984), or an index measure of neural processing speed (Reed & Jensen, 1992).

There have been many disputes over the utility of IQ and g. Some theorists are critical of the entire psychometric approach (e.g., Ceci, 1990; Gardner, 1983; Gould, 1978), while others regard it as firmly established (e.g., Carroll, 1993; Eysenck, 1973; Herrnstein & Murray, 1994; Jensen, 1972). The critics do not dispute the stability of test scores, nor the fact that they predict certain forms of achievement—especially school achievement—rather effectively (see Section 2). They do argue, however, that to base a concept of intelligence on test scores alone is to ignore many important aspects of mental ability.

2. Intelligence Tests and Their Correlates

The correlation coefficient, r, can be computed whenever the scores in a sample are paired in some way. Typically this is because each individual is measured twice: he or she takes the same test on two occasions, or takes two different tests, or has both a test score and some criterion measure such as grade point average or job performance. (In Section 3 we consider cases where the paired scores are those of two different individuals, such as twins or parent and child.)

The value of r measures the degree of relationship between the two sets of scores in a convenient way, by assessing how well one of them (computationally it doesn't matter which one) could be used to predict the value of the other. Its sign indicates the direction of relationship: when r is negative, high scores on one measure predict low scores on the other. Its magnitude indicates the strength of the relationship. If r = 0, there is no relation at all; if r is 1 (or -1), one score can be used to predict the other score perfectly. Moreover, the square of r has a particular meaning in cases where we are concerned with predicting one variable from another. When r = .50, for example, r^2 is .25; this means (given certain linear assumptions) that 25% of the variance in one set of scores is predictable from the correlated values of the other set, while the remaining 75% is not.

Basic Characteristics of Test Scores

Stability. Intelligence test scores are fairly stable during development. When Jones and Bayley (1941) tested a sample of children annually throughout childhood and adolescence, for example, scores obtained at age 18 were correlated r = .77 with scores that had been obtained at age 6 and r = .89 with scores from age 12. When scores were averaged across several successive tests to remove short-term fluctuations, the correlations were even higher. The mean for ages 17 and 18 was correlated r = .86 with the mean for ages 5, 6, and 7, and r = .96 with the
mean for ages 11, 12, and 13. (For comparable findings in a more recent study, see Moffitt, Causi, Harkness, & Silva, 1993.) Nevertheless, IQ scores do change over time. In the same study (Jones & Bayley, 1941), the average change between age 12 and age 17 was 7.1 IQ points; some individuals changed as much as 18 points.

It is important to understand what remains stable and what changes in the development of intelligence. A child whose IQ score remains the same from age 6 to age 18 does not exhibit the same performance throughout that period. On the contrary, steady gains in general knowledge, vocabulary, reasoning ability, etc. will be apparent. What does not change is his or her score in comparison to that of other individuals of the same age. A six-year-old with an IQ of 100 is at the mean of six-year-olds; an 18-year-old with that score is at the mean of 18-year-olds.

Factors and g. As noted in Section 1, the patterns of intercorrelation among tests (i.e., among different kinds of items) are complex. Some pairs of tests are much more closely related than others, but all such correlations are typically positive and form what is called a “positive manifold” (Spearman, 1927). This means that in any such manifold, some portion of the variance of scores on each test can be mathematically attributed to a “general factor,” or g. Given this analysis, the overall pattern of correlations can be roughly described as produced by individual differences in g plus differences in the specific abilities sampled by particular tests. In addition, however, there are usually patterns of intercorrelation among groups of tests. These commonalities, which played only a small role in Spearman’s analysis, were emphasized by other theorists. Thurstone (1938), for example, proposed an analysis based primarily on the concept of group factors.

While some psychologists today still regard g as the most fundamental measure of intelligence (e.g., Jensen, 1980), others prefer to emphasize the distinctive profile of strengths and weaknesses present in each person’s performance. A recently published review identifies over 70 different abilities that can be distinguished by currently available tests (Carroll, 1993). One way to represent this structure is in terms of a hierarchical arrangement with a general intelligence factor at the apex and various more specialized abilities arrayed below it. Such a summary merely acknowledges that performance levels on different tests are correlated; it is consistent with, but does not prove, the hypothesis that a common factor such as g underlies those correlations. Different specialized abilities might also be correlated for other reasons, such as the effects of education. Thus while the g-based factor hierarchy is the most widely accepted current view of the structure of abilities, some theorists regard it as misleading (Ceci, 1990). Moreover, as noted in Section 1, a wide range of human abilities—including many that seem to have intellectual components—are outside the domain of standard psychometric tests.

Tests As Predictors

School Performance. Intelligence tests were originally devised by Alfred Binet to measure children’s ability to succeed in school. They do in fact predict school performance fairly well: the correlation between IQ scores and grades is about .50. They also predict scores on school achievement tests, designed to measure knowledge of the curriculum. Note, however, that correlations of this magnitude account for only about 25% of the overall variance. Successful school learning depends on many personal characteristics other than intelligence, such as persistence, interest in school, and willingness to study. The encouragement for academic achievement that is received from peers, family, and teachers may also be important, together with more general cultural factors (see Section 5).

Years of Education. Some children stay in school longer than others; many go on to college and perhaps beyond. Two variables that can be measured as early as elementary school correlate with the total amount of education individuals will obtain: test scores and social class background. Correlations between IQ scores and total years of education are about .55, implying that differences in psychometric intelligence account for about 30% of the outcome variance. The correlations of years of education with social class background (as indexed by the occupation/education of a child’s parents) are also positive, but somewhat lower.

There are a number of reasons why children with higher test scores tend to get more education. They are likely to get good grades, and to be encouraged by teachers and counselors; often they are placed in “college preparatory” classes, where they make friends who may also encourage them. In general, they are likely to find the process of education rewarding in a way that many low-scoring children do not (Rehberg & Rosenthal, 1978). These influences are not omnipotent: some high scoring children do drop out of school. Many personal and social characteristics other than psychometric intelligence determine academic success and interest, and social privilege may also play a role. Nevertheless, test scores are the best single predictor of an individual’s years of education.

Social Status and Income. How well do IQ scores (which can be obtained before individuals enter the labor force) predict such outcome measures as the social status or income of adults? This question is com-
plex, in part because another variable also predicts such outcomes: namely, the socioeconomic status (SES) of one's parents. Unsurprisingly, children of privileged families are more likely to attain high social status than those whose parents are poor and less educated. These two predictors (IQ and parental SES) are by no means independent of one another; the correlation between them is around .33 (White, 1982).

One way to look at these relationships is to begin with SES. According to Jencks (1979), measures of parental SES predict about one-third of the variance in young adults' social status and about one-fifth of the variance in their income. About half of this predictive effectiveness depends on the fact that the SES of parents also predicts children's intelligence test scores, which have their own predictive value for social outcomes; the other half comes about in other ways.

We can also begin with IQ scores, which by themselves account for about one-fourth of the social status variance and one-sixth of the income variance. Statistical controls for parental SES eliminate only about a quarter of this predictive power. One way to conceptualize this effect is by comparing the occupational status (or income) of adult brothers who grew up in the same family and hence have the same parental SES. In such cases, the brother with the higher adolescent IQ score is likely to have the higher adult social status and income (Jencks, 1979). This effect, in turn, is substantially mediated by education: the brother with the higher test scores is likely to get more schooling, and hence to be better credentialled as he enters the workplace.

**Job Performance.** Scores on intelligence tests predict various measures of job performance: supervisor ratings, work samples, etc. Such correlations, which typically lie between \( r = .30 \) and \( r = .50 \), are partly restricted by the limited reliability of those measures themselves. They become higher when \( r \) is statistically corrected for unreliability: in one survey of relevant studies (Hunter, 1983), the mean of the corrected correlations was .54. This implies that, across a wide range of occupations, intelligence test performance accounts for some 29% of the variance in job performance.

**Social Outcomes.** Psychometric intelligence is negatively correlated with certain socially undesirable outcomes. For example, children with high test scores are less likely than lower-scoring children to engage in juvenile crime. In one study, Moffitt, Gabrielli, Mednick, and Schulzinger (1981) found a correlation of \( -19 \) between IQ scores and number of juvenile offenses in a large Danish sample; with social class controlled, the correlation dropped to \( -17 \). The correlations for most "negative outcome" variables are typically smaller than .20, which means that test scores are associated with less than 4% of their total variance. It is important to realize that the causal links between psychometric ability and social outcomes may be indirect. Children who are unsuccessful in—and hence alienated from—school may be more likely to engage in delinquent behaviors for that very reason, compared to other children who enjoy school and are doing well.

In summary, intelligence test scores predict a wide range of social outcomes with varying degrees of success. Correlations are highest for school achievement, where they account for about a quarter of the variance. They are somewhat lower for job performance, and very low for negatively valued outcomes such as criminality. In general, intelligence tests measure only some of the many personal characteristics that are relevant to life in contemporary America. Those characteristics are never the only influence on outcomes, though in the case of school performance they may well be the strongest.

**Test Scores and Measures of Processing Speed**

Many recent studies show that the speed with which people perform very simple perceptual and cognitive tasks is correlated with psychometric intelligence (for reviews see Ceci, 1990; Deary, 1995; Vernon, 1987). In general, people with higher intelligence test scores tend to apprehend, scan, retrieve, and respond to stimuli more quickly than those who score lower.

**Problems of Interpretation.** Some researchers believe that psychometric intelligence, especially g, depends directly on what may be called the "neural efficiency" of the brain (Eysenck, 1986; Vernon, 1987). They regard the observed correlations between test scores and measures of processing speed as evidence for their view. If choice reaction times, inspection times, and VEP latencies actually do reflect the speed of basic neural processes, such correlations are only to be expected. In fact, however, the observed patterns of correlation are rarely as simple as this hypothesis would predict. Moreover, it is quite possible that high- and low-IQ individuals differ in other ways that affect speeded performance (cf. Ceci, 1990). Those variables include motivation, response criteria (emphasis on speed vs. accuracy), perceptual strategies (cf. Mackenzie et al., 1991), attentional strategies, and—in some cases—differential familiarity with the material itself. Finally, we do not yet know the direction of causation that underlies such correlations. Do high levels of "neural efficiency" promote the
development of intelligence, or do more intelligent people simply find faster ways to carry out perceptual tasks? Or both? These questions are still open.

3. THE GENES AND INTELLIGENCE

... We focus here on the relative contributions of genes and environments to individual differences in particular traits. To avoid misunderstanding, it must be emphasized from the outset that gene action always involves an environment—at least a biochemical environment, and often an ecological one. (For humans, that ecology is usually interpersonal or cultural.) Thus all genetic effects on the development of observable traits are potentially modifiable by environmental input, though the practicability of making such modifications may be another matter. Conversely, all environmental effects on trait development involve the genes or structures to which the genes have contributed. Thus there is always a genetic aspect to the effects of the environment (cf. Fomin & Bergeman, 1991).

RESULTS FOR IQ SCORES
Parameter Estimates. Across the ordinary range of environments in modern Western societies, a sizable part of the variance in intelligence test scores is associated with genetic differences among individuals. Quantitative estimates vary from one study to another, because many are based on small or selective samples. If one simply combines all available correlations in a single analysis, the heritability (h^2) works out to about .50 and the between-family variance (c^2) to about .25 (e.g., Chipuer, Rivine, & Fomin, 1990; Loehlin, 1989). These overall figures are misleading, however, because most of the relevant studies have been done with children. We now know that the heritability of IQ changes with age: h^2 goes up and c^2 goes down from infancy to adulthood (McCartney, Harris, & Bernieri, 1990; McGue, Bouchard, Iacono, & Lykken, 1993). In childhood h^2 and c^2 for IQ are of the order of .45 and .35; by late adolescence h^2 is around .75 and c^2 is quite low (zero in some studies). Substantial environmental variance remains, but it primarily reflects within-family rather than between-family differences.

These adult parameter estimates are based on a number of independent studies. The correlation between MZ twins reared apart, which directly estimates h^2, ranged from .66 to .78 in five studies involving adult samples from Europe and the United States (McGuie et al., 1993). The correlation between unrelated children reared together in adoptive families, which directly estimates c^2, was approximately zero for adolescents in two adoption studies (Loehlin, Horn, & Willerman, 1989; Scarr & Weinberg, 1978). And 19 in a third (the Minnesota transracial adoption study; Scarr, Weinberg, & Waldman, 1993). These particular estimates derive from samples in which the lower socioeconomic levels were underrepresented (i.e., there were few very poor families), so the range of between-family differences was smaller than in the population as a whole. This means that we should be cautious in generalizing the findings for between-family effects across the entire social spectrum. The samples were also mostly White, but available data suggest that twin and sibling correlations in African American and similarly selected White samples are more often comparable than not (Loehlin, Lindzey, & Spuhler, 1975).

Why should individual differences in intelligence (as measured by test scores) reflect genetic differences more strongly in adults than they do in children? One possibility is that as individuals grow older their transactions with their environments are increasingly influenced by the characteristics that they bring to those environments themselves, decreasingly by the conditions imposed by family life and social origins. Older persons are in a better position to select their own effective environments, a form of genotype-environment correlation. In any case the popular view that genetic influences on the development of a trait are essentially frozen at conception while the effects of the early environment cumulate inexorably is quite misleading, at least for the trait of psychometric intelligence.

Implications. Estimates of h^2 and c^2 for IQ (or any other trait) are descriptive statistics for the populations studied. (In this respect they are like means and standard deviations.) They are outcome measures, summarizing the results of a great many diverse, intricate, individually variable events and processes, but they can nevertheless be quite useful. They can tell us how much of the variation in a given trait the genes and family environments explain, and changes in them place some constraints on theories of how this occurs. On the other hand they have little to say about specific mechanisms, i.e., about how genetic and environmental differences get translated into individual physiological and psychological differences. Many psychologists and neuroscientists are actively studying such processes; data on heritabilities may give them ideas about what to look for and where or when to look for it.

A common error is to assume that because something is heritable it is necessarily unchangeable. This is wrong. Heritability does not imply immutability. As previously noted, heritable traits can depend on learning, and they may be subject to other environmental effects as well. The value of h^2 can change if the distribution of environments (or genes) in the population
Intelligence: Knowns and Unknowns

is substantially altered. On the other hand, there can be effective environmental changes that do not change heritability at all. If the environment relevant to a given trait improves in a way that affects all members of the population equally, the mean value of the trait will rise without any change in its heritability (because the differences among individuals in the population will stay the same). This has evidently happened for height: the heritability of stature is high, but average heights continue to increase (Olivier, 1980). Something of the sort may also be taking place for IQ scores—the so-called "Flynn effect" discussed in Section 4.

4. ENVIRONMENTAL EFFECTS ON INTELLIGENCE

The "environment" includes a wide range of influences on intelligence. Some of those variables affect whole populations, while others contribute to individual differences within a given group. Some of them are social, some are biological, at this point some are still mysterious. It may also happen that the proper interpretation of an environmental variable requires the simultaneous consideration of genetic effects. Nevertheless, a good deal of solid information is available.

SOCIAL VARIABLES

It is obvious that the cultural environment—how people live, what they value, what they do—has a significant effect on the intellectual skills developed by individuals. Rice farmers in Liberia are good at estimating quantities of rice (Gay & Cole, 1967); children in Botswana, accustomed to story-telling, have excellent memories for stories (Dube, 1982). Both these groups were far ahead of American controls on the tasks in question. On the other hand Americans and other Westernized groups typically outperform members of traditional societies on psychometric tests, even those designed to be "culture-fair."

Cultures typically differ from one another in so many ways that particular differences can rarely be ascribed to single causes. Even comparisons between subpopulations can be difficult to interpret. If we find that middle-class and poor Americans differ in their scores on intelligence tests, it is easy to suppose that the environmental difference has caused the IQ difference (i.e., that growing up in the middle class produces higher psychometric intelligence than growing up poor). But there may also be an opposite direction of causation: individuals can come to be in one environment or another because of differences in their own abilities. Waller (1971) has shown, for example, that adult sons whose IQ scores are above those of their fathers tend to have higher social-class status than those fathers; conversely, sons with IQ scores below their fathers' tend to have lower social-class status. Since all the subjects grew up with their fathers, the IQ differences in this study cannot have resulted from class-related differences in childhood experience. Rather, those differences (or other factors correlated with them) seem to have had an influence on the status that they achieved. Such a result is not surprising, given the relation between test scores and years of education reviewed in Section 2.

Occupation. In Section 2 we noted that intelligence test scores predict occupational level, not only because some occupations require more intelligence than others but also because admission to many professions depends on test scores in the first place. There can also be an effect in the opposite direction, i.e., workplaces may affect the intelligence of those who work in them. Kohn and Schooler (1973), who interviewed some 3,000 men in various occupations (farmers, managers, machinists, porters, etc.), argued that more "complex" jobs produce more "intellectual flexibility" in the individuals who hold them. Although the issue of direction of effects was not fully resolved in their study—and perhaps not even in its longitudinal follow-up (Kohn & Schooler, 1983)—this remains a plausible suggestion.

Schooling. Attendance at school is both a dependent and an independent variable in relation to intelligence. On the one hand, children with higher test scores are less likely to drop out and more likely to be promoted from grade to grade and then to attend college. Thus the number of years of education that adults complete is roughly predictable from their childhood scores on intelligence tests. On the other hand, schooling itself changes mental abilities, including those abilities measured on psychometric tests. This is obvious for tests like the SAT that are explicitly designed to assess school learning, but it is almost equally true of intelligence tests themselves.

Schools affect intelligence in several ways, most obviously by transmitting information. The answers to questions like "Who wrote Hamlet?" and "What is the boiling point of water?" are typically learned in school, where some pupils learn them more easily and thoroughly than others. Perhaps at least as important are certain general skills and attitudes: systematic problem-solving, abstract thinking, categorization, sustained attention to material of little intrinsic interest, and repeated manipulation of basic symbols and operations. There is no doubt that schools promote and permit the development of significant intellectual skills, which develop to different extents in different children. It is because tests of intelligence draw on many of those same skills that they predict school achievement as well as they do.
To achieve these results, the school experience must meet at least some minimum standard of quality. In very poor schools, children may learn so little that they fall farther behind the national IQ norms for every year of attendance. When this happens, older siblings have systematically lower scores than their younger counterparts. This pattern of scores appeared in at least one rural Georgia school system in the 1970s (Jensen, 1977). Before desegregation, it must have been characteristic of many of the schools attended by Black pupils in the South. In a study based on Black children who had moved to Philadelphia at various ages during this period, Lee (1951) found that their IQ scores went up more than half a point for each year that they were enrolled in the Philadelphia system.

Interventions. Intelligence test scores reflect a child's standing relative to others in his or her age cohort. Very poor or interrupted schooling can lower that standing substantially; are there also ways to raise it? In fact many interventions have been shown to raise test scores and mental ability “in the short run” (i.e., while the program itself was in progress), but long-run gains have proved more elusive. One noteworthy example of (at least short-run) success was the Venezuelan Intelligence Project (Herrnstein, Nickerson, de Sanchez, & Swets, 1986), in which hundreds of seventh-grade children from under-privileged backgrounds in that country were exposed to an extensive, theoretically-based curriculum focused on thinking skills. The intervention produced substantial gains on a wide range of tests, but there has been no follow-up.

Children who participate in “Head Start” and similar programs are exposed to various school-related materials and experiences for one or two years. Their test scores often go up during the course of the program, but these gains fade with time. By the end of elementary school, there are usually no significant IQ or achievement-test differences between children who have been in such programs and controls who have not. There may, however, be other differences. Follow-up studies suggest that children who participated in such programs as preschoolers are less likely to be assigned to special education, less likely to be held back in grade, and more likely to finish high school than matched controls (Consortium for Longitudinal Studies, 1983; Darlington, 1986; but see Locurto, 1991).

More extensive interventions might be expected to produce larger and more lasting effects, but few such programs have been evaluated systematically. One of the more successful is the Carolina Abecedarian Project (Campbell & Ramey, 1994), which provided a group of children with enriched environments from early infancy through preschool and also maintained appropriate controls. The test scores of the enrichment-group children were already higher than those of controls at age two; they were still some 5 points higher at age 12, seven years after the end of the intervention. Importantly, the enrichment group also outperformed the controls in academic achievement.

Family Environment. No one doubts that normal child development requires a certain minimum level of responsible care. Severely deprived, neglectful, or abusive environments must have negative effects on a great many aspects—including intellectual aspects—of development. Beyond that minimum, however, the role of family experience is now in serious dispute (Baumrind, 1993; Jackson, 1993; Scarf, 1992, 1993). Psychometric intelligence is a case in point. Do differences between children's family environments (within the normal range) produce differences in their intelligence test performance? The problem here is to disentangle causation from correlation. There is no doubt that such variables as resources of the home (Gottfried, 1984) and parents' use of language (Fert & Risley, 1992) are correlated with children's IQ scores, but such correlations may be mediated by genetic as well as (or instead of) environmental factors.

**Biological Variables**

Every individual has a biological as well as a social environment, one that begins in the womb and extends throughout life. Many aspects of that environment can affect intellectual development. We now know that a number of biological factors—malnutrition, exposure to toxic substances, various prenatal and perinatal stressors—result in lowered psychometric intelligence under at least some conditions.

Nutrition. There has been only one major study of the effects of prenatal malnutrition (i.e., malnutrition of the mother during pregnancy) on long-term intellectual development. Stein, Susser, Saenger, and Marolla (1975) analyzed the test scores of Dutch 19-year-old males in relation to a wartime famine that had occurred in the winter of 1944–45, just before their birth. In this very large sample (made possible by a universal military induction requirement), exposure to the famine had no effect on adult intelligence. Note, however, that the famine itself lasted only a few months; the subjects were exposed to it prenatally but not after birth.

In contrast, prolonged malnutrition during childhood does have long-term intellectual effects. These have not been easy to establish, in part because many other unfavorable socioeconomic conditions are often associated with chronic malnutrition (Ricciuti, 1993;
but cf. Sigman, 1993). In one intervention study, however, preschoolers in two Guatemalan villages (where undernourishment is common) were given ad lib access to a protein dietary supplement for several years. A decade later, many of these children (namely, those from the poorest socioeconomic levels) scored significantly higher on school-related achievement tests than comparable controls (Pollint, Gorman, Engle, Martorell, & Rivera, 1993). It is worth noting that the effects of poor nutrition on intelligence may well be indirect. Malnourished children are typically less responsive to adults, less motivated to learn, and less active in exploration than their more adequately nourished counterparts. . . .

Lead. Certain toxins have well-established negative effects on intelligence. Exposure to lead is one such factor. In one long-term study (Baghurst et al., 1992; McMichael et al., 1988), the blood lead levels of children growing up near a lead smelting plant were substantially and negatively correlated with intelligence test scores throughout childhood. No "threshold" dose for the effect of lead appears in such studies. Although ambient lead levels in the United States have been reduced in recent years, there is reason to believe that some American children—especially those in inner cities—may still be at risk from this source (cf. Needleman, Geiger, & Frank, 1985).

Alcohol. Extensive prenatal exposure to alcohol (which occurs if the mother drinks heavily during pregnancy) can give rise to fetal alcohol syndrome, which includes mental retardation as well as a range of physical symptoms. Smaller "doses" of prenatal alcohol may have negative effects on intelligence even when the full syndrome does not appear. Streissguth, Barr, Sampson, Darby, and Martin (1989) found that mothers who reported consuming more than 1.5 oz. of alcohol daily during pregnancy had children who scored some 5 points below controls at age four. Prenatal exposure to aspirin and antibiotics had similar negative effects in this study.

Perinatal Factors. Complications at delivery and other negative perinatal factors may have serious consequences for development. Nevertheless, because they occur only rarely, they contribute relatively little to the population variance of intelligence (Broman, Nichols, & Kennedy, 1975). Down's syndrome, a chromosomal abnormality that produces serious mental retardation, is also rare enough to have little impact on the overall distribution of test scores.

The correlation between birth weight and later intelligence deserves particular discussion. In some cases low birth weight simply reflects premature delivery; in others, the infant's size is below normal for its gestational age. Both factors apparently contribute to the tendency of low-birth-weight infants to have lower test scores in later childhood (Lubchenko, 1976). These correlations are small, ranging from .05 to .13 in different groups (Broman et al., 1975). The effects of low birth weight are substantial only when it is very low indeed (less than 1,500 gm). Premature babies born at these very low birth weights are behind controls on most developmental measures; they often have severe or permanent intellectual deficits (Rosetti, 1985).

CONTINUOUSLY RISING TEST SCORES

Perhaps the most striking of all environmental effects is the steady worldwide rise in intelligence test performance. Although many psychometricians had noted these gains, it was James Flynn (1984, 1987) who first described them systematically. His analysis shows that performance has been going up ever since testing began. The "Flynn effect" is now very well documented, not only in the United States but in many other technologically advanced countries. The average gain is about 3 IQ points per decade—more than a full standard deviation since, say, 1940.

Although it is simplest to describe the gains as increases in population IQ, this is not exactly what happens. Most intelligence tests are "restandardized" from time to time, in part to keep up with these very gains. As part of this process the mean score of the new standardization sample is typically set to 100 again, so the increase more or less disappears from view. In this context, the Flynn effect means that if 20 years have passed since the last time the test was standardized, people who now score 100 on the new version would probably average about 106 on the old one.

The sheer extent of these increases is remarkable, and the rate of gain may even be increasing. The scores of 19-year-olds in the Netherlands, for example, went up more than 8 points—over half a standard deviation—between 1972 and 1982. What's more, the largest gains appear on the types of tests that were specifically designed to be free of cultural influence (Flynn, 1987). One of these is Raven's Progressive Matrices, an untimed nonverbal test that many psychometricians regard as a good measure of g.

These steady gains in intelligence test performance have not always been accompanied by corresponding gains in school achievement. Indeed, the relation between intelligence and achievement test scores can be complex. This is especially true for the Scholastic Aptitude Test (SAT), in part because the ability range of the students who take the SAT has broadened over time. That change explains some portion—not all—of the prolonged decline in SAT scores that took place from the mid-1960s to the early 1980s, even as IQ scores were continuing to rise.
5. GROUP DIFFERENCES

Group means have no direct implications for individuals. What matters for the next person you meet (to the extent that test scores matter at all) is that person's own particular score, not the mean of some reference group to which he or she happens to belong. The commitment to evaluate people on their own individual merit is central to a democratic society. It also makes quantitative sense. The distributions of different groups inevitably overlap, with the range of scores within any one group always wider than the mean differences between any two groups. In the case of intelligence test scores, the variance attributable to individual differences far exceeds the variance related to group membership (Jensen, 1980).

Besides European Americans ("Whites"), the ethnic groups to be considered are Chinese and Japanese Americans, Hispanic Americans ("Latinos"), Native Americans ("Indians"), and African Americans ("Blacks"). These groups (we avoid the term "race") are defined and self-defined by social conventions based on ethnic origin as well as on observable physical characteristics such as skin color. None of them are internally homogeneous. Asian Americans, for example, may have roots in many different cultures: not only China and Japan but also Korea, Laos, Vietnam, the Philippines, India, and Pakistan. Hispanic Americans, who share a common linguistic tradition, actually differ along many cultural dimensions. In their own minds they may be less "Latinos" than Puerto Ricans, Mexican Americans, Cuban Americans, or representatives of other Latin cultures. "Native American" is an even more diverse category, including a great many culturally distinct tribes living in a wide range of environments.

MEAN SCORES OF DIFFERENT ETHNIC GROUPS

Asian Americans. In the years since the Second World War, Asian Americans—especially those of Chinese and Japanese extraction—have compiled an outstanding record of academic and professional achievement. This record is reflected in school grades, in scores on content-oriented achievement tests like the SAT and GRE, and especially in the disproportionate representation of Asian Americans in many sciences and professions. Although it is often supposed that these achievements reflect correspondingly high intelligence test scores, this is not the case. In more than a dozen studies from the 1960s and 1970s analyzed by Flynn (1991), the mean IQs of Japanese and Chinese American children were always around 97 or 98; none was over 100. Even Lynn (1993), who argues...
Intelligence: Knows and Unknowns

for a slightly higher figure, concedes that the achievements of these Asian Americans far outstrip what might have been expected on the basis of their test scores.

It may be worth noting that the interpretation of test scores obtained by Asians in Asia has been controversial in its own right. Lynn (1982) reported a mean Japanese IQ of 111 while Flynn (1991) estimated it to be between 101 and 105. Stevenson et al. (1985), comparing the intelligence-test performance of children in Japan, Taiwan, and the United States, found no substantive differences at all. Given the general problems of cross-cultural comparison, there is no reason to expect precision or stability in such estimates. Nevertheless, some interest attaches to these particular comparisons: they show that the well-established differences in school achievement among the same three groups (Chinese and Japanese children are much better at math than American children) do not simply reflect differences in psychometric intelligence. Stevenson, Lee, and Stigler (1986) suggest that they result from structural differences in the schools of the three nations as well as from varying cultural attitudes toward learning itself. It is also possible that spatial ability—in which Japanese and Chinese obtain somewhat higher scores than Americans—plays a particular role in the learning of mathematics.

Hispanic Americans. Hispanic immigrants have come to America from many countries. In 1993, the largest Latino groups in the continental United States were Mexican Americans (64%), Puerto Ricans (11%), Central and South Americans (13%), and Cubans (5%) (U.S. Bureau of the Census, 1994). There are very substantial cultural differences among these nationality groups, as well as differences in academic achievement (Durán, 1983; United States National Commission for Employment Policy, 1982). Taken together, Latinos make up the second largest and the fastest-growing minority group in America (Davis, Haub, & Willette, 1983; Ede, 1992).

In the United States, the mean intelligence test scores of Hispanics typically lie between those of Blacks and Whites. There are also differences in the patterning of scores across different abilities and subtests (Hennessy & Merrifield, 1978; Lesser, Fifer, & Clark, 1965). Linguistic factors play a particularly important role for Hispanic Americans, who may know relatively little English. (By one estimate, 25% of Puerto Ricans and Mexican Americans and at least 40% of Cubans speak English "not well" or "not at all" [Rodriguez, 1992].) Even those who describe themselves as bilingual may be at a disadvantage if Spanish was their first and best-learned language. It is not surprising that Latino children typically score higher on the performance than on the verbal subtests of the English-based Wechsler Intelligence Scale for Children—Revised (WISC—R; Kaufman, 1994). Nevertheless, the predictive validity of Latino test scores is not negligible. In young children, the WISC—R has reasonably high correlations with school achievement measures (McShane & Cook, 1985). For high school students of moderate to high English proficiency, standard aptitude tests predict first-year college grades about as well as they do for non-Hispanic Whites (Penneck-Roman, 1992).

Native Americans. There are a great many culturally distinct North American Indian tribes (Driver, 1969), speaking some 200 different languages (Leap, 1981). Many Native Americans live on reservations, which themselves represent a great variety of ecological and cultural settings. Many others presently live in metropolitan areas (Brandt, 1984). Although few generalizations can be appropriate across so wide a range, two or three points seem fairly well established. The first is a specific relation between ecology and cognition: the Inuit (Eskimo) and other groups that live in the arctic tend to have particularly high visual-spatial skills. (For a review see McShane & Berry, 1988.) Moreover, there seem to be no substantial sex differences in those skills (Berry, 1974). It seems likely that this represents an adaptation—genetic or learned or both—to the difficult hunting, traveling, and living conditions that characterize the arctic environment.

On the average, Indian children obtain relatively low scores on tests of verbal intelligence, which are often administered in school settings. The result is a performance-test/verbal-test discrepancy similar to that exhibited by Hispanic Americans and other groups whose first language is generally not English. Moreover, many Indian children suffer from chronic middle-ear infection (otitis media), which is "the leading identifiable disease among Indians since record-keeping began in 1952" (McShane & Pas, 1984a, p. 84). Hearing loss can have marked negative effects on verbal test performance (McShane & Pas, 1984b).

African Americans. The relatively low mean of the distribution of African American intelligence test scores has been discussed for many years. Although studies using different tests and samples yield a range of results, the Black mean is typically about one standard deviation (about 15 points) below that of Whites (Jensen, 1980; Loehlin et al., 1975; Reynolds et al., 1987). The difference is largest on those tests (verbal or nonverbal) that best represent the general intelligence factor g (Jensen, 1985). It is possible, however, that this differential is diminishing. In the most recent restandardization of the Stanford-Binet test, the Black/White differential was 13 points for younger children and 10 points for older children (Thomndike, Hagen, & Sattler, 1986). In several other studies of
children since 1980, the Black mean has consistently been over 90 and the differential has been in single digits (Vincent, 1991). Larger and more definitive studies are needed before this trend can be regarded as established.

Another reason to think the IQ mean might be changing is that the Black/White differential in achievement scores has diminished substantially in the last few years. Consider, for example, the mathematics achievement of 17-year-olds as measured by the National Assessment of Educational Progress (NAEP). The differential between Black and White scores, about 1.1 standard deviations as recently as 1978, had shrunk to .65 SD by 1990 (Grissmer et al., 1994) because of Black gains. Hispanics showed similar but smaller gains; there was little change in the scores of Whites. Other assessments of school achievement also show substantial recent gains in the performance of minority children.

In their own analysis of these gains, Grissmer et al. (1994) cite both demographic factors and the effects of public policy. They found the level of parents’ education to be a particularly good predictor of children’s school achievement; that level increased for all groups between 1970 and 1990, but most sharply for Blacks. Family size was another good predictor (children from smaller families tend to achieve higher scores); here too, the largest change over time was among Blacks. Above and beyond these demographic effects, Grissmer et al. believe that some of the gains can be attributed to the many specific programs, geared to the education of minority children, that were implemented during that period.

Test Bias. It is often argued that the lower mean scores of African Americans reflect a bias in the intelligence tests themselves. This argument is right in one sense of “bias” but wrong in another. To see the first of these, consider how the term is used in probability theory. When a coin comes up heads consistently for any reason it is said to be “biased,” regardless of any consequences that the outcome may or may not have. In this sense the Black/White score differential is ipso facto evidence of what may be called “outcome bias.” African Americans are subject to outcome bias not only with respect to tests but along many dimensions of American life. They have the short end of nearly every stick: average income, representation in high-level occupations, health and health care, death rate, confrontations with the legal system, and so on. With this situation in mind, some critics regard the test score differential as just another example of a pervasive outcome bias that characterizes our society as a whole (Jackson, 1975; Mercer, 1984). Although there is a sense in which they are right, this critique ignores the particular social purpose that tests are designed to serve.

From an educational point of view, the chief function of mental tests is as predictors (Section 2). Intelligence tests predict school performance fairly well, at least in American schools as they are now constituted. Similarly, achievement tests are fairly good predictors of performance in college and postgraduate settings. Considered in this light, the relevant question is whether the tests have a “predictive bias” against Blacks. Such a bias would exist if African American performance on the criterion variables (school achievement, college GPA, etc.) were systematically higher than the same subjects’ test scores would predict. This is not the case. The actual regression lines (which show the mean criterion performance for individuals who got various scores on the predictor) for Blacks do not lie above those for Whites; there is even a slight tendency in the other direction (Jensen, 1980; Reynolds & Brown, 1984). Considered as predictors of future performance, the tests do not seem to be biased against African Americans.

Characteristics of Tests. It has been suggested that various aspects of the way tests are formulated and administered may put African Americans at a disadvantage. The language of testing is a standard form of English with which some Blacks may not be familiar; specific vocabulary items are often unfamiliar to Black children; the tests are often given by White examiners rather than by more familiar Black teachers; African Americans may not be motivated to work hard on tests that so clearly reflect White values; the time demands of some tests may be alien to Black culture. (Similar suggestions have been made in connection with the test performance of Hispanic Americans, e.g., Rodriguez, 1992.) Many of these suggestions are plausible, and such mechanisms may play a role in particular cases. Controlled studies have shown, however, that none of them contributes substantially to the Black/White differential under discussion here (Jensen, 1980; Reynolds & Brown, 1984; for a different view see Helms, 1992). Moreover, efforts to devise reliable and valid tests that would minimize disadvantages of this kind have been unsuccessful.

INTERPRETING GROUP DIFFERENCES

If group differences in test performance do not result from the simple forms of bias reviewed above, what is responsible for them? The fact is that we do not know. Various explanations have been proposed, but none is generally accepted. It is clear, however, that these differences—whatever their origin—are well within the range of effect sizes that can be produced by environmental factors. The Black/White differential amounts to one standard deviation or less, and we know that environmental factors have recently raised mean test scores in many populations by at least that much.
Intelligence: Knows and Unknowns

(Flynn, 1987; see Section 4). To be sure, the "Flynn effect" is itself poorly understood: it may reflect generational changes in culture, improved nutrition, or other factors as yet unknown. Whatever may be responsible for it, we cannot exclude the possibility that the same factors play a role in contemporary group differences.

Socioeconomic Factors. Several specific environmental/cultural explanations of those differences have been proposed. All of them refer to the general life situation in which contemporary African Americans find themselves. In the African American community, a child's IQ is often determined by his or her social class status. The correlations between IQ and socioeconomic status (SES) and scores on intelligence tests is well-known (White, 1982).

Several considerations suggest that this cannot be the whole explanation. For one thing, the Black/White differential in test scores is not eliminated when groups of individuals are matched for SES (Loehlin et al., 1975). Moreover, the data reviewed in Section 4 suggest that— if we exclude extreme conditions—nutrition and other biological factors that may vary with SES account for relatively little of the variance in such scores. Finally, the relationship between test scores and income is much more complex than a simple SES hypothesis would suggest. The living conditions of children result in part from the accomplishments of their parents: if the skills measured by psychometric tests actually matter for those accomplishments, intelligence is affecting SES rather than the other way around. We do not know the magnitude of these various effects in various populations, but it is clear that no model in which "SES" directly determines "IQ" will do.

A more fundamental difficulty with explanations based on economics alone appears from a different perspective. To imagine that any simple income- and education-based model can adequately describe the situation of African Americans is to ignore important categories of experience. The sense of belonging to a group with a distinctive culture—one that has long been the target of oppression—and the awareness or anticipation of racial discrimination are profound personal experiences, not just aspects of socioeconomic status. Some of these more deeply rooted differences are addressed by other hypotheses, based on caste and culture.

Caste-like Minorities. Most discussions of this issue treat Black/White differences as aspects of a uniquely "American dilemma" (Myrdal, 1944). The fact is, however, that comparably disadvantaged groups exist in many countries: the Maori in New Zealand, scheduled castes ("untouchables") in India, non-European Jews in Israel, the Burakumin in Japan. All these are "caste-like" (Ogbu, 1978) or "involuntary" (Ogbu, 1994) minorities. John Ogbu distinguishes this status from that of "autonomous" minorities who are not politically or economically subordinated (like Amish or Mormons in the United States), and from that of "immigrant" or "voluntary" minorities who initially came to their new homes with positive expectations. Immigrant minorities expect their situations to improve; they tend to compare themselves favorably with peers in the old country, not unfavorably with members of the dominant majority. In contrast, to be born into a caste-like minority is to grow up firmly convinced that one's life will eventually be restricted to a small and poorly-rewarded set of social roles.

Distinctions of caste are not always linked to perceptions of race. In some countries lower and upper caste groups differ by appearance and are assumed to be racially distinct; in others they are not. The social and educational consequences are the same in both cases. All over the world, the children of caste-like minorities do less well in school than upper-caste children and drop out sooner. Where there are data, they have usually been found to have lower test scores as well.

In explaining these findings, Ogbu (1978) argues that the children of caste-like minorities do not have "effort optimism," i.e., the conviction that hard work (especially hard schoolwork) and serious commitment on their part will actually be rewarded. As a result, they ignore or reject the forms of learning that are offered in school. Indeed they may practice a sort of cultural inversion, deliberately rejecting certain behaviors (such as academic achievement or other forms of "acting White") that are seen as characteristic of the dominant group. While the extent to which the attitudes described by Ogbu (1978, 1994) are responsible for African American test scores and school achievement has not been empirically established, it seems that familiar problems can take on quite a different look when they are viewed from an international perspective.

African American Culture. According to Boykin (1986, 1994), there is a fundamental conflict between certain aspects of African American culture on the one hand and the implicit cultural commitments of most American schools on the other. "When children are ordered to do their own work, arrive at their own individual answers, work only with their own materials, they are being sent cultural messages. When children come to believe that getting up and moving about the classroom is inappropriate, they are being
sent powerful cultural messages. When children come
to confine their 'learning' to consistently bracketed
time periods, when they are consistently prompted to
tell what they know and not how they feel, when they
are led to believe that they are completely responsible
for their own success and failure, when they are re-
quired to consistently put forth considerable effort for
their sake on tedious and personally irrelevant tasks
... then they are pervasively having cultural lessons
imposed on them" (Boykin, 1994, p. 125).

In Boykin's view, the combination of constriction
and competition that most American schools demand
of their pupils conflicts with certain themes in the
"deep structure" of African American culture. That
culture includes an emphasis on such aspects of
experience as spirituality, harmony, movement, verb,
affect, expressive individualism, communalism, oral-
ity, and a socially defined time perspective (Boykin,
1986, 1994). While it is not shared by all African
Americans to the same degree, its accessibility and
familiarity give it a profound influence.

The result of this cultural conflict, in Boykin's
view, is that many Black children become alienated
from both the process and the products of the educa-
tion to which they are exposed. One aspect of that
process, now an intrinsic aspect of the culture of most
American schools, is the psychometric enterprise it-
self. He argues (Boykin, 1994) that the successful edu-
cation of African American children will require an
approach that is less concerned with talent sorting
and assessment, more concerned with talent develop-
ment.

One further factor should not be overlooked. Only
a single generation has passed since the Civil Rights
movement opened new doors for African Americans,
and many forms of discrimination are still all too fa-
miliar in their experience today. Hard enough to bear
in its own right, discrimination is also a sharp re-
miniscence of a still more intolerable past. It would be
rash indeed to assume that those experiences, and
that historical legacy, have no impact on intellectual
development.

The Genetic Hypothesis. It is sometimes suggested
that the Black/White differential in psychometric in-
telligence is partly due to genetic differences (Jensen,
1972). There is not much direct evidence on this
point, but what little there is fails to support the ge-
etic hypothesis. One piece of evidence comes from a
study of the children of American soldiers stationed
in Germany after the Second World War (Eyferth,
1961); there was no mean difference between the test
scores of those children whose fathers were White
and those whose fathers were Black. (For a discussion
of possible confounds in this study, see Flynn, 1980.)
Moreover, several studies have used blood-group
methods to estimate the degree of African ancestry of
American Blacks; there were no significant correla-
tions between those estimates and IQ scores (Loehlin,
Vandenberg, & Osborne, 1973; Scarr, Paskins, Katz, &
Barker, 1977).

It is clear (Section 3) that genes make a substantial
contribution to individual differences in intelligence
test scores, at least in the White population. The fact
is, however, that the high heritability of a trait within
a given group has no necessary implications for the
source of a difference between groups (Loehlin et al.,
1975). This is now generally understood (e.g., Herrn-
stein & Murray, 1994). But even though no such im-
planation is necessary, some have argued that a high
value of h^2 makes a genetic contribution to group dif-
fences more plausible. Does it?

That depends on one's assessment of the actual
difference between the two environments. Consider
Lewontin's (1970) well-known example of seeds from
the same genetically variable stock that are planted in
two different fields. If the plants in field X are fertili-
ized appropriately while key nutrients are withheld
from those in field Y, we have produced an entirely
environmental group difference. This example works
(i.e., h^2 is genuinely irrelevant to the differential be-
tween the fields) because the differences between the
effective environments of X and Y are both large and
consistent. Are the environmental and cultural situa-
tions of American Blacks and Whites also substan-
tially and consistently different—different enough to
make this a good analogy? If so, the within-group her-
itability of IQ scores is irrelevant to the issue. Or are
those situations similar enough to suggest that the
analogy is inappropriate, and that one can plausibly
generalize from within-group heritabilities? Thus the
issue ultimately comes down to a personal judgment:
How different are the relevant life experiences of
Whites and Blacks in the United States today? At pre-
sent, this question has no scientific answer.

6. Summary and Conclusions

It is customary to conclude surveys like this one with
a summary of what has been established. Indeed,
much is now known about intelligence. A near-cen-
tury of research, most of it based on psychometric
methods, has produced an impressive body of find-
ings. Although we have tried to do justice to those
findings in this report, it seems appropriate to con-
clude on a different note. In this contentious arena,
our most useful role may be to remind our readers
that many of the critical questions about intelligence
are still unanswered. Here are a few of those questions:

1. Differences in genetic endowment contribute
substantially to individual differences in (psycho-
Intelligence: Knowns and Unknowns

metric intelligence, but the pathway by which genes produce their effects is still unknown. The impact of genetic differences appears to increase with age, but we do not know why.

2. Environmental factors also contribute substantially to the development of intelligence, but we do not clearly understand what those factors are or how they work. Attendance at school is certainly important, for example, but we do not know what aspects of schooling are critical.

3. The role of nutrition in intelligence remains obscure. Severe childhood malnutrition has clear negative effects, but the hypothesis that particular "micronutrients" may affect intelligence in otherwise adequately-fed populations has not yet been convincingly demonstrated.

4. There are significant correlations between measures of information-processing speed and psychometric intelligence, but the overall pattern of these findings yields no easy theoretical interpretation.

5. Mean scores on intelligence tests are rising steadily. They have gone up a full standard deviation in the last 50 years or so, and the rate of gain may be increasing. No one is sure why these gains are happening or what they mean.

6. The differential between the mean intelligence test scores of Blacks and Whites (about one standard deviation, although it may be diminishing) does not result from any obvious biases in test construction and administration, nor does it simply reflect differences in socioeconomic status. Explanations based on factors of caste and culture may be appropriate, but so far have little direct empirical support. There is certainly no such support for a genetic interpretation. At present, no one knows what causes this differential.

7. It is widely agreed that standardized tests do not sample all forms of intelligence. Obvious examples include creativity, wisdom, practical sense, and social sensitivity; there are surely others. Despite the importance of these abilities we know very little about them: how they develop, what factors influence that development, how they are related to more traditional measures.

In a field where so many issues are unresolved and so many questions unanswered, the confident tone that has characterized most of the debate on these topics is clearly out of place. The study of intelligence does not need politicized assertions and recriminations; it needs self-restraint, reflection, and a great deal more research. The questions that remain are socially as well as scientifically important. There is no reason to think them unanswerable, but finding the answers will require a shared and sustained effort as well as the commitment of substantial scientific resources. Just such a commitment is what we strongly recommend.

REFERENCES


PART IV MIDDLE CHILDHOOD


Helms, J. E. (1992). Why there is no study of cultural equiva-

lence in standardized cognitive ability testing? American Psychologist, 47, 1083–1101.


Intelligence: Knowns and Unknowns


---

**QUESTIONS**

1. What is intelligence?
2. Why is stability in intelligence considered an important characteristic of an IQ test score?
3. What do IQ tests predict well? What do they not predict very well?
4. What does research on environmental effects on intelligence indicate about the nature of human intelligence?
5. Why have IQ test scores increased over the twentieth century?
6. How does Lewontin's (1970) example of genetically identical seeds growing up in different fields relate to the debate about ethnic differences in IQ?