Psychometrics, Intelligence, and Public Perception

JOHN B. CARROLL
University of North Carolina at Chapel Hill

Herrnstein and Murray (1994, pp. 22–23) stated six propositions concerning a *g* factor of intelligence. Because these propositions had been widely criticized in public media as being false and pseudoscientific, they are examined here for support in the scientific literature. All are found to be reasonably well supported. Most experts agree that there *is* a general factor *g* on which human beings differ. It is measured to some degree by most tests of cognitive aptitude and achievement, but more accurately by tests designed to measure it. It corresponds to most people's concept of intelligence. It is quite stable over the life span, and properly constructed and administered IQ tests are not demonstrably biased against different social groups. It is substantially influenced by genetic factors, but also by environmental factors. Some psychometric findings about *g* have been poorly presented to the public or widely misunderstood. The public is urged to recognize that (1) psychometrics (literally, mental measurement) is a rigorous scientific discipline that has resolved many questions concerning cognitive abilities; (2) general ability scores should be taken not as direct measures of hereditary intelligence, but rather as measures of rate of progress over the life span in achieving full mental development; (3) there are many other cognitive abilities besides *g*; (4) important sources of variation in *g* or IQ are environmental; (5) the IQ is possibly more an indicator of how fast the individual can learn that it is of the individual's capability of learning; and (6) much more research is needed to resolve questions about the role of individual differences in cognitive abilities in a democratic society. These conclusions can be reached whatever one's views may be about the validity of Herrnstein and Murray's claims about the significance of variation in intelligence for social problems.

The publication of Herrnstein and Murray's *The Bell Curve: Intelligence and Class Structure in American Life* (1994) spawned a veritable cottage industry in which almost numberless reviews, critiques, editorials, and the like were written—but only rarely by informed specialists—to express (mainly) negative views about Herrnstein and Murray's data, analyses, and conclusions. Many of these, as collected by Fraser (1995) and Jacoby and Glauberman (1995), cast doubt on Herrnstein and Murray's emphasis on individual differences in intelligence as a factor in social success and failure, even to the extent of questioning the very concept of intelligence, the instruments used in measuring it, and the methodol-

Direct all correspondence to: John B. Carroll, 409 N. Elliott Road, Chapel Hill, NC 27514-7628.

INTELLIGENCE **24(1)** 25–52
ISSN: 0160-2896

ogy of psychometrics. Gould (1994), for example, claimed that Herrnstein and Murray were mistaken in assuming that intelligence "is depictable as a single number, capable of ranking people in linear order, genetically based, and effectively immutable" (p. 139).

Thus, one result of the appearance of *The Bell Curve* and its scrutiny by critics has been the general circulation among "public intellectuals" of the notion that psychometric research on cognitive ability is a discredited pseudoscience alien to the ideals of a democracy (Giroux & Searls, 1996). However one might regard the methodology and the conclusions of *The Bell Curve*, this result is unfortunate, to the extent that it ignores the real merits of psychological research, or, in particular, psychometric research, in describing and studying variations in the attributes of human individuals that may play important roles in their development and in their successes and failures in life.

Perhaps the methodology of *The Bell Curve* has shortcomings, and perhaps the book draws questionable conclusions about the significance of intellectual abilities in education, occupational success, and various aspects of social life. Those issues are not a focus of this article. Nevertheless, the book's treatment of psychometric research and results is hardly to be characterized as pseudoscientific. Certainly it is unjust and mistaken to claim, merely on the basis of *The Bell Curve*'s use of psychometric research, that all such research is unscientific and not to be taken seriously. It is important to reaffirm the proposition that psychometrics is a sound and fair-minded scientific discipline.

In hoping to do this, I focus on six conclusions about cognitive ability that Herrnstein and Murray stated in the early pages of their book—and that they regarded as being "by now beyond significant technical dispute" in psychometric research (p. 22). As they pointed out, "The received wisdom in the media is roughly 180 degrees opposite from each of the six points" (p. 23), but unfortunately their discussion of the evidence was not sufficient to persuade a general audience. It is useful to consider here the six points more fully from the standpoint of how they have been studied in psychometrics. Thus, each of these propositions is made a major subhead in this article. Each is examined with respect to accuracy and research support and with respect to prospects for further clarification and application of psychometric results.

Sternberg (1995) has already dealt with the issue of whether Herrnstein and Murray's characterization of these propositions as "beyond significant technical dispute" is factually correct. His conclusion, based an examination of many sources in the literature, is that none of them is actually beyond technical dispute, because it is possible to cite statements by various investigators to the effect that one or another of these propositions can be questioned on the basis of relevant theory, research, or analysis. This is perhaps a reflection of the fact that almost nothing in science (at least in the social sciences) is beyond technical dispute.

Nevertheless, scientists do develop consensual views about issues in their fields. It is my goal here to state the current consensus in psychometrics with respect to the initial propositions about cognitive ability that Herrnstein and Murray set forth. Note that these propositions do not touch on some of the more controversial claims that Herrnstein and Murray make, for example, the claim that low intelligence is in part responsible for criminality, or the claim that there are genetically caused racial differences in average intelligence—a claim on which there is certainly no consensus among specialists in psychometrics.

This article is addressed chiefly to the general reader, rather than the specialist. As a consequence, for certain statistical procedures commonly used in psychometrics, explanations are offered that are intended to be simple enough for general readers to follow.

PROPOSITION 1

There is such a thing as a general factor of cognitive ability on which human beings differ. (Herrnstein & Murray, 1994, p. 22)

To discuss this proposition, it is worthwhile to look back at the history of psychometrics. Essentially, the proposition was first introduced by the British psychologist Charles Spearman (1904), though not exactly in those words. Examining small sets of data on test scores and school marks that he had collected in several village schools in the English countryside, Spearman concluded that the correlations among the variables could best be explained by assuming that there was a single factor of mental ability that underlay them. He called this single factor *g*, for "general intelligence," and he developed a mathematical formulation whereby the "saturations" of any two variables with *g* could be multiplied together to predict, approximately, the actual correlation found between them. For example, if the *g* saturation (or as we now call it, the loading) of grades in English was .80, and the *g* saturation of a test of musical pitch discrimination was .67, the correlation between English grade and pitch discrimination was predicted to be $(.80 \times .67) = .536$, very close to the actual observed correlation of .54. In a table showing all the correlations among six variables, Spearman found that the predicted correlations were indeed all close to the actual observed correlations, thus establishing, to his mind, the generality of his formulation.

Spearman's early work formed the basis of a specialty within psychometrics later called *factor analysis*. Over the years since 1904, this specialty has burgeoned enormously, giving rise to dozens of textbooks, hundreds of articles about developments in the technique, and literally thousands of studies of psychological and other kinds of data (even data from chemistry!). The data that Spearman collected and analyzed were later seen to be flawed and inappropriate in certain ways, but the basic idea has remained: that correlations among measured variables (like test scores and ratings) can be shown to be explained by postulating the existence of "factors" or latent sources of variance in the people or objects studied. When correlational data can be well fitted using factor-analytic models, it is scientifically appropriate to accept the existence and functioning of the postulated factors. Factors are not "things," as alleged by Gould (1981, 1994; see also Carroll, 1995), but at least some of them may well correspond to characteristics of individuals that predict their behavior to a substantial extent, in the sense that they describe how well individuals can learn, remember, and perform in a great variety of real-life situations beyond the situations in which individuals are tested.

As applied to data using measures of cognitive abilities and achievements, factor-analytic research proceeds on the basis of a mathematical model that states that a test score (or other observed variable) is a function of the one or more latent abilities that contribute to performance on the test. Factor analysis attempts to identify these abilities, describe their relationships, and show how they contribute to test scores. The basic computational problem of a factor-analytic study is to convert data on the correlations of a series of variables to an appropriate model that will account for those correlations in terms of one or more factors.

Here I give a simple example (much simpler, for the purposes of this exposition, than what is usually studied in actual research). Also, for the sake of simplicity I ignore several technical problems that arise in this example but that need not cause concern.

Suppose that we have given six tests to a large group of people sampled from the general population. Test 1 is a test of reading comprehension, Test 2 is a measure of vocabulary knowledge, Test 3 is a measure of listening comprehension, Test 4 is a formboard test requiring people to see how geometric forms fit together, Test 5 is one requiring people to compare letters of the alphabet rotated to different positions, and Test 6 is one requiring people to predict how a piece of paper folded several times and punched with a hole will look when it has been unfolded. Suppose further that the correlations among the six tests are found to be as in Table 1 (given in what is called a correlation matrix):

The correlation matrix in Table 1 would be factor analyzed to produce what is called a factor matrix. The goal of such an analysis is to determine how many factors would be needed to explain the correlations and then to specify the

TABLE 1
Correlation Matrix for Six Tests

Test		1	2	3	4	5	6
Reading comprehension	1	1.00	.72	.59	.42	.24	.30
Vocabulary	2	.72	1.00	.50	.35	.20	.25
Listening comprehension	3	.59	.50	1.00	.28	.16	.20
Paper formboard	4	.42	.35	.28	1.00	.58	.65
Letter rotation	5	.24	.20	.16	.58	1.00	.56
Paper folding	6	.30	.25	.20	.65	.56	1.00

TABLE 2
Final Estimated Factor Matrix

	Factor		
	I	II	III
Test 1—Reading comprehension	.62	.69	.01
Test 2—Vocabulary	.52	.58	.00
Test 3—Listening comprehension	.42	.48	.00
Test 4—Paper formboard	.59	.08	.59
Test 5—Letter rotation	.44	−.05	.55
Test 6—Paper folding	.51	−.03	.60

simplest set of loadings of the tests on these factors to best fit or predict the data in the correlation matrix. For this particular correlation matrix, factor analysis would find that only two uncorrelated factors would be needed to account for the correlations to a sufficient degree of accuracy, but it would also find that these two uncorrelated factors should optimally be "rotated" in the factorial space in such a way that these factors themselves would be correlated. The resulting correlated factors could then produce a third factor (labeled Factor I in the matrix that is Table 2). Finally, the three factors would be further transformed so that none of them are correlated. (Readers should not be concerned if they fail to understand these procedures completely. The procedures depend on fairly complex mathematical models that need not be explained here.) The final estimated factor matrix would be as shown in Table 2.

The factor matrix specifies the weights or loadings of each test on each factor. If the factor matrix has been correctly estimated, these loadings can be used to see how well the factor matrix accounts for the correlations. That is, each correlation between two different tests should be predictable by summing the products of the corresponding weights for the corresponding tests. For example, readers can verify for themselves that the correlation between Test 1 and Test 2 is predicted to be $(.62 \times .52) + (.69 \times .58) + (.01 \times .00) = .7226$, close to .72 as shown in the correlation matrix (Table 2). Similarly for other pairs of tests. In general, the correlations are well predicted by the loadings in the factor matrix. Note, however, that if one tries to use the factor matrix to estimate the correlation between a test and itself (which must actually be 1.00), the estimates will in general not be 1.00. Instead, they will be values that are called *communalities*; these are measures of the degree to which scores on a given test can be accounted for by the underlying common factors. For the six tests in the example, the communality values are .85, .61, .41, .74, .52, and .61, respectively. What is left over from this (for Test 1: $1.00 - .85 = .15$) is a measure of test *uniqueness*, that is, the variance contributed by anything that is uniquely measured by that test, including error of measurement.

Some critics of factor analysis have complained that the factor matrix produced using standard factor-analytic procedures is not the only factor matrix that can reproduce or predict the correlation matrix. Indeed, there is an infinite number of such matrices. However, standard factor-analytic procedures now include a requirement of what Thurstone (1938, 1947) called *simple structure*, which may be roughly described as a requirement that the factor matrix include a maximal number of near-zero loadings. In the factor matrix in Table 2, the effect of this requirement can be seen in the fact that both Factors II and III include three zero or near-zero loadings.

Note that the factors are, up to this point, labeled arbitrarily with the Roman numerals I, II, and III. The next problem would be to "interpret" the factors. This is done by inspecting the matrix, bearing in mind the characteristics of the tests. It appears that Factor I is measured to some extent by all six tests, with loadings ranging from .42 to .62. It might well be interpreted as a "general intelligence" or *g* factor. Factor II has high loadings only for Tests 1, 2, and 3, and near-zero loadings for Tests 4, 5, and 6. What is measured uniquely by Tests 1, 2, and 3, and not by the remaining tests? Because Tests 1 to 3 involve knowledge of language, it is reasonable to interpret Factor II as a factor of "verbal ability" that could be labeled *V*. Factor III appears to pertain only to what is measured in common among Tests 4, 5, and 6. It seems that because these tests measure, in part, an ability to deal with spatial forms that can be rotated in two dimensions, we may conclude that Factor III could be labeled as a "spatial ability" factor, *S*.

My purpose here has been to give the reader a concrete idea of how factor analysis works and how its results are achieved and interpreted. In actuality, factorial studies would hardly ever be limited to such a simple set of data. To adequately sample several domains or subdomains of cognitive ability, studies usually involve much larger sets of data—involving anywhere from, say, 10 variables up to 100 or even more. Requirements for satisfactory analysis include the need for each factor to be represented by at least three variables. The number of variables studied is limited chiefly by the number of tests or subtests that can be feasibly administered to groups of participants—whether they are volunteers or members of "captive groups" (as in the military). An important aspect of the design of factorial studies is the selection of tests; there must be common elements in groups of tests, preferably varied in format and content to test hypotheses about the nature of the factors. Other aspects of the design of factorial studies concern the selection of participant samples to be tested and details about how the tests are administered—for example, whether they are given with a time limit. Finally, because of errors of measurement and sampling, factorial studies produce factor matrices that do not precisely reproduce the correlation matrices from which they are derived, but the fits between correlations produced by the factor matrices and the actual data are generally close, and it is possible to measure the significance and goodness of fit.

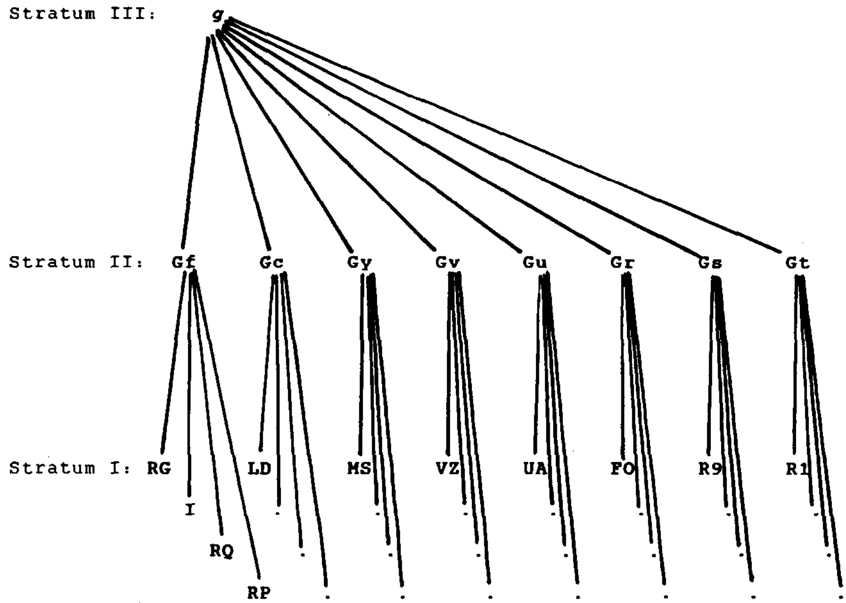

Figure 1. Schematic structure of the domain of cognitive abilities as postulated by Carroll's (1993) three-stratum theory. The two-character symbols at Stratum II represent broad abilities. Only part of the structure at Stratum I is spelled out; the one- or two-character symbols and periods represent narrow, Stratum I factors.

Whereas Spearman, in much of his work (Spearman, 1927), in effect postulated that only one factor, *g*, was needed to account for correlations among cognitive ability tests, it is now known that many more factors are often needed. In a survey and reanalysis of more than 460 sets of data accumulated over the past six or seven decades (Carroll, 1993), I concluded that the total domain of cognitive abilities may be described in terms of what I called a three-stratum model. This model is illustrated here in Figure 1. At Stratum I there are a fairly large number (around 60 are now known) of "narrow" abilities, like native-language vocabulary knowledge, ability in performing basic arithmetical operations, or ability to discriminate musical pitches. At a second stratum, there may exist about 10 "broad" abilities, such as *Gf* (fluid intelligence) and *Gc* (crystallized intelligence), as proposed by Cattell (1971) and others. Finally, at the apex of the structure, Stratum III consists of only a single factor, *g*, much like what Spearman originally proposed. It is "general" in the sense that it is likely to be present, in some degree, in nearly all measures of cognitive ability. Furthermore, it is an important factor, because on the average over many studies of cognitive ability tests it is found to constitute more than half of the total common factor variance in a test.

Let us remind ourselves again that the factorial model assumes that a test score (or other variable) is a function of one or more different uncorrelated factors. In the simple example given earlier, it was found that scores on Tests 1, 2, and 3 were functions of two factors—a general factor and a verbal factor, and scores on Tests 4, 5, and 6 were functions of this same general factor and a spatial ability factor. In the process of doing the factor analysis, the factors were transformed to be uncorrelated. Many investigators prefer to base interpretations on factors made to be uncorrelated, even though it is normally impossible to calculate, from actual data on test scores, factor scores that are uncorrelated. Conceptually, orthogonality (noncorrelation) of factors makes sense in interpreting a person's profile on a series of factors. For example, a person could be relatively high on a general factor while being relatively low on a particular narrow factor like vocabulary knowledge. The general factor can be regarded as being associated with the covariance between tests that is left over after the influence of lower stratum factors has been statistically controlled.

Early in the history of factor analysis—around 1935 to 1945—there was a controversy about the existence of *g*. The American psychologist Thurstone (1938) claimed to find no *g* in a large set of data that he analyzed, but Spearman (1939) and one of his colleagues (Eysenck, 1939) reanalyzed Thurstone's data and found a general factor. The controversy arose because of certain technical problems in Thurstone's then newly developed procedures for performing factor analysis studies. These technical problems were eventually resolved, and Thurstone (1947) himself acknowledged that a general factor *g* could be found in the correlations of "primary" or first-order factors. One of these technical problems—the rotation of factors to achieve simple structure—has now been reasonably well solved, and it is no longer a major obstacle to the analysis of factorial data.

At the present time the evidence for a general factor can be said to be overwhelming. In my reanalyses (Carroll, 1993) of factorial studies done since about 1925 up to the 1990s, some kind of a general factor was found in nearly all of these studies. The nature of a general factor can vary somewhat with the sample of tests included in a test battery. For example, if nearly all the tests in a battery that is analyzed are paper-and-pencil tests with verbal instructions, the general factor may be biased toward incorporating verbal ability. The nature and extent of the general factor can also be affected by the kind of participant population given the tests. For example, there is less variability in the general factor if the participant population is relatively homogeneous in educational level, as opposed to a participant population that is widely heterogeneous in educational level—even when the influence of age is statistically controlled.

It must be pointed out that some factor analysts are dubious about the existence of a general factor. Partly this is because we still do not have an adequate data base from which we might identify and characterize a truly "general" factor. An adequate data base would be one in which there would be adequate measurements of the total diversity of cognitive abilities and in which there would be adequate sampling of individuals from the general population with respect to age, social class, amount of education, and other demographic characteristics that might correlate with cognitive abilities. The debate about the general factor is also the result of differences in specialists' definitions of the formal requirements for identifying a general factor. Horn and Noll (1994), for example, concluded that no truly general factor that meets Spearman's original requirements has yet been found. Nevertheless, they supported the existence of a broad factor of ability they called fluid intelligence (*Gf*), and such a factor could form the basis of the general factor cited by Herrnstein and Murray. Some investigators (Gustafsson, 1984) believe that *Gf* is identical to the general factor *g*. The consensus of most investigators is that some kind of general factor of cognitive ability exists and that it can be estimated satisfactorily from currently available measurements. In this sense Herrnstein and Murray's Proposition 1 can be taken as valid.

PROPOSITION 2

> *All standardized tests of academic aptitude or achievement measure th[e] general factor to some degree, but IQ tests expressly designed for that purpose measure it most accurately.* (Herrnstein & Murray, 1994, p. 22)

Several years ago (Carroll, 1987), I examined data on reading skill attainment in the United States from the National Assessment of Educational Progress (NAEP) (1985) and from a study (Kirsch & Jungeblut, 1986) of a representative national sample of young adults. I introduce my findings here because they are pertinent to considering whether the data on reading skill attainment also reflect the influence of a general factor of intelligence.

Figure 2 shows levels of reading skill attainment for selected percentiles of the relevant sampled participants at ages 9, 13, 17, and 23 on a Reading Proficiency Scale constructed by Educational Testing Service that specifies levels as rudimentary, basic, intermediate, adept, and advanced. (See the NAEP report for fairly satisfactory definitions of the levels rudimentary, basic, intermediate, adept, and advanced in terms of typical test items at each of those levels.) I regarded this plot as "of signal, almost historic importance because it shows for the first time a meaningful overall view of the status and variability of literacy in the U.S.—at least for schoolchildren and young adults." To quote further from my report:

> What we see is a gradual development of reading skill, from age 9 through about age 23 (the young adult sample ranged from age 21 to age 25). At age 9, the median performance is a little beyond the Basic level; at age 13, it is a little beyond Intermediate; at age 17, it is almost up to Adept; and at age 23, it is somewhat better than Adept.
>
> But what is even more striking is the tremendous variability in reading skill at all ages. At age 9, the top 5% of children are already reading at a level as high as that of the 25th percentile of the adult population. At age 17, the bottom 5% are still reading at levels no

better than those attained by about 40% of the 9-year-olds. Even greater variability is evident in the young adult group, in which the middle 99% of the population ranges from just about the Rudimentary level to far beyond the Advanced level, and the middle 50% ranges from just above Intermediate to a little below Advanced.

It appears that reading comprehension ability is, on the average, rather slow to develop. Under the present conditions of education in the U.S., it takes the *average* person about 14 years to progress from a skill level just above Basic (at age 9) to a level just above Adept (at age 23). But once again the variation is enormous: about 1% of 9-year-olds have already reached the Adept level; about 44% of young adults have not yet attained that level. (Carroll, 1987, p. 426)

In many respects these data are analogous to what we might expect if we made a similar plot for mental ages on standard intelligence tests such as the Stanford-Binet (to be discussed). The data might even lead us to suspect that the Reading Proficiency Scale actually measures intelligence or the general factor *g*. Let us consider to what extent this might be the case. For various reasons the NAEP Reading Proficiency Scale has never appeared as a variable in a broad factor-analytic study, one reason being that the scale does not give scores for individuals, but only for groups. But the format and content of the reading skill exercises on which the scale is based are very similar to what is found in standardized tests of reading comprehension.

Let us assume that Test 1, in the small example of a factor analysis presented earlier, is such a standardized test of reading comprehension. In the factor analy-

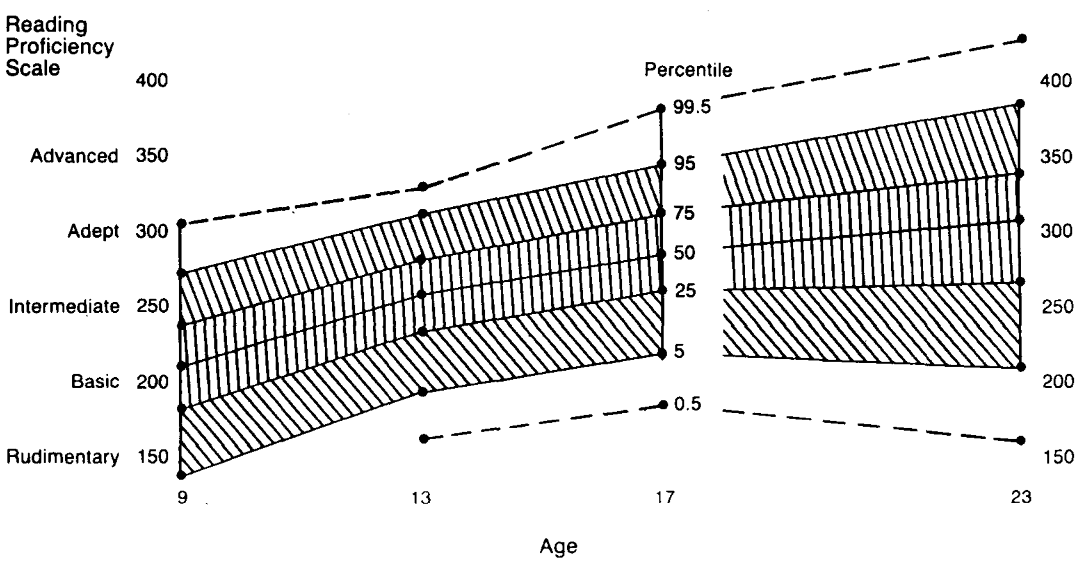

Figure 2. Trends in reading proficiency for schoolchildren aged 9, 13, and 17 tested in the National Assessment of Educational Progress (1984), and a sample of young adults (1985), at selected percentiles. (Adapted by permission from Carroll, 1987.)

sis, we found that Test 1 had a loading of .62 on a general factor, and a loading of .69 on a factor of Verbal Ability. Such loadings would be typical of those found in actual studies. Therefore, we could conclude that the Reading Proficiency Scale measures the general factor to a considerable extent, but that it is also strongly influenced by the individual's reading ability (holding the *g* factor constant). Indeed, in the report of my analysis of the NAEP and Young Adult data (Carroll, 1987), I suggested that these data tended to indicate the status of verbal intelligence in the U.S. population, at least over the ages represented in the samples studied.

Recently, educational policy specialists Ralph, Keller, and Crouse (1994) have gone even further in suggesting that achievement growth in reading, mathematics, and science, as indicated by NAEP results in those subjects, may possibly be constrained by the growth and distribution of general intelligence and by the failure of school practices to boost general intelligence. Such a conclusion may go too far, in my opinion, in that reading, science, and mathematics skills can clearly be improved by education and training, but it is not clear that general intelligence can similarly be trained. An important research hypothesis that remains to be thoroughly investigated is, however, that the degree to which reading, math, and science skills can be improved is dependent on general intelligence levels.

Obviously, a reading proficiency test, even when very well constructed according to all the criteria of excellence developed in psychometrics, is an imperfect measure of *g*, even though influenced by it.

Now, suppose we are interested in finding a better, purer measure of *g*. For this we might turn to tests that are specifically designed to measure *g*. One such test could be one of Raven's (1938–65) *progressive matrices* tests, developed under the supervision of Spearman himself. Items in these tests show 3×3 tables with geometric designs that change in systematic ways over the rows and the columns of the table. The bottom right space in each such table is blank; the participant is asked to figure out which one of eight possible alternative designs belongs in that space, given the rules (which the participant must discover) that govern how the designs change over the rows and columns. Carpenter, Just, and Shell (1990) analyzed the cognitive processing involved in this test and concluded that it measures primarily the ability to induce abstract relations and the ability to manage a large set of problem-solving goals in working memory. However, my analysis of the status of the test in factor-analytic investigations indicates that it is not necessarily the pure test of *g* that some authors (e.g., Jensen, 1980, p. 646) think it is; scores tend to be influenced by spatial abilities, and loadings on a General Ability factor are not always as high as one would desire for a test of *g*.

The *g* Factor in the Stanford-Binet

Thus, I turn to the widely known Stanford-Binet Intelligence Scale (Terman & Merrill, 1960; Thorndike, Hagen, & Sattler, 1985) as an example of a test of general ability. In point of strict historical accuracy, it was not designed as a test of the general factor *g*; the concept of *g* as a "factor" had not developed sufficiently at the time of the original creation of the Stanford-Binet series (Terman, 1916) to allow one to make such an assertion. Nevertheless, its format and content justify the belief that it measures general ability, and its IQ scores generally have high loadings on *g* in factorial studies (see Carroll, 1993, p. 701, for details). The Stanford-Binet does not require any degree of reading proficiency on the part of the examinee. All questions are presented orally by a person trained to do so, and it is administered individually, one to one, rather than as a test that can be administered to a group.

The Stanford-Binet Intelligence Scale illustrates the use of a particular technology that was developed in psychometrics, originally by the French psychologist Alfred Binet. The key elements in this technology were (a) the creation of a series of "mental tasks" and (b) the age grading of these tasks. Terman (1916), following Binet's lead, developed or adapted a series of mental tasks that were graded in terms of the average ages at which children could begin to perform them successfully. Criteria were established for deciding whether a task was "mental," in order to eliminate tasks that represented mere physical maturation, but there were no criteria for distinguishing between, for example, verbal, spatial, memory, and reasoning tasks. As a result, intelligence test scores reflected average progress over different abilities, some of which were not especially well correlated with each other. In administering these tests, the problem was to estimate the "mental age" of a child, in terms of the age assignments of the tasks, without reference to the child's chronological age. For example, if a child could pass all the tasks up to those assigned to age 8, but tended to fail on tasks beyond that age, the child was given a mental age of 8. This score was then evaluated by considering the child's chronological age. If the child was only age 6, the child was regarded as being well advanced in mental development, whereas if the child was 10 years old, the child was regarded as being delayed in mental development. An intelligence quotient (IQ) was often used in this evaluation, obtained by dividing the mental age by the chronological age and multiplying the result by 100. (Finding IQs in this way has generally been superseded by various other procedures to evaluate mental ages, such as the use of "derived scores," but the principle has remained the same—to evaluate the mental age with reference to the mental development of "average" children.)

These procedures imply a kind of primitive theory of mental development, one that says that (a) different aspects of mental development proceed in parallel, and (b) there is a norm of mental development represented by the performances of average children.

This primitive theory says nothing, really, about how mental development proceeds, except that on the average, mental ability increases. It does not say whether mental development has genetic or environmental causes, or some combination of such causes; and it gives only a crassly empirical basis for specifying a norm of development. As a result, the way in which IQ tests such as the Stanford-Binet are created provides little basis for defining what the tests measure or for specifying the standards of mental development in any directly interpretable way. Operationally, one can only appeal to the character of the tasks contained in the tests and the norms provided. Despite the somewhat primitive quality of the developmental theory underlying these general intelligence tests, the tests have been widely accepted because of the excellent psychometric attributes of their scores (high reliability and stability coefficients and the strong correlations with school success and many other variables). Clinicians and school psychologists find them very useful (Kaufman, 1994; Reynolds, 1994). Nevertheless, the procedures in selecting items for these tests are open to debate. For example, some have argued that IQ tests are unduly loaded with tasks that are taught in schools.

The Stanford-Binet Intelligence Scale thus comprises two types of measures: (1) a mental age (M.A.) scale and (2) an IQ scale. Both are based on the same data—that is, the responses of the participants to the tasks presented and the questions asked. Mental age scores indicate the *absolute* amount of progress achieved, *by whatever means* (genetic or environmental), in attaining the mental skills and knowledges that can be expected to be attained by individuals who have had adequate exposure to, and have been able to exploit, the total range of information provided in an advanced culture. Obviously, young children receive relatively low M.A. scores, partly because they have not lived in the culture for very long. Older children can and do receive much higher mental ages, but there is wide variation in those scores. Some older children receive M.A. scores considerably above their chronological age; others may receive scores substantially below their chronological age. IQ scores indicate the *relative* amount of progress made by a child as compared with that of average children.

We can depict the theoretical development of intelligence, as measured by IQ tests like the Stanford-Binet, by plotting the expected values of mental ages for children with given chronological ages and values of IQ, according to an equation for mental growth proposed by Sagiv (1979), as I have adapted it for this purpose; Figure 3 presents such a plot. The curve for a child with IQ = 100 represents the expected course of mental age growth over chronological ages 3 to 20 for the average child in a culture such as ours, whereas the other curves are for children with IQs higher or lower than 100. The ordinate of the plot represents mental age in two ways: (first, at the left) in terms of an equal-unit scale developed by Thurstone and Ackerson (1929) and (second, at the right) in terms of mental age as commonly determined from IQ tests. Note that these curves are only *theoretical*; actually, progress in mental age for an individual child tends to fluctuate somewhat, presumably because of variations in the extent to which, for whatever reasons, the child is able to learn and profit from the environment at any given time (Moffitt, Caspi, Harkness, & Silva, 1993; Pinneau, 1961). Also, there is a

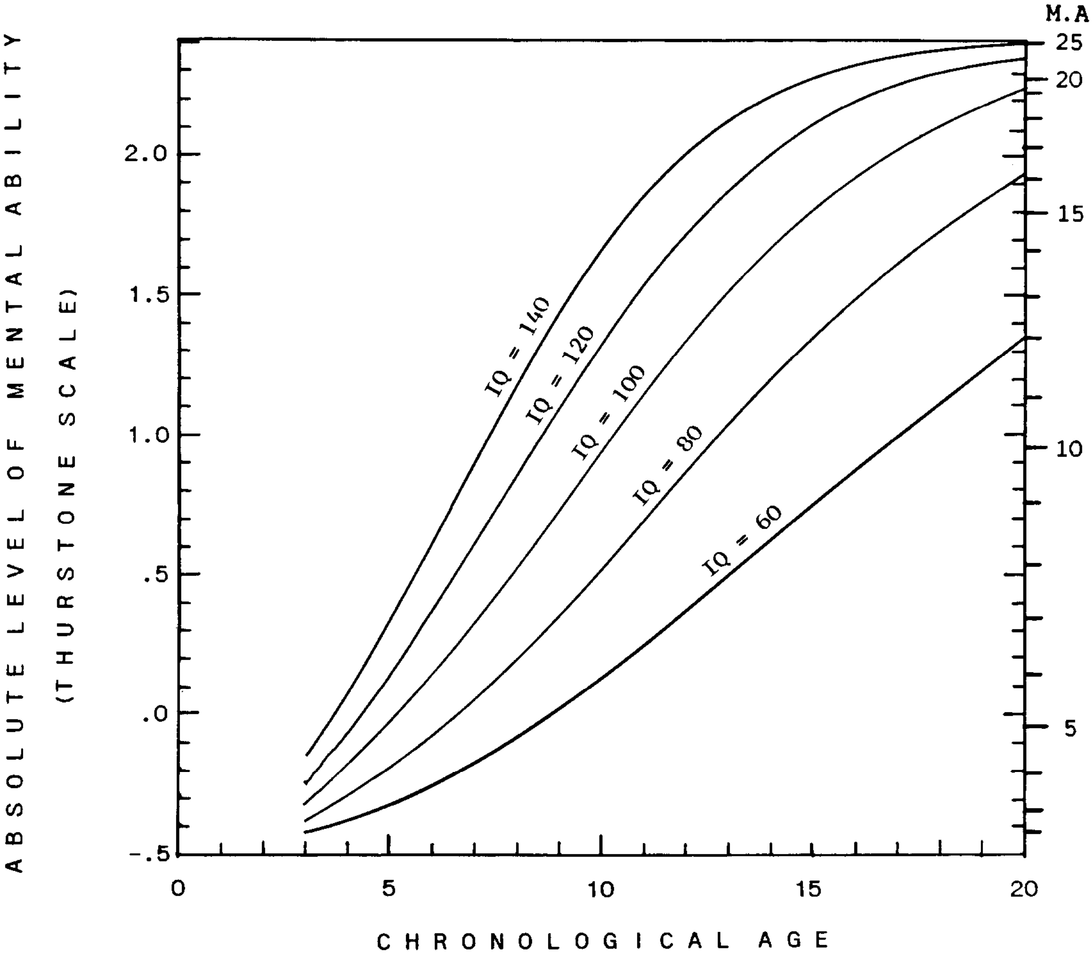

Figure 3. Theoretical curves showing expected trends in mental age scores for children with IQs of 60, 80, 100, 120, and 140.

problem in using Sagiv's formulation for different IQs, in that the curves may not properly suggest what levels of ability can be attained at chronological ages after about 16 for persons with IQs substantially different from 100.

What is important about Figure 3 is that it implies that the IQ can be interpreted as a measure of the *rate* at which the child's mental age improves as a result of exposure to the culture and learning from it. The IQ score expresses the *relative* degree of progress attained in comparison with the progress achieved by the typical or modal ("average") individual in the culture. Because, over historical time, the average levels of progress attained by individuals in a culture can change—upwards or downwards, but usually upwards—the average IQs can change as well, as documented, for example, by Flynn (1987).

The basis for identifying a *g* factor in scores on the Stanford-Binet is that factor-analytic studies have shown that *g* is present to some degree in any kind of mental task contained in the test and that the *g* factor can therefore be estimated by averaging over all such tasks, even if they measure different abilities that may not be particularly well correlated among themselves.

The *g* Factor in the ASVAB

A slightly different technology has been used in the construction of mental ability tests for adults, that is, persons for whom the concept of an increasing mental age is inapplicable. Let us consider this technology in the case of the Armed Services Vocational Aptitude Test (ASVAB), data from which were extensively used by Herrnstein and Murray (1994) in considering correlations with such demographic and social phenomena as social class, poverty, and crime. This technology depends on grading tasks not by age but by the proportions of adults (in representative samples of defined populations) successfully performing them. Tasks that are passed by most people are regarded as "easy," and tasks that are performed successfully by few people are regarded as "difficult," even though in actuality some "difficult" tasks may be easy, in some sense, but for various reasons are learned by only relatively few persons. The theoretical model underlying this technology is similar to that for the Stanford-Binet—one that assumes that a series of tasks can measure a single ability—or at least a composite of different abilities that contains an underlying single ability.

As described by Foley and Rucker (1989), potential tasks for 10 subtests of the ASVAB were selected from banks of items that have been found predictive of successful performance in a variety of occupations or military specialities. They were pretested and "item analyzed" to determine their "difficulty" and the extent to which each item correlated with a total score. Procedures developed in item response theory (Lord & Novick, 1968) were used to establish subtests with desired properties such as high reliabilities and appropriate score distributions. The content of the subtests is indicated fairly well by their names: General Information, Numerical Operations, Attention to Detail, Word Knowledge, Arithmetic Reasoning, Space Perception, Mathematics Knowledge, Electronics Information, Mechanical Comprehension, General Science, Shop Information, and Automotive Information.

As may be seen, the subtests measure knowledges and "aptitudes" in a variety of domains; the subtest scores are correlated in such a way as to indicate that at least the following special abilities are measured: V (verbal ability), RQ (quantitative reasoning ability), N (numerical facility), MK (mechanical knowledge), KM (knowledge of mathematics), and G0 (general information, including scientific information). In addition, all subtests measure, in various degrees, a general factor *g*, in that the correlations of their scores are structured in such a way as to suggest that a general factor is present. The measure of a *g* factor most used by

Herrnstein and Murray is the standardized total score on a subset of the ASVAB subtests—subtests that had been found to be most highly *g* loaded. This total score is referred to as being from the AFQT (Armed Forces Qualifying Test).

There is one special consideration in the case of the ASVAB as a whole, and the AFQT score derived from a subset of the ASVAB tests: the ASVAB is a paper-and-pencil test battery that can be administered to groups. Most of the subtests contain much printed verbal material, and, therefore require, on the part of the examinee, proficiency in reading, to perform the tasks successfully. All the subtests contributing to the AFQT require reading proficiency. Data presented above (Figure 2), on variations in reading proficiency, show that there are wide variations in reading skills in the U.S. population (on which the ASVAB was normed). It is almost undoubtedly the case that deficits in reading proficiency (for whatever reasons they occur) are reflected in low *g* scores from the ASVAB, regardless of the intelligence of individuals as it might be measured by the Stanford-Binet IQ tests, which does not require reading proficiency. The ASVAB was designed to include measures of reading proficiency because military psychologists had found that reading proficiency was important for success in many military occupational specialties (Sticht, 1992). Thus, Herrnstein and Murray's findings of correlations between *g* (as measured by the AFQT) and various social indicators need to be qualified to some extent, in that low-scoring individuals are not *necessarily* low in general mental ability; they may be low mainly in reading proficiency (for various reason, e.g., lack of exposure to good reading instruction, or some kind of special reading disability, or dyslexia). On the other hand, in the population as a whole, general mental ability is substantially correlated with reading proficiency.

It is fair to say that a current research problem is how best to determine and measure the general factor of intelligence; it is not equally well measured using various standardized tests.

As Herrnstein and Murray have tried to demonstrate, the distribution of the general factor of intelligence in the total U.S. population approximately follows the bell-shaped curve, the so-called normal distribution formulated by statisticians. From the standpoint of a democracy dedicated to the creed that "all men are created equal" in rights before the law, the conclusion to be drawn from this distribution of *g* is that men (or women) are far from equal to each other in cognitive ability. This is not the place to philosophize about the implications of this conclusion; I can only call attention to the fact that it imposes major problems for our society—problems that psychometrics may have some role in solving. Some critics have complained that the distribution of intelligence scores does not truly conform to the bell-shaped curve as defined by statisticians (the "Gaussian" distribution), but psychometricians would generally regard this as an unimportant issue. The important point is that the distribution of intelligence is *roughly* "bell-shaped," with diminishing frequencies as one departs from the middle or average.

PROPOSITION 3

IQ scores match, to a first degree, whatever it is that people mean when they use the word intelligent *or* smart *in ordinary language.* (Herrnstein & Murray, 1994, p. 22)

This proposition speaks to the meaning of *g*, or IQ (as the term is loosely used by Herrnstein and Murray). There has been relatively little research on what the term *intelligence* means in ordinary language used by people in everyday life. A study by Sternberg, Conway, Ketron, and Bernstein (1981) is relevant here, however. Sternberg and his colleagues collected their data by getting people at various sites (a commuter train station, a supermarket, and a college library) to list behaviors characteristic of "intelligent," "academically intelligent," and "everyday intelligent" persons. Behaviors rated as important by experts in the measurement of intelligence were factor analyzed and interpreted as being associated with three types of intelligence: verbal, problem solving, and "practical." Apparently, the first two types are fairly well tested using standard intelligence tests, but the consensus is that "practical" intelligence is somewhat different from what is measured by such tests. There is ongoing research on how "practical intelligence" (the ability to size up everyday situations or to display interest in the world around one) could be measured (Sternberg & Wagner, 1986; Sternberg, Wagner, Williams, & Horvath, 1995).

In the meantime, specialists in intelligence measurement have long speculated on the meaning of intelligence as measured using standard tests. More than 70 years ago, a symposium (Thorndike et al., 1921) attempted to yield acceptable definitions, but if anything, there was more disagreement than agreement. A more recent symposium (Sternberg & Detterman, 1986) came far from reaching a satisfactory consensus. In a volume on theories of intelligence, edited by Detterman (1994a), most of the contributors agreed on the existence of *g*, although they characterized it in somewhat different ways. For example, Humphreys (1994) conceived of *g* as the total intellectual repertoire of behavioral responses that an individual has attained at any given point of time. Jensen (1994) stated his belief that *g*, as identified in factorial studies, reflects "some general property or quality . . . of the brain" (p. 268), and Eysenck (1994) attempted to give *g* a biological interpretation by referring to its substantial correlations with a number of reaction-time and physiological measures. Detterman (1994b) was concerned principally with whether *g* is a single factor; in his view the evidence suggests that whereas *g* may appear as a single factor in factorial studies, it actually represents many different information-processing abilities. Horn and Noll (1994) were concerned with what types of psychological tests conform to a strict Spearman model in which there is one and only one general factor; in their opinion only certain tests of what they call fluid intelligence (*Gf*) do so, in contrast to tests of other broad intellectual abilities, which apparently do not. Ceci and Bruck (1994) agreed that there is evidence for a general factor of intelligence, but they urged that it be reinterpreted in the light of what they called a bio-ecological theory that takes cognizance of developmental and contextual perspectives. Only Krechevsky and Gardner (1994), in presenting the latter's theory of multiple intelligences (Gardner, 1983), appeared to deny or lay aside the existence of *g*, insisting that this concept need not be taken into account in planning educational curricula.

At this time, therefore, specialists studying different manifestations of intelligence do not present anything like a united front on the meaning of the general factor. Still, I would contend that there are many common elements in their approaches, and the notion of general ability—whether it is called *g*, or whatever—is one of these. Even Krechevsky and Gardner may not be able to deny that there might be a *g* factor present in correlations among measures of their seven kinds of intelligence; adequate data on this point are apparently not yet available.

In my opinion, however, from the earliest days of the discipline up to the present time, psychometricians have failed to take a sufficiently broad perspective on the way in which progress in mental development occurs, the types of forces that encourage or inhibit that development, and the social context in which it takes place. The norms of mental development intrinsic to the mental age and the intelligence quotient result from the fact that *on the average*, as it were, children are brought up and taught their native language in "typical" families, go to school from an early age, and are exposed to "typical" curricula that teach the "three Rs," and later get into social studies, science, and the rest—all promoting mastery of knowledges and skills that should permit them to perform the tasks posed at different ages on intelligence tests. This occurs in advanced societies—in the United States and elsewhere—which show broad similarities in family structures and systems of education. It is difficult to deny that passing items on intelligence tests depends on *learned* skills—skills learned in the daily life of families and in school. Hart and Risley (1995), for example, showed how preschool children aged 2 to 4 in the United States acquire their oral vocabularies from their parents; the sizes of the vocabularies they acquire depend largely on the kinds of talk their parents use with them. Size of vocabulary is certainly one aspect of intelligence (Olson, 1986). The norms of mental development implicit in the IQ, as it is usually measured, might in fact be somewhat different in a culture in which family groups and schooling are structured differently from those in our own culture (Cole & Means, 1981). But psychometricians often overlook this fact, seeming to claim that intelligence develops independently of family structure or schooling. Note that Ceci (1991) found it necessary to question the traditional assumption that, as he put it, "IQ influences how many years of school one completes but is itself uninfluenced by how many years of schooling one has already completed." Ceci's review tends to support the "alternative possibility" that "schooling exerts a substantial influence on IQ formation and maintenance" (p. 703), a conclusion that seems only too obvious.

Psychometricians have emphasized the finding that, under schooling that is roughly comparable in quantity and quality for all children up to the minimal school-leaving age in typical advanced societies, there are large variations in actual attainment, as we have already seen. The big question has always been: What are the sources of these variations? Are they more in heredity, or in the environment, that is, in how children are treated or exposed to opportunities to learn? I cannot deal with this question here, though it comes up in connection with Proposition 6, to be discussed briefly. First I want to mention a problem about what we might mean by "ability to learn."

IQ as Learning Ability Versus Rate of Learning

Attempts to define intelligence have sometimes included the notion that IQ is related to "ability to learn." For example, in the 1921 symposium on intelligence and its measurement (Thorndike et al., 1921), this idea was mentioned by several of the contributors—S.S. Colvin, W.F. Dearborn, and H. Woodrow. However, Woodrow (1917a, 1917b) had performed several studies that suggested that it was at least difficult to find any significant correlation between IQ and ability to learn simple tasks, and the idea of such a correlation was in effect dropped from further study for many years (Woodrow, 1946). In the symposium that attempted to update the 1921 symposium (Sternberg & Detterman, 1986), the only essays that specifically considered this idea again were those by Brown and Campione (1986) and by Butterfield (1986). Furthermore, only the first of these essays reported any results to support a relation between IQ and learning—in particular, for tasks that involved understanding and "principled transfer" of knowledge rather than the rote associative learning that had been studied by Woodrow.

But what, actually, is "ability to learn"? If there is a correlation between IQ and learning ability, does this mean that individuals with low IQs are *unable* to learn to perform tasks that appear to require higher levels of mental skills? This, indeed, seems to have been the interpretation to which many educators have been inclined. It underlay the use of mental tests in the military, both in World War I and World War II, to screen out individuals believed unable to learn the skills required in various military specialties, and it has similarly been the basis for the use of mental tests in schools to classify students into different "tracks" of curricular content and difficulty.

I believe this interpretation has been a major error throughout the history of psychometrics—at least as understood by people who use IQ tests. If there is any connection between IQ and learning, it is more likely to be with the *rate with which learning occurs* or *the time required for learning*. Such a conclusion seems obvious when one considers the meaning of IQ as it emerges from our analysis of the relation between IQ and mental growth. Essentially, the evidence tends to show that people with high IQ scores are likely to learn more, and remember more, than people with low IQs, and that moreover they are able to learn things faster than people with lower IQs. Up to now, research has not yet established how much it matters what things are to be learned—that is, whether they are easy or difficult tasks.

In 1963 I published a brief essay on what I called "a model of school learning" (Carroll, 1963). I postulated, among other things, that at least two kinds of aptitude tended to govern how much time an individual would need to learn something: (1) ability to understand instruction on the task (an ability that is probably related to *g*, or some component of *g*) and (2) aptitudes that are specific or intrinsic to the task to be learned, in the sense that possession of a reasonably high level of aptitude is necessary for learning the task. (If I were rewriting this essay today, I would lay more stress on the influence of *g* itself.) My model of school learning was adopted and expanded by Bloom (1976), Block and Burns (1976), and others, and validated in a series of studies. From Bloom's standpoint, intelligence tests measure "cognitive entry behaviors," that is, the prerequisite learning held to be necessary for a learning task. His research suggested, furthermore, that necessary learning time tended to decrease as the individual accumulated more cognitive entry behaviors.

All these considerations are relevant to evaluating Herrnstein and Murray's (1994, chap. 17) discussion of raising cognitive ability. In the main, they are correct in asserting that raising cognitive ability is not easy, in that numerous experiments and projects that had that as their goal have produced only small and uncertain results, far short of what might be achieved, they claim, by near-birth adoption of children from disadvantaged families into better advantaged families. As some readers may have interpreted their discussion, however, the implication seemed to be that trying to raise cognitive skills, or even to institute training programs for teaching skills to those of low cognitive ability, is not worthwhile. This, I think, would be a mistake, because low cognitive ability does not necessarily imply untrainability. It only implies the probability that the training time that would be necessary would be longer, and more expensive, than for persons with higher cognitive skills. Military training programs described by Sticht (1992) were often successful, in a wide range of military specialties, in raising literacy and producing individuals who were well trained. There is no reason why similar programs could not succeed in the civilian world. The important goal is not to raise IQs but to put whatever cognitive skills do exist to work.

To summarize: Experts have largely neglected what seems to be an obvious conclusion to be drawn from the evidence from IQ tests: that IQ represents the degree to which, and the rate at which, people are able to learn, and retain in long-term memory, the knowledge and skills that can be learned from the environment (that is, what is taught in the home and in school, as well as things learned from everyday experience). As shown in Figure 3, children with high IQs rather quickly attain mental ages well above those of their age peers, and they maintain competence over many years, whereas children with low IQs are much slower in learning what the environment exposes them to, and they are deficient in retaining what they learn. Differences in IQ among adults are the resultant of these differences in learning rates over the years of childhood, adolescence, and later on.

To the extent that people can judge the degree to which others can learn and retain those knowledges and skills offered to them by the environment, the general factor *g* does indeed correspond to what people conceive of as intelligence or "being smart." It would be highly useful to pursue research devoted to obtaining more details about people's concepts of intelligence and the extent to which those concepts correspond to what is measured using different kinds of intelligence tests. Even more useful would be research on the degree to which *g* corresponds to people's rates of learning different kinds of tasks or tasks of different degrees of complexity.

For more discussion of the relation between IQ and learning, see Gottfredson (1997).

PROPOSITION 4

> *IQ scores are stable, although not perfectly so, over much of a person's life.* (Herrnstein & Murray, 1994, p. 23)

This proposition is well supported in the literature, which says that although abilities (in an absolute sense) may increase over time with maturation, education, and other effects, individuals tend to hold approximately the same position relative to their age cohort throughout their life. Thus, the correlations of abilities from one age to another are generally high. Early on, however, it was observed by Anderson (1939) that the correlations decreased with the number of years separating any two measurements. On this basis Anderson formulated a so-called overlap hypothesis that implied that a correlation between ability at one time and ability at another time was solely the result of the overlap between the skills known at the earlier time and the skills known at the later time, which were assumed always to include the skills known at the earlier time. This formulation implied that the "constancy of the IQ" was merely an artifact. More recent research (Cronbach & Snow, 1977), however, has shown that the overlap hypothesis itself is flawed and that IQ scores are indeed relatively stable over time, probably because of the basic individual characteristics that are preserved throughout life. For example, in a well-known longitudinal study (Jones & Bayley, 1941; Pinneau, 1961), the correlations between IQs averaged over different ages and the IQs at ages 17 and 18 were .41 (for months 10, 11, and 12), .62 (for months 42, 48, and 54), .86 (for years 5, 6, and 7), and .96 (for years 14, 15, and 16). Thus, IQs taken just before the child reached age 1 showed an appreciable correlation with IQ at age 17 or 18, and IQs taken at later ages showed even much higher correlations. For further discussion, see Carroll (1993, pp. 662–664) and Moffitt et al. (1993).

PROPOSITION 5

Properly administered IQ tests are not demonstrably biased against social, economic, ethnic, or racial groups. (Herrnstein & Murray, 1994, p. 23)

This proposition, also, is well supported by massive evidence from psychometric studies, as summarized and extensively discussed by Brody (1992), Jensen (1980), and others. The bottom line is that IQ scores from most standard tests of intelligence correctly assess, within small standard errors of measurement, the individual's amount of progress, relative to his or her age cohort, in achieving the mental proficiency that it is possible for one to attain in an advanced culture such as ours. To the extent that such scores can predict success in school, in a training course, or in an occupation, they tend to make similar predictions for different social, economic, ethnic, or racial groups, regardless of the fact that average scores for those different groups may differ for one or more reasons. Agencies that construct and develop standardized tests of intelligence or scholastic aptitude make every effort to minimize bias in such tests.

PROPOSITION 6

Cognitive ability is substantially heritable, apparently no less than 40 percent and no more than 80 percent. (Herrnstein & Murray, 1994, p. 23)

I do not attempt to present the evidence supporting this proposition, for that is done by Plomin and Petrill (1997) in another article published in this issue of *Intelligence*. I may, however, point out that there is no major discrepancy between Herrnstein and Murray's proposition and Plomin and Petrill's article. Plomin and Petrill attest that "genetic influence on individual differences in intelligence is significant and substantial" (p. 56). Their Figure 2 leads to the conclusion that "about half of the variance of IQ scores is attributable to genetic factors," but their Figure 3 suggests that the heritability of IQ actually rises as age increases, from about .4 at 4 to 6 years to about .8 for older adults (precisely the values mentioned in Herrnstein and Murray's proposition). They emphasize that the same genetic influences affect IQ at different ages, that they affect different cognitive abilities in the same way, and that they affect both IQ and scholastic achievement.

There is one caution to be observed, however. Heritability is a statistic that applies to a defined population, not to individual members of that population. This caution needs to be borne in mind in interpreting an IQ score for an individual. In the early days of psychometrics, investigators often made the mistake of suggesting that an IQ score could be directly interpreted as a measure of the individual's genetic inheritance. The effect of this error survives to the present day in the public's perception of IQ scores, or scores on such tests as the Scholastic Aptitude Test (the SAT), which tend to be highly correlated with IQ scores. For example, the typical response that one makes on learning one's IQ or one's SAT score is to assume that it represents a measure of one's hereditary intelligence, with no implication about possible environmental effects. Many critics of the *g* described by Herrnstein and Murray seem to have made this same kind of mistake. In the case of an individual IQ score, the genetic influence might be small, or it might be large. It is only on the average that it might be about half of the total variance. It may be the case that for individuals who have for one reason or another suffered disadvantages in opportunities to learn, the effect of nurture (or its lack) might be large.

I would also like to point out that, whereas Herrnstein and Murray seemed to suggest that genetic influences may be responsible for at least part of the acknowledged differences in Blacks' and Whites' mean IQ scores, the psychometricians Bock and Moore (1986) came to a different conclusion. They did this despite the fact that they used the same NLSY data that Herrnstein and Murray used. In an extensive discussion of their findings (pp. 90–96), Bock and Moore rejected three possible theories of the mean IQ differences among Whites, Blacks, and Hispanics: genetic endowment theories, formal models of status attainment, and "linguistic theories." Instead, they stated their belief that the evidence supports what they called a "community norm theory," such that "[a]ll members of the school social system participate in the development and maintenance of the school's normative climate and find ways to constrain performance demands from individual teachers that are perceived as inappropriate according to existing norms" (p. 93).

Bock and Moore's community norm theory fits into the conception outlined here, whereby intellectual development as indexed using IQ tests is in part the resultant of the opportunities afforded by schools, families, and everyday experience. The lower average IQ scores of Blacks, thus, could be caused more by lower community norms and poorer educational facilities than by any presumed genetic inferiority. Nevertheless, the question of whether race differences in intelligence are to be regarded as partly the result of genetics cannot yet be taken to be settled by research; for a further view, see Rowe (1997).

CONCLUSIONS

I have reviewed Herrnstein and Murray's (1994, pp. 22–23) six propositions about cognitive ability and the *g* factor, and I find that all of them are well supported in psychometric and behavioral genetic research. (This says nothing, however, about the validity of the conclusions that Herrnstein and Murray draw about the consequences of variations in intelligence.) Along the way, I have tried to explain how these propositions should be interpreted and understood by the general public. At various points I have introduced thoughts about why psychometric findings have been poorly presented and misunderstood by the public, and I have suggested some possibly new interpretations of these findings. The major points that I hope will be noticed and remembered are the following:

(1) Psychometrics is a rigorous and scientifically respectable discipline that has developed a sound technology for constructing and analyzing psychological variables, such as scores on tests of cognitive skills, and for identifying and describing different abilities. It is in no sense a "pseudoscience." Over the years, however, psychometricians have struggled with a number of conceptual problems that have caused the general public to misunderstand or misperceive their findings.

(2) Psychometrics correctly reports that a major portion of the variation in intelligence can be interpreted as a *g* or general ability factor that pervades measures of intelligence, mental growth, and cognitive achievement. It also shows that there are numerous cognitive abilities other than *g*.

(3) Intelligence or general ability test scores should not be taken as direct measures of hereditary intelligence. Rather, they should be taken as measures of an individual's progress, at a given point of time and for whatever reasons (genetic or environmental), in attaining the full range of mental development that is possible in an advanced society. Over the years of childhood to young adulthood, mental ages report the absolute amount of this progress, and IQs report the relative rate of that progress as compared with that of persons of comparable chronological age, with an IQ of 100 supposedly representing the average rate.

(4) There are numerous causes or sources of mental development, only some of which may stem from the individual's biological, genetic makeup. From a social and educational standpoint, important sources reside in parenting and home influences in the individual's infancy and childhood, in schooling, in experiences with communication media and other societal facilities, and in general social interaction. At least some of the persons who test low in IQ are those who have not had the advantages and opportunities that persons with higher IQs have had.

(5) Measures of *g* or general ability do not necessarily indicate anything about the individual's absolute ability to learn. It is more likely that they indicate something about the amount of time that an individual needs to master a given task or to successfully complete a course of learning. All individuals are trainable or educable to a certain extent; they may vary in their rate of learning and the level of mastery they can easily achieve.

(6) Psychometrics could provide better information than it now does on the real life meanings of psychological test scores. In particular, it should provide better information about what scores mean in terms of individuals' probabilities of success in different types of learning and the time needed for that learning to be successful. This will require much further research, including meta-analysis of results already available.

(7) Because of the need for more knowledge about cognitive skills and their properties, such as their improvability and trainability, the nature of intelligence

and the identification and measurement of cognitive skills are important topics for further scientific research.

Acknowledgment: I wish to thank Linda Gottfredson, John Horn, Lloyd Humphreys, Robert Plomin, and Valerie Williams for comments on earlier drafts. None of these, of course, is to be held responsible for errant statements and opinions that may still remain.

REFERENCES

Anderson, J.E. (1939). The limitations of infant and preschool tests in the measurement of intelligence. *Journal of Psychology, 8*, 351–379.

Block, J.H., & Burns, R.B. (1976). Mastery learning. *Review of Research in Education, 4*, 3–49.

Bloom, B. (1976). *Human characteristics and school learning*. New York: McGraw-Hill.

Bock, R.D., & Moore, E.G.J. (1986). *Advantage and disadvantage: A profile of American youth*. Hillsdale, NJ: Erlbaum.

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

Brown, A.L., & Campione, J.C. (1986). Academic intelligence and learning potential. In R.J. Sternberg & D.K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex.

Butterfield, E.C. (1986). Intelligent action, learning, and cognitive development might all be explained with the same theory. In R.J. Sternberg & D.K. Detterman (Eds.), *What is intelligence? Contemporary viewpoints on its nature and definition*. Norwood, NJ: Ablex.

Carpenter, P.A., Just, M.A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404–431.

Carroll, J.B. (1963). A model of school learning. *Teachers College Record, 64*, 723–733.

Carroll, J.B. (1987). The national assessments in reading: Are we misreading the findings? *Phi Delta Kappan, 68*, 424–430.

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

Carroll, J.B. (1995). Reflections on Stephen Jay Gould's *The Mismeasure of Man* (1981): A retrospective review [Editorial]. *Intelligence, 21*, 121–134.

Cattell, R.B. (1971). *Abilities: Their structure, growth, and action*. Boston: Houghton Mifflin.

Ceci, S.J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology, 27*, 703–722.

Ceci, S.J., & Bruck, M. (1994). The bio-ecological theory of intelligence: A development-contextual perspective. In D.K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence*. Norwood, NJ: Ablex.

Cole, M., & Means, B. (1981). *Comparative studies of how people think: An introduction*. Cambridge, MA: Harvard University Press.

Cronbach, L.J., & Snow, R.E. (1977). *Aptitudes and instructional methods: A handbook for research on interactions*. New York: Irvington.

Detterman, D.K. (Ed.). (1994a). *Current topics in human intelligence: Vol. 4. Theories of intelligence*. Norwood, NJ: Ablex.

Detterman, D.K. (1994b). Theoretical possibilities: The relation of human intelligence to basic cognitive abilities. In D.K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence*. Norwood, NJ: Ablex.

Eysenck, H.J. (1939). Review of Thurstone's *Primary Mental Abilities. British Journal of Educational Psychology, 9*, 270–275.

Eysenck, H.J. (1994). A biological theory of intelligence. In D.K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence*. Norwood, NJ: Ablex.

Flynn, J.R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101*, 171–191.

Foley, P., & Rucker, L.S. (1989). An overview of the Armed Services Vocational Aptitude Battery (ASVAB). In R.F. Dillon & J.W. Pellegrino (Eds.), *Testing: Theoretical and applied perspectives*. New York: Praeger.

Fraser, S. (Ed.), (1995). *The Bell Curve wars: Race, intelligence, and the future of America*. New York: Basic Books.

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

Giroux, H.A., & Searls, S. (1996). *The Bell Curve* debate and the crisis of public intellectuals. In J.L. Kincheloe, S.R. Steinberg, & A.D. Gresson (Eds.), *Measured lies:* The Bell Curve *examined*. New York: St. Martin's Press.

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

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

Gould, S.J. (1994, November 28). Curveball: Review of R.J. Herrnstein & C. Murray, *The bell curve: Intelligence and class structure in American life* (New York: Free Press, 1994). *The New Yorker*, 139–149. [Reprinted in Jacoby & Glauberman (1995), pp. 3–13.]

Gustafsson, J.-E. (1984). A unifying model for the structure of intellectual abilities. *Intelligence, 8*, 179–203.

Hart, B., & Risley, T.R. (1995). *Meaningful differences in the everyday experience of young American children*. Baltimore: Brookes.

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

Horn, J., & Noll, J. (1994). A system for understanding cognitive capabilities: A theory and the evidence on which it is based. In D.K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence*. Norwood, NJ: Ablex.

Humphreys, L.G. (1994). Intelligence from the standpoint of a (pragmatic) behaviorist. In D.K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence*. Norwood, NJ: Ablex.

Jacoby, R., & Glauberman, N. (Eds.). (1995). *The* Bell Curve *debate: History, documents, opinions*. New York: Times Books.

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

Jensen, A.R. (1994). Phlogiston, animal magnetism, and intelligence. In D.K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence*. Norwood, NJ: Ablex.

Jones, H.E., & Bayley, N. (1941). The Berkeley growth study. *Child Development, 12*, 157–173.

Kaufman, A.S. (1994). *Intelligent testing with the WISC-III*. New York: Wiley.

Kirsch, I.S., & Jungeblut, A. (1986). *Literacy: Profiles of America's young adults*. Princeton: Educational Testing Service. (ETS Report No. 16-PL-02)

Krechevsky, M., & Gardner, H. (1994). Multiple intelligences in multiple contexts. In D.K. Detterman (Ed.), *Current topics in human intelligence: Vol. 4. Theories of intelligence*. Norwood, NJ: Ablex.

Lord, F.M., & Novick, M.R. (1968). *Statistical theories of mental test scores*. With contributions by Allan Birnbaum. Reading, MA: Addison-Wesley.

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

National Assessment of Educational Progress (1985). *The reading report card: Progress toward* *excellence in our schools; Trends in reading over four national assessments, 1971–1984.* Princeton: Educational Testing Service. (ETS Report No. 15-R-01)

Olson, D.R. (1986). Intelligence and literacy: The relationships between intelligence and the technologies of representation and communication. In R.J. Sternberg & R.K. Wagner (Eds.), *Practical intelligence: Nature and origins of competence in the everyday world.* Cambridge: Cambridge University Press.

Pinneau, S.R. (1961). *Changes in intelligence quotient: Infancy to maturity: New insights from the Berkeley Growth Study, with implications for the Stanford-Binet scales, and application in professional practice.* Boston: Houghton Mifflin.

Plomin, R., & Petrill, S.A. (1997). Genetics and intelligence: What's new? *Intelligence, 24*(1), 53–77.

Ralph, J., Keller, D., & Crouse, J. (1994). How effective are American schools? *Phi Delta Kappan, 76*, 144–150.

Raven, J.C. (1938–65). *Progressive matrices.* New York: Psychological Corporation.

Reynolds, C.R. (Ed.). (1994). *Cognitive assessment: A multidisciplinary perspective.* New York: Plenum.

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

Sagiv, A. (1979). General growth model for evaluation of an individual's progress in learning. *Journal of Educational Psychology, 71*, 866–881.

Spearman, C. (1904). "General intelligence," objectively determined and measured. *American Journal of Psychology, 13*, 201–293.

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

Spearman, C. (1939). Thurstone's work reworked. *Journal of Educational Psychology, 30*, 1–16.

Sternberg, R.J. (1995). For whom the bell curve tolls: A review of the book *The Bell Curve. Psychological Science, 6*, 257–261.

Sternberg, R.J., Conway, B.E., Ketron, J.L., & Bernstein, M. (1981). People's conceptions of intelligence. *Journal of Personality and Social Psychology, 41*, 37–55.

Sternberg, R.J., & Detterman, D.K. (Eds.). (1986). *What is intelligence? Contemporary viewpoints on its nature and definition.* Norwood, NJ: Ablex.

Sternberg, R.J., & Wagner, R.K. (Eds.). (1986). *Practical intelligence: Origins of competence in the everyday world.* Cambridge: Cambridge University Press.

Sternberg, R.J., Wagner, R.K., Williams, W.H., & Horvath, J.A. (1995). Testing common sense. *American Psychologist, 50*, 912–927.

Sticht, T.G. (1992). Military testing and public policy: Selected studies of lower aptitude personnel. In B.R. Gifford & L.C. Wing (Eds.), *Test policy in defense: Lessons from the military for education, training, and employment.* Boston: Kluwer.

Terman, L.M. (1916). *The measurement of intelligence: An explanation of and a complete guide for the use of the Stanford revision and extension of the Binet-Simon intelligence scale.* Boston: Houghton Mifflin.

Terman, L.M., & Merrill, M.A. (1960). *Stanford-Binet intelligence scale: Manual for the Third Revision Form L-M.* Boston: Houghton Mifflin.

Thorndike, E.L., et al. (1921). Intelligence and its measurement: A symposium. *Journal of Educational Psychology, 12*, 123–147, 195–216, 271–275.

Thorndike, R.L., Hagen, E.P., & Sattler, J.M. (1985). *Stanford-Binet Intelligence Scale* (4th ed.). Chicago: Riverside.

Thurstone, L.L. (1938). Primary mental abilities. *Psychometric Monographs*, No. 1.

Thurstone, L.L. (1947). *Multiple factor analysis: A development and expansion of* The Vectors of Mind. Chicago: University of Chicago Press.

Thurstone, L.L., & Ackerson, L. (1929). The mental growth curve for the Binet tests. *Journal of Educational Psychology*, *20*, 569–583.

Woodrow, H. (1917a). Practice and transference in normal and feeble-minded children. 1. Practice. *Journal of Educational Psychology*, *8*, 85–96.

Woodrow, H. (1917b). Practice and transference in normal and feeble-minded children. 2. Transference. *Journal of Educational Psychology*, *8*, 151–165.

Woodrow, H. (1946). The ability to learn. *Psychological Review*, *53*, 147–158.