Psychological Inquiry
1994, Vol. 5, No. 3, 179–192

TARGET ARTICLE

Intelligence From the Standpoint of a (Pragmatic) Behaviorist

Lloyd G. Humphreys
University of Illinois, Urbana–Champaign

Intelligence is defined as a phenotypic behavioral trait. It is the size of the intellectual repertoire measured at a particular point in time. Circularity is avoided by defining intellectual by consensus among persons doing research in the area. Information about the intercorrelations of elements in the intellectual repertoire is used to delineate further the construct. When assumptions about human development are added, the resulting theory allows testable predictions to be made about the trait. The theory is, however, one of small scale. It provides no theorems about how problems are solved, and it is not antithetical to research that leads to understanding the anatomical and biochemical mechanisms underlying the behavioral trait. The latter is assumed to have both genetic and environmental substrates. Intelligence so defined has many important correlates that are congruent with the theory.

Intelligence from the standpoint of even a pragmatic behaviorist may seem like a contradiction in terms for many readers, but "proof of the pudding is in the eating." Any behavioristic approach seems outmoded in today's climate of opinion in which cognitive psychology is ascendant, but psychology is still the science of behavior, not of cognition per se. Our task is to explain and predict behavior, with or without hypothetical mental constructs.

Choice of vocabulary to be used in psychological theories is an important problem. Our vocabulary is shared with the humanities, theology, law, and people at large to a greater extent than is the case in other sciences. There are two realistic choices—either adopt a new vocabulary or redefine words as necessary for scientific purposes. I have chosen the latter alternative. A third alternative that requires molding psychological theory to fit the definitions of the man-in-the-street is not acceptable.

Basic Approach

Intelligence is defined as a phenotypic behavioral trait. A phenotypic trait is an observable characteristic, such as height or length, of a biological organism. Physical traits differ in most cases from behavioral ones, however, in being easier to define and measure. The former tend to stand out perceptually from one another so that agreement on what is being measured is more readily attained, but there are exceptions. The distinction between systolic and diastolic blood pressure and the importance of measuring each required both research and conceptual development.

Measurement Methodology for a Behavioral Trait

After a phenotypic trait has been tentatively defined and a measure proposed, certain functional properties of the measures of the trait must be investigated. Among these are the reliability of the measurements, the robustness of the measurements under a variety of conditions, and the stability of measurements over time as maturation and learning take place. Because behavioral traits are generally measured by psychological tests, a fourth category, the homogeneity of the test items, must be added. Each item must measure a component of variance common to all the items that have been proposed as indicators of the trait. The trait variance in the items, as long as it is not zero, is not required to be of any minimum size. If it is small, the trait variance in the total score on the test can be pushed to an acceptable level by adding the needed number of parallel items. High correlations among items are not required, but high homogeneity of the total score (coefficient alpha) is needed for satisfactory measurement of a hypothetical trait.

The next step is to study the correlates of the trait. Correlations are obtained with other traits, including both physical and behavioral ones, and with experimental treatments. Without knowledge of measurement characteristics, interpretation of these correlations is uncertain. An important interpretation requires, for a specified population, the amount of variance in the measure of the trait that is not shared with measures of established traits.

Scope of Definition

Defining intelligence as a phenotypic behavioral trait is not a substitute for understanding how people or other biological organisms solve problems that are said to require intelligence. Neither is it a substitute for understanding the anatomical and biochemical mechanisms underlying the behavioral trait. These problems are legitimate and important, but they supplement the present approach. Research in those areas, if it is to be related to a behavioral trait, requires definition and measurement of the trait.

Outline of Discussion

In the discussion that follows, I define the trait of intelligence, describe the requirements for measurement, develop in greater detail how the construct is organized and measured, make some assumptions about its bases in development, and describe hypotheses inferred from the ground work thus laid. I conclude with remarks concerning the importance of the construct in human activities.

Construct of Intelligence

I start with the definition of *intelligence* that appeared some years ago (Humphreys, 1971). Intelligence is the acquired repertoire of all intellectual (cognitive) skills and knowledge available to the person at a particular point in time. Individual differences in intelligence are monotonically related to the size of this repertoire. To avoid circularity, *intellectual* is defined by the consensus among experts working in the area. The repertoire is acquired during development, but it is acquired, stored, and retrieved by a biological organism. Thus, there are both a genetic substrate and an environmental substrate for the trait.

Measuring the Repertoire

Because intelligence is defined as the total of all intellectual skills and knowledge, the only restriction placed on a measure of intelligence is that it assess the repertoire veridically. A test cannot measure the entire repertoire, but it can measure a broad representative sample of the elements. Similarly, a rater cannot have observed ratees exhibiting all the knowledge and skills in their repertoires, but the number and breadth of situations in which performance was observed are critical.

Under appropriate circumstances, tested intelligence and rated intelligence converge on the same construct. On the test side, the appropriate circumstance is use of a standard test such as the Stanford–Binet intelligence test or one of the Wechsler scales. The appropriate circumstances for ratings are (a) the use of multiple, independent raters who have had adequate opportunity to observe the ratees and (b) the education of raters to distinguish intellectual behaviors from character and personality characteristics.

Because the repertoire can only be known behaviorally, examinees must be both able and willing to produce the behaviors to be measured. Thus, motivation of examinees to perform well is critical. Physical characteristics, such as blindness and deafness, and psychoeducational characteristics, such as bilingualism, complicate the measurement problem.

Inequality of Opportunity

The measurement of intelligence has been complicated by loose thinking concerned with opportunity to acquire the behaviors in the repertoire. Some have defended tests of intelligence on grounds that opportunity of potential examinees in a particular culture was approximately equal. Others have rejected the tests because opportunity was clearly unequal. Still others have tried to devise items that would be culture free. All have missed an essential point: It is the phenotype that is being measured. Performance in life is a function of phenotypic traits, not estimated genotypes.

The behaviors observed and measured when assessing the phenotypic trait of intelligence are acquired by a biological organism dependent on the luck of the genetic draw. Development also occurs in widely different environments and in constantly changing environments. Opportunity is obviously not equal between or within any grouping of a population or for a given person over time. Whether phenotypic traits can be modified by environmental manipulation, by how much, by what methods, in what period of time, at what point in development, and at what costs are appropriate research questions.

Exposure to Content

Of course, there is an experiential requirement for a valid test, although equality of experience is not necessary. If the test is in English, familiarity with that language is required. The examinee must have had potential access to the information, knowledge, and skills required by the test questions. Exposure requires little more than presence in the environment in which examinees' repertoires were acquired. Compulsory education, TV, radio, newspapers, magazines, and public libraries provide widespread potential access.

Organization of the Repertoire

Elements of the repertoire form a positive manifold (intercorrelations are positive) that is the sole basis needed for the use of the word *general* to describe intelligence (Humphreys, 1979). The correlations among individual acts or items are, however, quite small. Intelligence is general, but, at this level of detail, the amount of specificity in any element is far and away more in evidence than is generality. When test items are linearly combined to form a total score, or when individual acts are summated in the form of ratings, grades in school, or performance in industrial or military training, correlations among the aggregates of the elements become much higher. There is still unevenness in the level of performance of persons from one aggregate measure to another, but generality is now substantial. For example, the correlation between total score on verbal and quantitative tests is larger than the item correlations within either of the tests. Even so, verbal items are more like one another than they are like quantitative items.

Factors in the Repertoire

The considerable variation in the size of the correlations among aggregates of the repertoire is systematic, so that multiple factors can be defined. The positive manifold supports the use of the modifier *general* to describe intelligence; the systematic variation in size defines multiple factors. That the mathematics of factor extraction defines orthogonal factors from this systematic variation has been widely misinterpreted. There have also been widely disparate claims concerning their number. Unfortunately, the number depends on the selection of the tests to be factored. I (Humphreys, 1981) described a compelling case for the numbers to be in the thousands if one starts with a wide enough sample of tests and a large enough sample of examinees, so that small differences in correlations become stable. Factor analysis of tests in the first order of factoring, and stopping there, represents a dead end psychologically because the broader dimensions of greater psychological and social importance are lost. The definition of large numbers of intellectual factors is useful only as a way of characterizing the extent of the intellectual repertoire.

There is a second way of describing the systematic variation in size of the correlations in the intellectual repertoire that is independent of ordinary factor analysis. This is the radex model of Guttman (1954; Marshalek, Lohman, & Snow, 1983). To a first approximation, the elements of the repertoire can be located in two dimensional space in accordance with a small number of assumptions and principles. Each aggregate measure is assumed to be error free, and the sample of measures is assumed to be intensive as well as extensive. Intensive sampling has the effect of providing little room for non–error-specific variance, unless specificity is characteristic of the measure as distinguished from being a function of the sampling of tests. The cognitively most complex measures are grouped closely in the center of the space whether their content is verbal, numerical, or figural. Variation in complexity is portrayed along radii leading out from the center, with the simplest measures in the periphery. True specifics are due to superficial aspects of measures and therefore are cognitively simple. Content does have an effect on the size of correlations, so that verbal, quantitative, and figural items are clustered in pie-shape segments of the area of the space. Distance between points is inversely related to the size of the correlations between measures. For example, a measure of quantitative reasoning would be close to the center of the space, and accuracy in clerical number-checking would be in the same segment of the space but close to the periphery.

Defining a General Factor

The intercorrelations of the elements of the repertoire define a general factor on which every element has a positive loading. Also, when the contribution of the general factor to covariation is held constant, there is no residual negative covariation. The size of the loadings varies with the variation in size of the mean correlations of the individual elements, which in turn are a function of complexity. Given adequate methodology (described in Humphreys, 1982), the general factor can be defined uniquely from one sample of intellectual measures to another and from one sample of the general population to another. (Restriction of range of talent has effects on factors that are beyond the scope of this discussion.) The general factor represents a substantial amount, but not all, of the common variance in the repertoire (Carroll, 1993).

The construct of general intelligence does not necessarily represent anything more than a mathematical dimension. As such, it is "real," even though it cannot be observed under a microscope. Knowing that there is a highly replicable general factor that describes a substantial proportion of the covariance among the elements of the repertoire is both a necessary and a sufficient condition for the construct and its measurement.

Tests of the General Factor

The general factor of intelligence cannot be measured directly because each and every item includes variance that is orthogonal to the general factor. The latter can only be estimated by a test, but, given a sufficient number of items that sample the domain with sufficient breadth, the "noise" in the items contributes progressively less to their linear combination (the total score) as the number of items increases. It may seem paradoxical, but the variance of the total amount of noise shrinks, and the variance associated with the central construct increases as the number of items increases—as long as each item measures the general factor to an appreciable extent and as the heterogeneity of the residual noise is maximized. (For a fuller discussion, see Humphreys, 1985.) The indeterminancy in estimating the general factor can approach zero closely if there is a sufficient number of carefully selected items.

Returning to Guttman's model, the two-dimensional plot of the elements in the repertoire serves as a guide to the selection of the most effective types of items to include in an intelligence test. Items close to the centroid of the space have the highest loadings on the general factor, so that a relatively small number of them provides a valid estimate of that factor. Tests of knowledge of the meaning of abstract words, arithmetic reasoning, and spatial reasoning are close to the centroid. Items like these appeared in the original Binet–Simon scale.

Alternative Samples From the Repertoire

Although the types of items included in so-called aptitude and achievement tests are not identical, many tests labeled *achievement* lie very close to the centroid of the space. Reading comprehension is the example par excellence. If a composite is formed from the individual tests of a standard academic achievement series, the correlation of that composite with a standard test of intelligence will be almost as high as the correlation between two standard tests. An achievement composite obtained from the 12th-grade data for Project Talent has a correlation with the intelligence composite of .850 and .825 in males and females, respectively. These correlations are evaluated in the light of the reliability of the intelligence composite, which is estimated as .899 and .887 in the same two samples. The achievement composite, furthermore, is not fully representative of the intellectual repertoire.

If one selects tests from an array somewhat removed from the centroid and obtains a composite, another high correlation is obtained. Nonacademic-information tests are not near the centroid, but a composite of such tests from the 12th grade of Project Talent, also not fully representative of the space, has correlations with the intelligence composite of .787 and .777 in the two sexes, and the correlations with academic achievement are .826 and .814. Furthermore, most of the nonacademic-information tests show substantial sex differences in both directions, although there are more tests in the set analyzed showing male versus female superiority. Virtually every item includes a substantial systematic-bias component of variance.

Within each sex, as well as between sexes, there are wide differences in the opportunity for a given person to acquire information about topics such as the Bible, hunting, fishing, dance, exotic foods, and so forth. Equality of experience with the content of the items is not required in estimating the size of the repertoire if the sampling is broad enough. Although it has not been done, a composite formed from an even larger number of tests widely scattered around the periphery of the space and heterogeneous in methodology as well as content would produce similar correlations. Such a test would not be practical either in time or conditions to administer, but it should be formed experimentally. Because tests of short-term memory are prominent in the periphery, obtaining the predicted outcome would have theoretical significance.

Ratings in the Test Space

Ratings of intelligence are also in the space defined by the test scores. Their location depends on several factors. If the rater does not know the ratee well, or if the acquaintance is based on a limited behavioral sample, the ratings are likely to be peripheral. If raters are allowed to select their own individual definitions of intelligence but are otherwise well qualified to rate, ratings will scatter in the space as a function of the definition adopted. If well-qualified raters are instructed to disregard desirable qualities of the ratee that are not a part of the intellectual repertoire and to rate on use of symbols, solving problems, and so forth, ratings will be closer to the centroid of the space. Each of the ratings made independently by *n* qualified raters will have a substantial correlation with the test of intelligence, and the aggregate of the *n* ratings will be more highly correlated with the scores on the standard test than the median of the individual ratings. The gain will depend on the independence of the raters and the breadth of the behavior sampled.

Performance Measures in the Test Space

Academic grades that are based on proficiency in the subject matter lie fairly close to the centroid of the test space but are also affected by content. English grades lie in the verbal segment, arithmetic grades in the quantitative segment, and engineering grades in design in the visuospatial segment. If letter grades had the same meaning for degree of proficiency in the course for all courses and all instructors, which they decidedly do not, a composite of grades would be closer to the centroid than grades in any single course. This does not deny that grades based on proficiency include other variance, such as motivation and hard work, but nonintellectual traits have limited impact in the short run on level of proficiency attained in a new cognitive task. Individual differences on the task are not static over time, but change in the cognitive domain takes place slowly.

Measures of proficiency in occupations, including supervisor and peer ratings, that include tasks from modest to high levels of complexity also appear in the test space. If performance on the job is cognitively complex, one knows in advance that correlations of that measure with scores on an intelligence test will be positive and of a size influenced primarily by the reliability and validity of the performance measure. Performance in different occupations such as mechanical maintenance and repair, on the one hand, and clerical occupations, on the other, falls in different segments of the test space but is close enough to the centroid in the respective segments that correlations with general intelligence are quite substantial.

This is not a revival of the identical-elements theory of transfer of training. There is no assumption that the elements on the test overlap directly with the elements of the job. Both the test and the job sample a much broader repertoire (by several orders of magnitude). It is the overlap in repertoires required that is responsible for the predictive validities.

Standard Tests of Intelligence

Although there are many tests labeled *intelligence,* two stand out from the rest in terms of their item sampling as good approximations of the general factor—the Stanford–Binet intelligence test and the various Wechsler scales. Based on content, they can be called *standard* tests of intelligence, but both share a limitation. Neither furnishes a useful score that measures the intelligence herein defined. Intelligence is the size of the repertoire—or the level on the general factor that reflects the size of the repertoire. Both these tests were designed to furnish IQs, which are measures of relative intelligence and do not reflect growth.

When IQs were computed by dividing mental age by chronological age, mental age was a measure of intelligence. An IQ is a useful addition, just as knowledge of height relative to an age group is a useful addition to knowledge of height. Mental-age units were abandoned because it was difficult to meet the requirements of an invariant scale for IQ with their use. The lack of a replacement is a serious gap because it confuses ability to perform with performance relative to age. For children, one recourse is to multiply the obtained IQ by the child's chronological age. Mental age, of course, was also not useful for adults, but a common scale for all age groups that would reflect the size of the repertoire could be formed. It is fortunate that test publishers retained grade-equivalent scores for achievement tests in spite of psychometric criticisms of their measurement properties. They do define a scale on which growth can be measured.

The standard tests and their close relatives obtain high marks on the four criteria described under the earlier heading of measurement methodology. There is no problem in measuring estimates of the general factor reliably as long as sufficient time and items are devoted to the task. Measuring a behavioral trait reliably is time consuming and requires standardization of critical conditions, but such measurement does not suffer otherwise in comparison to the use of physical scales. If examinees are well motivated—a critical condition—measurement is also remarkably robust to conditions in the examinee and the environment that are considered stressful. Among these are minor illness, fatigue, loss of sleep, ambient temperature, and ambient noise. These are not excuses for measuring intelligence unpleasantly, but they do show that the trait itself is robust.

There is no requirement that a behavioral trait be fixed and unchanging, just as there is no similar requirement for a trait of physique. Information about stability is required for theoretical and applied interpretations and is available for standard tests. Last, the standard tests are approximately homogeneous with respect to general factor content, but only because they are so heterogeneous with respect to the content introduced by their minor factors. The traditional criticism that a given score can be attained by different patterns of response is beside the point as long as the score is used and interpreted only as an estimate of the general factor.

It is fortunate in many ways that the two standard tests do a reasonably good job in estimating the intellectual repertoire. Item types have been sufficiently constant over 80 years that the large number of correlates of scores of these tests and their close relatives can be considered comparable over the same time period. It is more fruitful to retain the standard tests and to redefine what developers and most users thought the tests were measuring than to continue the search for the Holy Grail of "real intelligence."

Assumptions About Development

My basic definition of intelligence required a single paragraph. The subsequent discussion depended very largely on empirical observations of correlations among behaviors in fleshing out the construct. Still needed, however, are some assumptions about human development.

Genetic Substrate

The genetic substrate for general intelligence is typically polygenic. I exclude cases involving pairs of recessive genes and chromosomal abnormalities. In a limited sense, because intelligence is a behavioral trait, the substrate includes the genetics of the whole organism. Obviously, this point of view can be pushed too far. On the other hand, it is simplistic to assume that the genetic substrate is restricted to anything like a bodily organ. It seems more reasonable that the genetically determined substrate consists primarily of the structural and chemical characteristics of the central nervous system, but the entire sequence from stimulus through to response should not be forgotten. The assumption that many genes are involved must be taken very literally—as well as the many, many products of the large number of genes.

An assumption that the genes in the substrate for intelligence do not all "fire" at the moment of conception is also reasonable in the light of numerous late-appearing genetic effects. Relative individual differences in physical traits are not stable during development and decay.

Although a great deal of evidence concerning the heritability of intelligence has been published, behavioral geneticists have only recently (Cardon, Fulker, DeFries, & Plomin, 1992) considered whether group factors, holding the general factor constant, have a specific genetic substrate. Cardon et al.'s methodology showed a larger contribution to variance of group factors of the genetic substrate than did mine (Humphreys, 1991). More research is required, but it is highly probable that the genetic contribution is greater for the general factor than for group factors. It also follows that narrower factors defined by the difference between a measure's communality and its reliability (specifics) include much larger environmental contributions to variance.

Environmental Substrate

Elements in the repertoire are acquired. Principles of learning and motivation are directly implicated, but environmental influences can also have indirect effects. Both types of influences are multiple. Polygenic determination is accompanied by polyenvironmental determination, and the combined causes have polybehavioral manifestations.

A detailed theory of learning and motivation is not required. If learning takes place effectively, elements will be added to the repertoire more rapidly than when the process is ineffective. If the content to be learned is cognitively complex, general intelligence will benefit more than if the content is simple. This does not mean that all persons acquire new elements in the repertoire with equivalent speed and sureness. As described earlier, there is a genetic substrate for the acquisition of the intellectual repertoire.

There is also an organic substrate that is environmental in origin. Differential experience produces differences in dendritic growth. It also seems probable that early experience is more effective than experience that comes later in maturity. The extended period of infancy and dependency in the human is accompanied by a high degree of plasticity, but this property wanes with increasing age. In addition, there are the accidents, diseases, and other physiological pathologies that affect the biological organism and may well affect in turn the acquisition of the intellectual repertoire (Lubinski & Humphreys, 1992). The objective of testing the repertoire, whether in the physically or psychologically handicapped, is to sample the repertoire, not to estimate the genotype.

Size of the Repertoire

Development of the intellectual repertoire starts very early in development but is also small at the outset. Infants are attentive to changes in stimulation revealed by turning of the head. Individual differences in this behavior are positively correlated with scores on tests administered later in development. Humphreys and Davey (1988) suggested that the general factor measured by a test at 12 months could be identified with the general factor measured several years later, even if the stability of individual differences over the time interval is quite modest. Although individual differences in both absolute and relative scores change rapidly early in development, the change appeares to be gradual and lawful. From this point of view, the general factor gradually evolves into its later manifestations by cumulative accretions.

By the age of 2, the repertoire is already large. Responses to verbal stimuli precede use of words, and nonverbal problem-solving has been prominent for many months. With almost 100% of American children in school from age 6 to age 14 and exposed to a broad, relatively homogeneous curriculum, growth tends to take place on a broad front. Both home and school require continuous practice of earlier elements. By age 18, the intellectual repertoire is huge.

Differential experience in the home and differential reinforcement of early school experiences provide a possible basis for the initial differentiation of group factors as well. The high-school curriculum provides somewhat greater opportunity for specialized growth (Atkin et al., 1977), whereas specialization becomes more apparent in postsecondary education. Even specialization increases the total repertoire, but growth in general intelligence slows markedly. Ultimately, decline occurs from disuse and organic change.

Relation Between the Repertoire and Gain

There is ample basis in the discussion of the genetic and environmental substrates for intelligence for the following conclusion: True score gains between Times 1 and 2 are necessarily imperfectly correlated with true score of the repertoire at Time 1. Another way of stating this involves the distinction between reliability and stability over time. The latter statistic will always be smaller than the root of the product of the reliabilities at Times 1 and 2, providing a corrected stability coefficient of less than 1.00.

The correlations of scores at base periods with gains and the correlations between successive gains can be quite small. Genes that "fire" late in development are determiners of bodily structures and functions, but, being independent of other determinants in their action, they introduce seemingly stochastic changes in development. Similarly, the environment is constantly changing—and, to a large extent, in a stochastic manner insofar as the individual child is concerned. Many years ago, following a proposal by Anderson (1939), Roff (1941) computed raw-score gains with raw-score bases and found largely small negative correlations. Because there is a bias in raw-score data toward negative values, it seems probable that the true-score correlations might have been no larger than small positives.

Hypotheses Derived From the Theory

At this point, it is appropriate to discuss a theory about the trait of intelligence as distinct from its definition. As such, it should lead to testable hypotheses about relations involving the trait. These relations constitute only a limited segment of the phenomena subsumed under the rubric of intelligence by psychologists. The process by which elements of the repertoire are acquired or how the repertoire is used in solving problems is clearly important, but it is outside the scope of the theory of the trait.

Data are available for many of the testable hypotheses about the trait. Eighty years of testing children with tests that differ little from what I have called the standard ones have provided a wealth of information. In an important sense, the theory was derived from the large body of data that has been accumulated. This approach may differ from a great deal of psychological theorizing, but it provides a coherent structure for the data and leads to additional research.

I discuss hypotheses under three headings—group means and changes in group means, stability and instability of relative intelligence of individuals, and some selected correlates of intelligence. I concentrate on describing the hypotheses and only incidentally on the evidence relevant to their validity.

Group Means

These hypotheses are admittedly difficult to test with confidence, but they are also related to important social issues. Imposition of adequate experimental controls, of which random assignment to treatments is crucial, is frequently not feasible. Statistical control is never completely adequate. Ability to test latent-trait models was an important advance, but such models only control the attenuation introduced by measurement error. There remain more important problems of the construct validities of the latent traits identified and whether an important one has not been identified. Because the problems are important, research must continue. Investigators must be aware of methodological weaknesses and temper their conclusions accordingly.

1. *The mean level of intelligence in a population can change over time.* Intelligence is the intellectual repertoire. The amount of education is related to the size of the repertoire, and amount of education is easy to measure. Quality of family life and education are presumptively involved also but are more difficult to measure. The enlisted population in U.S. military services had attained approximately 3 more years of education when they entered a service during World War II than during World War I (Tuddenham, 1948). More recently, Flynn (1987) reported similar increases, including specifically the perceptuovisual component of the repertoire. Critics use such data to denigrate intelligence tests, claiming that "real intelligence" did not change. A construct of "real intelligence" cannot be inferred from any measurement operation.

It is salutary to point out that what goes up can come down. Within recent memory, there has been a well-publicized decline in scores on the Scholastic Aptitude

Test. Jensen (1977) reported decreases in IQs for Black students in rural Southern schools from Grade 1 to Grade 6. Those students were deprived of the growth being shown in average American schools by deficits in their schooling.

2. *When the size of the intellectual repertoire changes, there will be corresponding changes in educational achievement and occupational proficiency.* Concern is expressed when scores go down, but increases are discounted. It is essential to the theory that World War II draftees and enlistees were more able than their World War I counterparts in acquiring the cognitive knowledge and skills in their military assignments, but data are lacking.

To test this hypotheses, it is essential to distinguish between a gain in score based on practice or coaching on a given set of items and a gain in the total repertoire. Current discussions concerning the importance of functional-literacy training for employment and industrial productivity assume the accuracy of the hypothesis.

3. *Psychologically important gains in intelligence will occur only with the expenditure of substantial effort in time and resources.* If the gain between World Wars I and II was largely due to the increase in years of education, consider the economic cost of keeping a growing population of children in school for approximately 3 additional years. One cannot expect to obtain a large gain in intelligence from an experimental program the duration of which is measured in months.

4. *For a given level of effort, there will be greater effects on young children than on older children.* Growth in intelligence is measured by growth in the intellectual repertoire. Producing a measurable increment to that repertoire by a given intervention can be accomplished more easily when the repertoire is relatively small.

5. *For effective growth in intelligence, there must be a continuous supportive psychosocial substrate.* Preschool intervention programs typically have short-term effects. There must be continuing exposure and continuing effort, but the latter is required of the learner as well as of society. Social effort that does not affect individual effort is not sufficient.

6. *Changes in intelligence are a function of the kind of intervening educational experiences.* In the system of education studied by Harnquist (1968), students could be grouped in the following categories—compulsory level, vocational, lower secondary, and gymnasium. Estimated gains, after allowing for either imperfect stability over time or imperfect reliability on pretest and posttest, increase monotonically from compulsory level to gymnasium. This listing of educational categories appears to be monotoic with respect to the intellectual content of the curriculum, the intellectual standards imposed, and the intellectual competition among students.

7. *Intervention can change the level of an individual component of intelligence more rapidly than the total.* Narrow repertoires can be increased more easily than broad ones. Even so, some components, such as aural and visual comprehension of language, are so central to the repertoire that obtaining more growth than typically occurs is slow at best. The intervention problem is quite different, however, among the small number of individuals who are substantially retarded in reading as compared to their level of aural comprehension.

8. *The components of general intelligence do not all grow at the same rate.* The age of leveling off varies from one component to another (Horn, 1989). These differences are correlated with differential exposure to the different types of content. Engineers and physical scientists, on the one hand, and humanists, on the other, will show different patterns of growth and decline for verbal, quantitative, and visuospatial components. The last of these is prominently involved in measures of so-called fluid intelligence.

9. *There are mean differences in intelligence among groups defined demographically.* Such groups do not experience identical environments and cannot be expected to have precisely equivalent repertoires. There may be genetic differences as well. In a relatively free society, groups defined by level of socioeconomic status have a genetic component to the variance of their scores on standard tests of intelligence. Groups whose gene pools have been partially segregated historically are likely to be genetically somewhat different. Differences in the size of the broad intellectual repertoire, or in a major component, are socially significant.

10. *It is difficult to overcome in adults deficits in phenotypic intelligence.* This hypothesis follows from the size of the intellectual repertoire. It is also possible that inadequate learning early in development has lasting effects. Thus, this hypothesis is independent of the nature–nurture problem. Observe the many problems encountered by adults in the acquisition of a second language. The popular view of intelligence is one of innate power or capacity that can be released at any time by providing opportunity and that can quickly come to full flower. There are no data to support this view. The intelligence I have defined develops slowly with support from the society and with effort on the part of the learner.

11. *For effective growth in intelligence, the specific curriculum and standards of achievement should be set somewhat above the current mean level of the group.* Individuals in the group must be currently motivated or can be motivated to work at the learning of the material, whatever it may be. If the students within the group are highly heterogeneous in preparation for learning the material, both the highly prepared and the poorly prepared are disadvantaged. It is especially disadvanta-

geous to be so far below the mean that competition is essentially impossible.

The preceding discussion seemingly conflicts with a great deal of the literature on the effects of ability grouping. That literature, however, is in conflict with the fundamental educational principle that learning should start at the level at which students currently are academically and should be pushed ahead (provided with opportunity, incentive, and skilled teaching) as far and as rapidly as possible. The educational operations suggested by the principle can be approximated by a well-designed ability-grouping program.

One explanation for failure to find effects in the expected direction of even well-designed programs is a fixation on making the evaluation instruments applicable to all participants. If there are no differences in content of the evaluation instruments, the evaluation was not well designed. Given valid instruments, if there are no differences in the expected direction for all groups, the program was not well designed and executed.

Stability of Individual Differences

Hypotheses in this area are concerned with change in relative intelligence over time. Time is not, of course, the effective variable; in the absence of information about the actual determiners, time is the appropriate dependent variable. Longitudinal research over time does require examinees who are relatively homogeneous in chronological age. Thus, it is immaterial whether a growth or relative score is used.

1. *Scores on a test of general intelligence intercorrelated over occasions will produce a quasi-simplex matrix.* Such a matrix tends to have its smallest correlation in the upper right-hand or lower left-hand corner and the largest correlations adjacent to the principal diagonal. If the correlations are obtained on a large sample and are corrected for the attenuation produced by measurement error, the preceding statement can be made without the qualification introduced by the word *tends.* There should be no break in a monotonic decrease in the size of the correlations from the principal diagonal to the periphery of the matrix along both rows and columns. Stability coefficients drop monotonically as time between occasions increases.

The corrections for attenuation also produce an R-matrix that can be compared with Guttman's definition of a true simplex. The latter requires a zero correlation between the true-score base on a first occasion and the true-score gain to a following occasion. Tests of the Guttman model show that it is not easy to reject independence of base and gain (Humphreys, Park, & Parsons, 1979; Humphreys & Parsons, 1979). Other models may produce acceptable fits also, but no alternative model assuming large correlations with gains is in that category. Low correlations with gains are congruent with intelligence as size of intellectual repertoire having multiple genetic and environmental determinants.

2. *Stability over time is a function of the initial age of the examinee.* As the repertoire increases in size, it becomes increasingly difficult to obtain a gain proportionate to the size of the base. The corollary of this hypothesis is that a gain of the same absolute size will have a larger effect when the repertoire is relatively small. It also follows, in the absence of special interventions, that the stability of intelligence over a given unit of time should increase throughout maturation.

3. *Low correlations between intelligence in the first few years and adult intelligence reflect the normal growth of intelligence.* This hypothesis is contrary to the common belief that early tests do not measure intelligence, but the hypothesis is viable (Humphreys & Davey, 1988).

The simplex model is the basis for an interpretation of continuity in the general factor of intelligence during early development despite modest correlations over the span of the preschool years. Guttman applied the model originally to concurrently obtained correlations among selected tests that differed in the complexity of the operations required. Intellectual development can be conceived as also arranged along a continuum of increasing complexity. Finding a behavioral item in infancy more highly correlated with a standard test at a later age than at age 2 would reject this hypothesis.

4. *Change is more rapid for the narrow components of intelligence than for the total score.* There may be differences from one component to another, but the results of simplex-fitting attempts by Humphreys et al. (1979) for 16 tests in male and female samples seem to be in line with this prediction.

5. *Change in individual differences is a function of the intervening psychosocial substrate.* More change is expected for students in an academic curriculum than in a narrowly oriented trade school.

6. *A learning environment that produces a larger mean gain in intelligence than that produced by a second environment will also produce more change.* Because gains have small correlations with initial bases, the larger the gain, the greater the amount of individual change in relative intelligence. The stability of individual differences will be lower when learning is most effective.

Socially Important Correlates of Intelligence

Intelligence tests have a huge number of correlates, and few are trivial in size. Here I discuss a few of these, all derivable from the definition and supporting assumptions.

1. *General intelligence and other continuously distributed psychological traits exhibit a modest degree of similarity among relatives.* Psychological traits have multiple genetic and environmental determinants, and these determinants can overlap only moderately among relatives. Differences, whatever the sources may be, produce resemblance coefficients that are a good deal less than unity. Lack of identity requires that children of extreme parents, on average, regress toward the childrens' mean on each trait as the parents of extreme children, on average, regress toward the parents' mean. For children reared with parents, the correlation between either of the two parents and a single child for intelligence in a wide range of talent is about .50.

2. *Proficiency of any performance in education, industry, and the military requiring intellectual (cognitive) tasks is correlated with scores on a standard test of intelligence.* Intellectual is again defined by consensus, and proficiency is represented by an aggregate measure, not a single act. The "validity generalization" doctrine was established empirically (Hunter, 1980; Schmidt & Hunter, 1977) but is readily derived from present theory. Educational and occupational tasks that broadly sample the repertoire are necessarily highly correlated with a test that also broadly samples the repertoire.

3. *Maximum predictive validity is obtained when the repertoire sampled by the test matches the repertoire sampled by the criterion performance.* Performance in a blue-collar mechanical occupation is predicted more accurately by tests of mechanical reasoning and visuospatial problem-solving than by a test of general intelligence.

4. *Performance measures at age 18 are predicted with increasing error the earlier the age of administration of the predictor test.* As predictors of intelligence at age 18, parental intelligence and socioeconomic status are only moderately valid (.50 and .40, respectively), and the child's intelligence at age 2 is no more accurate. If a program is widely adopted in which parents are expected to start saving for the expenses of higher education during the child's preschool period, problems will be created. Should all such children be admitted to college if they wish to attend and have a high school diploma? More important, should children whose parents made no attempt to save be discriminated against?

5. *Predictions of criterion performance from tests at any age will not have a correlation of constant size with successive occasions of criterion measurement.* All behavioral measures taken over several occasions show a quasi-simplex pattern of correlations. A predictor administered at a given point in time is correlated, as it were, with a moving target (Hulin, Henry, & Noon, 1990). Criterion performances can be conceived as moving from more to less complexity along the same continuum (acquiring a narrow skill), less to more complexity along the same continuum (from freshman mathematics to upper division courses), or gradually in the portion of the repertoire principally sampled (from reliance on verbal comprehension early in acquisition to spatial relations late). The rate of change in the size of correlates depends on the rate of change in individual differences on both the predictor test and the criterion measure.

6. *Decreased accuracy of prediction over time also requires information about the mean performance of the group.* Decreasing accuracy results in increasing amounts of regression of extreme persons toward the group mean. It is widely overlooked, however, that the group mean at the end of the time interval may itself be extreme with respect to the total population. The fundamental question is whether the members of a selected group regress toward a superior mean or toward the mean of the population from which they were selected. The level of performance in a select group is hypothesized to remain high as long as there is appropriate institutional support. Thus, the usefulness of the test whose predictive validity drops substantially with the passage of time cannot be evaluated solely by correlations computed within the group. There are numerous essentially stochastic events that occur during a given interval of time and that produce changes in individual performance, both up and down, that affect within-group correlations but that do not necessarily effect a superior performance by a selected group.

7. *Scores on an appropriate test have the same significance for criterion performance across different demographic groups.* Appropriateness is judged in terms of the exposure criterion. It is also judged in terms of the match in sampling the cognitive space of the test and the performance required. This hypothesis translates into an expectation of equality of intercepts in the regression of performance on the test after allowance is made for the effect on the intercepts of measurement error in the predictor. If an intercept difference remains, it is more productive to search for an additional valid predictor than for biased items in the present predictor.

8. *There are differences among socially important criterion performances in their predictability by an intelligence test or its most central components.* Among these are artistic, musical, and athletic abilities. The expectation, however, is not that predictive correlations will be zero—but that they will be much smaller than those with achievement in the so-called learned professions. In some cases, lower correlations are not produced by specific abilities in the periphery of the cognitive space but are produced by traits more appropriately called *personality* or *temperament.* Success in certain kinds of selling or leadership is an example. The importance of specifying kind is indicated by the difference between leadership in a learned profession and

in the command of an infantry platoon on the battlefield. Intelligence tests predict later ratings of officer effectiveness but do so less accurately than they predict either precommissioning or postcommissioning grades in technical training. The effectiveness ratings are also known to be predictable by rated differences in personality traits.

Importance of General Intelligence

General intelligence is one of many human traits, but it is clearly an important one. It is highly related to educational success and to occupational attainment. The latter occurs despite inadequate educational guidance and inadequate support in our society for children of high intelligence from working class and lower class family backgrounds. It is related to economic productivity, and, although it is a difficult research problem, it is undoubtedly related to the ability of a democratic society to function as such. Remember, in this connection, that there is little difference between functional literacy and general intelligence.

Illustrating the Importance

The traditional method of presenting correlations between psychological tests, including those that measure general intelligence, and important social criteria has led the general populace, politicians, and reporters to underestimate the importance of intelligence. Individual-difference correlations between intelligence and various military performance criteria may, on average, be as high as .70, whereas correlations with academic grades in college, a restricted range of talent, may be represented by values around .40. In both cases, standard errors of estimate are large so that there is a great deal of uncertainty in predicting individual performance.

Now consider an alternative (Lubinski & Humphreys, 1993). Divide the predictor test into multiple class intervals and compute a mean criterion score in each class interval. If there are adequate ceiling and floor for both the test and the criterion measure, the regression of criterion score on the test will be approximately linear. Now compute the product–moment correlation between the test and the criterion based on the means only, which is what engineers and experimental psychologists do when they wish to express the functional relation between two variables when each data point represents several observations.

When the preceding steps are followed, it is not unexpected in large samples to find a correlation of .99 for the means of tests and criteria. The sampling error of a given mean is equal to the standard error of estimate in the distribution of individuals divided by the square root of the number of observations in the mean.

These correlations have interesting properties. They are independent of the size of correlations computed in samples of individuals, but they are dependent on the accuracy of the assumption of linearity in the population and on sample size. Under conditions of linearity and large samples, correlations between distributions of means will inevitably approach unity. Several critics have incorrectly called this property a *fatal flaw,* but it is actually the principal asset of the methodology. Under favorable circumstances, group means can be predicted with almost perfect accuracy. To hide this fact by presenting only the correlation based on individuals has resulted in many errors in interpreting the correlates of intelligence.

Given linearity and large sample size, the squared correlation in the sample of persons can be approximated with high accuracy by the ratio of the variance of the criterion means to the variance of persons. However, the correlation based on individuals is not needed in making a judgment as to whether the gain in performance for each unit of change in the predictor in the raw-score regression equation is important practically and theoretically. The utility of the test in making decisions about groups or in forecasting social trends is the issue, and the utility judgment is not complicated by the costs of false-negative and false-positive errors in prediction. The regression estimate of a performance mean in good data is made with high accuracy in both senses of the use of that term—an almost zero standard error of estimate and a small sampling error.

Given high confidence in the gain in performance that can be obtained by a given increase in intelligence or a component ability, how does a democratic society deal with individual differences produced by the combined genetic and environmental determiners of intelligence? The problem may be in conflict with the egalitarian ideal but not with the basic democratic value that stresses individual worth. Acceptance of the reality of the problem is the first step in dealing with it democratically. The next step is to know its manifestations and to deal with them rationally.

Instability Over Generations

As discussed earlier, parents and children are only similar, never identical, with respect to traits that are determined by many genes and many environmental influences. Given correlations in the population between a selected parent and a selected child of less than unity, and typically a good deal less, regression in both directions from one generation to the other takes place. If the correlation is .50, regression is halfway back to the population mean. If the generational differences involve relatives less closely related genetically and environmentally, the resemblance coefficient is substantially reduced.

Persons with an emotional bias toward equality in outcomes are due to be forever disappointed with respect to individual differences within and between present social strata but need not be disappointed with respect to equality of outcomes of descendants two to three generations removed. Regression toward the population mean is progressive. The key is to maximize equality of opportunity in each generation. Genetics alone cannot produce a rigid class structure, but social inventions can.

Misconceptions About Heritability

In considering the social consequences of some degree of heritability of general intelligence, understanding the definition of the term is essential. Heritability is the proportion of genetic variance in the total variance of a trait in a defined population at a particular point in time. Heritability of intelligence can vary from one population to another and from one time period to another.

As long as there is a genetic component in general intelligence, the goal in a democratic society is to reduce environmental variance as much as possible. Increasing equality of opportunity for all citizens increases heritability. If the accurate figure were .80, as some have claimed, a democratic society could be proud of what it had achieved. If heritability in this country is lower today than it was a generation ago, as others have suggested, liberals have no cause for rejoicing. The probable explanation is that our class structure has become more rigid and that freedom of opportunity has decreased.

Mean levels do not enter the definition of heritability. Adding a constant amount to each person's intellectual repertoire by age 18 by strictly environmental means would not change the heritability coefficient. If we added a substantially larger amount to the intelligence of persons in the lower half of the distribution, heritability would increase.

Some conservatives have argued that a revolution in an existing aristocratic society would merely exchange one ruling class for another. Several comments are pertinent to this assertion. If the prerevolutionary society had had a rigid class structure, the new ruling class would be more able than the old. If the new rulers formed a rigid class structure in the generation of their grandchildren or great grandchildren, the level of genetic endowment in the ruling class would approach the society's average. It is doubtful that environmental variance can adequately compensate. It also follows that highly able persons have appeared in the lower classes. They are also frustrated by the shackles imposed on them by the class structure and are ready to lead a new revolution.

Perhaps the most dramatic evidence of genetic influence is typically overlooked. That it does not enter an estimate of heritability may be the explanation. The sheer number of children who exhibit high levels of talent and who are found in the most unlikely environmental circumstances is very impressive (Lubinski & Humphreys, 1990).

Differences Among Demographic Groups

As I have defined intelligence, there is no doubt about the reality of demographic differences. A democracy must deal with these differences. Value judgments can and should override at times the implications of quantitative data, but such judgments should not ignore those data. Even worse is the strong tendency among many policymakers and policy shapers to substitute myths for data. More and better data are available concerning Black–White comparisons, so I shall limit my discussion to these two groups.

An important empirical generalization is known and widely accepted among persons who read the research literature. When Blacks and Whites compete directly with each other and are evaluated by the same standard in education, industry, and the military services, Blacks do not perform quite as well generally as the estimate from their supposedly biased scores on valid predicter tests indicates (Linn, 1982, Hartigan & Wigdor, 1989). However, the difference tends to be about the size expected from measurement error in the predictor. That is, there is typically not sufficient evidence for a second deficit over and beyond that measured on a well-developed and well-researched predictive measure of ability.

It follows from this empirical generalization that a group of Blacks whose mean on a valid predictor is substantially below that of a group of Whites will also have a mean on a performance criterion substantially below that of their White counterparts. Thus, use of a lower standard for the selection of Blacks in education and in occupations that are to a degree cognitively complex inevitably places Blacks at a competitive disadvantage. I have described use of a lower standard for selection as the strong form of affirmative action. In contrast, the original definition required active recruiting of qualified minorities and selecting a minority person when qualifications were equal.

Affirmative Action and Race Relations

It seems probable that strong forms of affirmative action are determinants, among others, of increasing tensions between Blacks and Whites. Over and beyond the violation of the widely accepted value of rewarding merit, the inability of Blacks, on average, to be competitive with Whites increases or even introduces White prejudice. That same inability is also recognized by Blacks themselves. After acquiring a belief derived from educators, politicians, journalists, and many scientists that they are equal in fundamental capacity, Blacks look for reasons in their present educational or occupational situation for their criterion performance problems. There must be something that prevents them from performing to their full potential, and that something, one way or another, is White.

Our military services practice affirmative action as originally defined. Selection and assignment to specialties require Blacks and Whites to meet the same test standards, and promotion is as objective as it well can be. There are differences in the proportions of Whites and Blacks among specialties, but competition between the races within assignments is more or less equalized by the setting of the same entrance standard for a given assignment, even though these standards vary from one assignment to another. A testable hypothesis is that there is less separatism and more integration in the military services than on university campuses.

The organization of the military services does contribute to the success in following the original definition of affirmative action. Each military assignment has its own career ladder leading to promotion to the same ranks as in other assignments, and pay depends on rank rather than on assignment. Creative industrial personnel policies and federal income tax policies would allow civilian industry to approximate the military success.

The important problem faced by a minority group is not a deficit on a predictor test but a deficit in proficiency on a socially important performance. It is as difficult to identify Black exceptions to the expectation based on test score as it is for Whites. Item bias is a trivial problem when the focus is on outcome. Discovery and use of tests that reduce the number of errors of prediction for Blacks would be desirable, but such tests would do the same for Whites. Present theory does not reject the possibility of an environmental solution to the problem, but it does strongly suggest that intervention should start early in pregnancy and be continuous thereafter. Entrance in the first grade is already late. Early intervention should focus on the acquisition of aural comprehension of language. Reading, writing, and arithmetic follow closely. Experiences that enhance spatial and mechanical comprehension also make important contributions to the development of general intelligence.

Individuals Are Important

There are, of course, many individual exceptions to trends based on means. The distributions of persons about their respective means are approximately equivalent from group to group, and mean differences on any phenotypic behavioral trait are not so large that a great deal of overlap between distributions does not occur. The variability from person to person in all groups demonstrates the error of substituting either a verbal stereotype or a quantitative mean for information about individuals. Variability about means demonstrates the need for the original definition of affirmative action. A search for talented persons must be both intensive and extensive.

Summation

I have defined the phenotypic behavioral trait of intelligence that is measured by standard tests of intelligence. Genetic and environmental substrates affect the acquisition, storage, and retrieval of the intercorrelated intellectual (cognitive) knowledge and of the skills that form a behavioral repertoire. Intercorrelations define a factor common to the elements of the repertoire that support designation of the trait as general intelligence.

With the addition to the definition of some assumptions about learning and development, a number of testable hypotheses can be derived. These are categorized into those that concern mean changes in intelligence in groups, changes in individual differences on the trait, and correlates with performances in various social roles.

Because intelligence is general, individual and group differences on the trait have widespread effects on the individual and on society. General intelligence is far from being of sole importance in human affairs, but it is one of a relatively small number of important human traits. The number of traits that have appreciable effects on behavior outside the room in which tests are given is far smaller than the numbers that some ability and personality theorists have discussed.

Because intelligence is important, many liberals have felt pressure to reject any importance of the genetic substrate in the formation of individual differences in the trait. They have added an assumption, frequently implicitly, that environmental determination allows quick and easy compensation for earlier environmental deficits. The definition and theory discussed herein reject the latter assumption, and the theory does not require a resolution of the relative contributions of nature and nurture to the development of individual differences in intelligence. A democratic society can deal with the reality of an unknown, probably substantial genetic contribution to variance. As a matter of fact, the goal should be to maximize that contribution by increasing environmental opportunities.

Notes

Writing of this target article was supported by the Department of Psychology, University of Illinois, Urbana–Champaign, and by the Pioneer Fund.

Lloyd G. Humphreys, Department of Psychology, University of Illinois, 603 East Daniel Street, Champaign, IL 61820.

References

Anderson, J. E. (1939). The limitations of infant and preschool tests in the measurement of intelligence. *Journal of Psychology, 3,* 351–379.

Atkin, R., Bray, R., Division, M., Herzberger, S., Humphreys, L., & Selzer, U. (1977). Ability factor differentiation, Grades 5 through 11. *Applied Psychological Measurement, 1,* 65–76.

Cardon, L. R., Fulker, D. W., DeFries, J. C., & Plomin, R. (1992). Multivariate genetic analysis of specific cognitive abilities in the Colorado Adoption Project at age 7. *Intelligence, 16,* 383–400.

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

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin, 101,* 271–291.

Guttman, L. (1954). A new approach to factor analysis: The radex. In P. F. Lazarsfeld (Ed.) *Mathematical thinking in the social sciences* (pp. 258–348). Glencoe, IL: Free Press.

Harnquist, K. (1968). Relative changes in intelligence from 13 to 18. *Scandinavian Journal of Psychology, 9,* 50–82.

Hartigan, J., & Wigdor, S. (1989). *Fairness in employment testing: Validity generalization, minority issues, and the General Aptitude Test Battery.* Washington, DC: National Academy Press.

Horn, J. (1989). Models of intelligence: In R. L. Linn (Ed.) *Intelligence: Measurement, theory and public policy* (pp. 29–73). Urbana: University of Illinois Press.

Hulin, C. L. Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictor relationships. *Psychological Bulletin, 107,* 328–340.

Humphreys, L. G. (1971). Theory of intelligence. In R. Cancro (Ed.) *Intelligence: Genetic and environmental influences* (pp. 31–42). New York: Grune & Stratton.

Humphreys, L. G. (1979). The construct of general intelligence. *Intelligence, 3,* 105–120.

Humphreys, L. G. (1981). The primary mental ability. In M. P. Friedman, J. P. Das, & N. O'Connor (Eds.), *Intelligence and learning* (pp. 87–102). New York: Plenum.

Humphreys, L. G. (1982). The hierarchical factor model and general intelligence. In N. Hirschberg & L. G. Humphreys (Eds.) *Multivariate applications in the social sciences* (pp. 223–240). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Humphreys, L. G. (1985). General intelligence: An integration of factor, test, and simplex theory. In B. B. Wolman (Ed.), *Handbook of intelligence: Theories, measurement, and applications* (pp. 201–224). New York: Wiley.

Humphreys, L. G. (1991). Some unconventional analyses of resemblance coefficients for male and female monozygotic and dizygotic twins. In W. M. Grove & D. Cicchetti (Eds.), *Personality and psychopathology* (pp. 158–197). Minneapolis: University of Minnesota Press.

Humphreys, L. G., & Davey, T. C. (1988). Continuity in intellectual growth from 12 months to 9 years. *Intelligence, 12,* 183–197.

Humphreys, L. G., Park, R. K., & Parsons, C. K. (1979). Application of a simplex process model to six years of cognitive development in four demographic groups. *Applied Psychological Measurement, 3,* 51–64.

Humphreys, L. G., & Parsons, C. K. (1979). A simplex process model for describing differences among cross-lagged correlations. *Psychological Bulletin, 86,* 325–334.

Hunter, J. E. (1980). *Test validation for 12,000 jobs: An application of synthetic validity and validity generalization to the General Aptitude Test Battery (GATB).* Washington, DC: U.S. Employment Service.

Jensen, A. R. (1977). Cumulative deficit in IQ of Blacks in the rural South. *Developmental Psychology, 13,* 184–191.

Linn, R. (1982). Ability testing: Individual differences, prediction, and differential prediction. In A. K. Wigdor & W. R. Garner (Eds.), *Ability testing: Uses, consequences, and controversies. Part II: Documentation section* (pp. 335–388). Washington, DC: National Academy Press.

Lubinski, D., & Humphreys, L. G. (1990). A broadly based analysis of mathematical giftedness. *Intelligence, 14,* 327–355.

Lubinski, D., & Humphreys, L. G. (1992). Some bodily and medical correlates of mathematical giftedness and commensurate levels of socioeconomic status. *Intelligence, 16,* 99–115.

Lubinski, D., & Humphreys, L. G. (1993). *Seeing the forest from the trees: Compute correlations among means.* Unpublished manuscript.

Marshalek, B., Lohman, D. F., & Snow, R. E. (1983). The complexity continuum in the radex and hierarchical models of intelligence. *Intelligence, 7,* 107–127.

Roff, M. (1941). A statistical study of the development of intelligence test performance. *Journal of Psychology, 11,* 371–386.

Schmidt, F. L., & Hunter, J. E. (1977). Development of a general solution to the problem of validity generalization. *Journal of Applied Psychology, 62,* 529–540.

Tuddenham, R. D. (1948). Soldier intelligence in World Wars I and II. *American Psychologist, 3,* 54–56.