Psychological Reports, 2010, 106, 3, 643-664.

IQ VARIATIONS ACROSS TIME, RACE, AND NATIONALITY: AN ARTIFACT OF DIFFERENCES IN LITERACY SKILLS[1, 2]

DAVID F. MARKS

London

Summary.—A body of data on IQ collected over 50 years has revealed that average population IQ varies across time, race, and nationality. An explanation for these differences may be that intelligence test performance requires literacy skills not present in all people to the same extent. In eight analyses, population mean full scale IQ and literacy scores yielded correlations ranging from .79 to .99. In cohort studies, significantly larger improvements in IQ occurred in the lower half of the IQ distribution, affecting the distribution variance and skewness in the predicted manner. In addition, three Verbal subscales on the WAIS show the largest Flynn effect sizes and all four Verbal subscales are among those showing the highest racial IQ differences. This pattern of findings supports the hypothesis that both secular and racial differences in intelligence test scores have an environmental explanation: secular and racial differences in IQ are an artifact of variation in literacy skills. These findings suggest that racial IQ distributions will converge if opportunities are equalized for different population groups to achieve the same high level of literacy skills. Social justice requires more effective implementation of policies and programs designed to eliminate inequities in IQ and literacy.

Since the early research of Spearman (1904, 1927), intelligence test performance has been widely assumed to reflect a single intelligence factor, *g*, which underlies the performance of all cognitive tests and tasks. It has been suggested that *g* is a fundamental property of brain size (Jensen, 1998; Vernon, Wickett, Bazana, & Stelmack, 2000; Rushton & Rushton, 2003) or plasticity (Shaw, Greenstein, Lerch, Clasen, Lenroot, Gogtay, *et al.*, 2006). Other theorists have proposed multiple kinds of intelligence. Gardner proposed seven types of intelligence, namely, logical, linguistic, spatial, musical, bodily–kinesthetic, intrapersonal, and interpersonal (Gardner, 1983), then added an eighth, "naturalistic intelligence," and was considering a ninth, "existential intelligence" (Gardner, 1999). Sternberg (1980) proposed three kinds: analytic, creative, and practical intelligence. More recently, Sternberg (1999) proposed a definition of successful intelligence as: "(1) the ability to achieve one's goals in life, given one's sociocultural context; (2) by capitalizing on strengths and correcting or compensating for weaknesses; (3) in order to adapt to, shape, and se-

[1]Address correspondence to David F. Marks, Ph.D., 12a Stanhope Road, London N6 5DB, UK or e-mail (david_francis_marks@yahoo.com).
[2]The author thanks Emee V. G. Estacio for her assistance in locating datasets, and Conor V. Dolan, Brian Evans, James R. Flynn, and anonymous reviewers for their comments on earlier versions of this paper.

DOI 10.2466/PR0.106.3.643-664 ISSN 0033-2941

lect environments; and, (4) through a combination of analytical, creative, and practical abilities" (p. 189). Theories of multiple intelligences imply that the measurement of intelligence must go well beyond traditional IQ tests. It is the contention of this paper that literacy ability is an important "strength," enabling an individual to demonstrate intelligence using a "combination of analytical, creative, and practical abilities." An intelligence test is only valid for those individuals who have sufficiently high literacy to access the information and skills needed to complete the test. If, and only if, an individual has the ability to read and comprehend both the instructions and the item content of an intelligence test can that individual respond appropriately to the items in the test.

When intelligence tests are factor analyzed, different factors can be extracted (e.g., verbal meaning, verbal fluency, reasoning, number, space, memory, and perceptual speed). However, these factors overlap with one another so that what is common to all of them accounts for most of the variance in test scores. This general factor is called "*g*." The *g* factor is manifested in tests to the extent that they involve mental manipulation of the input elements. Any task involving conscious mental effort is *g* loaded and the more complex a task, the higher the *g* loading tends to be. Thus, the more complexity an individual can deal with, the higher will be his IQ on an intelligence test.

A huge body of data collected during the last 50 years has revealed that the population distribution of intelligence test scores shows significant variations over time, place, and race. Three well-established phenomena of intelligence tests remain unexplained. Firstly, during the period 1950–2000, in all countries studied to date, large increases in IQ occurred over time, known as the "Flynn effect" (Flynn, 1984, 1987; Neisser, 1998). The IQ gain has been approximately three points per decade. Yet there is no real evidence that people were becoming more intelligent. A few tentative but unverified explanations for the IQ gains have been offered. Mingroni (2004) argued that the Flynn effect is explained by "heterosis," a genetic effect that results from matings between members of genetically distinct subpopulations, owing to urbanization and greater population. Flynn (2007) argued that the secular IQ gains were a consequence of recent cultural learning experiences of applying logic on a formal level, rather than a concrete level, through an increasingly sophisticated understanding of science. These two explanations mirror the tenet that intelligence has a principally hereditary or environmental foundation.

A second replicable, but unexplained, phenomenon is that different ethnic/racial groups within a country have typically shown large differences in mean IQ. For several decades in the United States, the mean IQ of the African-American population had been averaging about one standard

deviation below the average for the Euro-American population (Herrnstein & Murray, 1994). Recent analyses of standardization samples suggest that this gap had decreased five to six points from 1972 to 2002 (Dickens & Flynn, 2006). On the other hand, Rushton and Jensen (2006) have rebutted this claim and Murray (2006) has argued that the IQ gap between these populations has remained stable.

A third consistent empirical finding is the observation of substantial national differences in average IQ, with sub-Saharan African countries showing a significantly lower average IQ (Lynn & Vanhanen, 2002, 2006; Lynn, 2006, 2008). This, along with racial differences in IQ, had been interpreted as supporting the view proposed in the 1850s that intellectual and brain capacity vary across race with Black people having lesser capacity than Whites (Agassiz, 1850; Nott & Gliddon, 1854; Jensen, 1998; Vernon, *et al*., 2000; Rushton & Rushton, 2003; Eberhardt, 2005; Rushton & Jensen, 2005; Lynn & Vanhanen, 2006).

These three empirical phenomena are not mutually exclusive, as they occur in combination, but, until now, no single explanation for them has been available.

The majority of relevant data have been collected using the Wechsler Adult Intelligence Scale (WAIS; Wechsler, 1955) and the Wechsler Intelligence Scale for Children (WISC; Wechsler, 1949). The WAIS is a general test of adult intelligence, while the WISC is an intelligence test for children between the ages of 6 and 16 years. In both cases, the test has been altered by the designer in light of results obtained, and it is essential that comparisons over time and race compare "like with like." IQ scores can be computed using all of the responses to the test, the Full Scale IQ, or the responses to particular subscales. Ten subtests of the WISC and the WAIS have been used by Flynn (2007) and others to analyze secular changes which have varying effect sizes according to the type of test. There are Verbal IQ subtests (Information, Vocabulary, Comprehension, and Similarities), and subtests for Perception (Picture Completion, Block Design, and Picture Arrangement), Working Memory (Arithmetic), and Processing Speed (Object Assembly and Coding). Flynn (2007, Table 1) presented data for the various versions of the WISC over a 54-year period that showed an IQ gain ranging from only 0.14 standard deviation for the Information subscale to 1.59 standard deviation for the Similarities subscale. The effect size for the White–Black Full Scale IQ gap is claimed to be in the region of 1.0 standard deviation (Herrnstein & Murray, 1994).

Research on secular and racial IQ differences has been mainly empirical in nature with little theoretical rationale. Recently, some theorists interested in a multifactorial approach to intelligence have integrated two key theories into a single structure. The first theory developed by Carroll (1993) is a hierarchical, three-stratum theory of cognitive abilities suggest-

ing three strata, with each layer attempting to account for the variation in factor loadings at the next lower level. The three strata represent general, broad, and narrow cognitive ability:

General (Stratum III): general intelligence factor.

Broad (Stratum II): fluid intelligence, crystallised intelligence, general memory and learning, broad visual perception, broad auditory perception, broad retrieval ability, and broad cognitive speediness and processing speed.

Narrow (Stratum I): 69 narrow abilities, each related to a specific Stratum II domain.

The second theory was proposed by Cattell and Horn concerning fluid and crystallized intelligence (Horn & Noll, 1997). Fluid intelligence (Gf) is the ability to draw inferences and understand the relationships between concepts, independent of acquired knowledge. Crystallized intelligence (Gc) is the ability to use skills, knowledge, and experience. It should not be equated with memory or knowledge, but it does rely on accessing information from long-term memory. The two approaches have been combined into the Cattell–Horn–Carroll (C–H–C) theory that aims to provide a comprehensive theory of the structure of human cognitive abilities, which can be applied to studies relevant to these topics (Flanagan & Ortiz, 2004; Alfonso, Flanagan, & Radwan, 2005; McGrew, 2005). Based on psychometric studies, the subscales of the WISC and WAIS have been mapped onto the 10 components of the C–H–C theory (McGrew & Flanagan, 1998). This mapping of IQ subscales onto the factors of the C–H–C theory is most helpful to the theory tested in this paper.

A recent study on the relationship between literacy skills and cognitive abilities suggests that there is a strong direct effect of *g* on basic reading skills at Kindergarten to Grade 3 and Grades 4 to 6 while, at Grades 7 to 12, *g* has a weak direct effect but a strong indirect effect on basic reading skills (Benson, 2008). The directionality of any causal relationship between *g* as a measure of general intelligence and literacy skills, however, remains uncertain.

It is clear that intelligence research contains some striking but unexplained phenomena and some highly contested theoretical formulations that raise questions about the nature of intelligence and intelligence testing. To date, theories offered to explain IQ variations have focused on environmental factors such as social class (Huang & Hauser, 1998) and education (Brody, 1992; Greenfield, 1998; Williams, 1998; Flynn, 2007) or biological theories including heredity (Jensen, 1998; Vernon, *et al.*, 2000; Lynn & Vanhanen, 2002; Rushton & Rushton, 2003; Rushton & Jensen, 2005), dysgenic factors (Herrnstein & Murray, 1994), and nutrition[3] (Lynn,

[3]The nutrition hypothesis suggests that the observed secular increases in IQ, like physical characteristics such as height, are caused by improvements in nutrition.

1998; Martorell, 1998; Sigman & Whaley, 1998; Colom, Lluis-Font, & Andrés-Pueyo, 2005).

Supporting the nutrition hypothesis, Colom, *et al.* (2005) argued that average IQ gains are caused by decreasing variance in the lower half of the distribution. The nutrition hypothesis predicts a primary effect on the most deprived, producing disproportionate gains in the portion of the population with low intelligence. Colom, *et al.* (2005) compared a sample of children tested in 1970 with an equivalent sample tested 30 years later for whom the entire IQ distributions were available for analysis. The gains were mainly concentrated in the lower and medium parts of the distribution and negligible in the highest part of the distribution. The authors stated, "an impressive gradual decrease in the gains was observed from the lower half to the top half of the distribution" (Colom, *et al.*, 2005, p. 83). Flynn (2009) argued against the nutrition hypothesis of Colom, *et al*. (2005) and Lynn (1998) on the grounds that "there has been no consistent tendency for the dietary gap between the classes to narrow since 1929" (p. 22), and he believes that the nutrition hypothesis receives no support from the analysis of British IQ gains. It is clear that no consensus has been reached, and that no theory is consistent with all of the available evidence.

Literacy Hypothesis for Intelligence Test Scores

This study is the first to analyze systematically the statistical association between literacy skills and IQ across time, nationality, and race. Successful performance on an intelligence test requires literacy skills not present in all people and populations to the same extent. At its most basic level, a person is literate if he or she can read, write, and understand a short simple statement about a situation or event drawn from everyday life (e.g., "The cat sat on the mat"). At more advanced levels, a literate person can write, read, and comprehend prose, documents, and quantitative expressions of high complexity. Many intelligence tests require high literacy because they entail the understanding and application of complex information to the solution of prose, numeracy problems, and abstract thinking tasks. A person who lacks literacy skills is ill-equipped to complete an intelligence test. Consequently, if the proportion of people with low literacy in a population is relatively high, the average test performance of the population will tend to be relatively low. If the average literacy skills in a population increase, as a consequence of improvements in education or literacy programs, then the average IQ of the population should also increase. However, the increased IQ will not be equally distributed through the entire population. IQ gains will be most evident in the lower half of the IQ distribution because this is the population sector that previously would have obtained relatively low scores as a consequence of their inability to comprehend the intelligence test's instructions.

In this study, ecological methods were used in which statistical associations are computed using data from samples that are representative of whole populations rather than individuals. For example, analyses are presented concerning the association between IQ and literacy rates across different countries. Ecological methods are employed in intelligence studies focused on correlating IQ differences measured across time, race, or place with theoretically important measures such as economic prosperity, health, or educational achievement (e.g., Hunt & Wittmann, 2006; Rindermann, 2007a; Gelade, 2008; Lynn & Mikk, 2009). It must be acknowledged from the outset that the ecological approach is not sufficient for a definitive test of the predictions, but it is the approach that can be used with currently available data. One must also observe the relationships in individuals in longitudinal studies of interventions. Care is certainly needed in interpretation because there is potential for confounding effects of age and sex or biases that can occur if test scoring criteria differ from one place to another or among tests. However, the method is valuable, and, in fact, indispensable, for investigating questions when controlled studies are not feasible for ethical, political, or logistical reasons.

Studies by Rindermann (2007b) on national group means are consistent with the hypothesis that "student achievement assessments and intelligence tests primarily measure a common cognitive ability *at the macro-social level.* This ability consists of the ability to think (intelligence) and of knowledge (extent of true and relevant knowledge, the ability to acquire and use knowledge)" (p. 667). Rindermann (2007b) reported moderate to high correlations between skill measures and intelligence test scores (r = .60 to .86) and interpreted these correlations as empirical verification of Spearman's *g*. An alternative interpretation is that intelligence test scores are measures of, or are highly associated with, the testees' literacy skills.

There are an estimated 774 million illiterate adults in the world, about 64% of whom are women.[4] It is commonly assumed that the literacy skills of people in developed countries are close to the maximum of 100%. However, in the 2003 National Assessment of Adult Literacy measuring the English skills of individuals from the United States aged 16 and older living in households and prisons (Kirsch, Jungeblut, Jenkins, & Kolstad, 1993; National Assessment of Adult Literacy, 2004), 14% of adults had below-basic literacy, and therefore possessed insufficient literacy and numeracy skills to participate fully in employment, legal, health care, and retirement systems. The survey found that 5% of the adult population in the United States were completely nonliterate. Similar to intelligence test scores, literacy skills vary according to socioeconomic status, sex, race, and occupation (Perie & Moran, 2005).

[4]UNESCO Institute for Statistics. (2007) Accessed at http://stats.uis.unesco.org/unesco/TableViewer/tableView.aspx?ReportId=201.

Self-reported literacy skills of both White and Black populations of the United States have increased steadily since 1870 (National Center for Education Statistics, 1993). National surveys using more objective measures have shown that, between 1992 and 2003, the gap in reading scores between Black and White 17-year-olds decreased by 11% (National Assessment of Adult Literacy, 2004). Some (but not all) recent analyses suggest that the gap between Black and White Full Scale IQ also has been narrowing progressively (Dickens & Flynn, 2006).[5] Fig. 4 of Dickens and Flynn's scale (2006) is interesting because it shows that the IQ and literacy scores of Blacks were increasing precisely in parallel over the period 1980–2000. According to the literacy hypothesis, the two upward trends are related: improved literacy skills in a population enable the population's average IQ to increase and its variance to decrease. The literacy hypothesis generates two predictions: (1) changes in differences on Full Scale IQ will be unevenly distributed, with greater increases in the lower half of the IQ distribution; (2) patterning of differences will be similar when comparing Verbal and Performance scales. Flynn's (2007) hypothesis, that cognitive stimulation is causing the IQ gains over time, predicts gains along the entire intelligence distribution of the population. However, if literacy improvements are strongly associated with the observed gains in mean IQ, as hypothesized, then one would expect to observe three phenomena within the distribution of Full Scale IQ. First, the standard deviation of the IQ distribution should become smaller over a period of time as the average IQ is progressively increasing, specifically because the literacy skills of people previously nonliterate or with low literacy skills improve, so the standard deviation in the IQ distribution decreases. Second, larger IQ gains should occur in the lower half and not the upper half of the IQ distribution. Third, the skewness of the distribution should change from positive to negative as a consequence of declining numbers of people with relatively poor literacy skills in the lower half of the IQ distribution.

The literacy hypothesis implies that secular IQ gains and racial IQ differences are not disparate phenomena but have a common origin, the acquired ability of people to comprehend and apply the verbal instructions in intelligence tests. Another independent test of the hypothesis is at the subscale level related to the fact that the pattern of effect sizes across different intelligence test subscales should show similarities between data sets showing the Flynn effect and White/Black IQ differences. The literacy hypothesis predicts that secular and racial differences will both be maximized on verbal subscales of intelligence tests while other subscales are predicted to show minimal differences or no orderly patterning. If a simi-

[5]One of the tests used in Dickens and Flynn's (2006) analysis was the Armed Forces Qualifications Test, which, as will become apparent later, is a measure of literacy, not IQ.

lar patterning is observed in secular and racial differences, this will have evidential value in support of the literacy hypothesis regarding the origin of the two phenomena, i.e., the hypothesis that both phenomena are associated with differences in literacy skills. The analysis reported here investigates these predictions using existing data on the different subscales of the WAIS and WISC.

Method

Three types of data analyses were carried out: (i) analyses of the relationship between mean Full Scale IQ and literacy skills in different cohorts and populations; (ii) analyses of the standard deviations and skewness of Full Scale IQ test score distribution in different cohorts; (iii) analyses of the similarity in the pattern of effects across verbal and nonverbal subscales of intelligence tests in both Flynn effect and White–Black datasets.

Procedures

The analyses were carried out on data from a variety of existing sources. In the majority of cases the original IQ data were derived from reports of national screening programs carried out by military or educational services. Literacy skill data were obtained from a variety of published estimates. All IQ or ability data were obtained from primary sources and were independent of the sources used for the literacy data. All data sources are fully within the public domain. For the majority of analyses, IQ and literacy scores were regressed against time and then against each other. In all cases, linear or quadratic functions provided the best fit. Whenever possible, descriptive statistics (means, standard deviations, skewness) derived from the original sources of raw data were employed. In a few cases, literacy and IQ data were available over the same period of time but for different combinations of years. In such cases, linear interpolation was used to provide a pairing of scores for the same set of years to enable regression to be carried out. All statistical analyses were carried out using SPSS Version 12.0.0 (release 4 September 2003).

Results

Correlations Between Intelligence Test Scores and Literacy Skill Measures

IQ and literacy skills in the United States.—IQ and literacy data for the White population of the United States are available for the period 1932–1978 (Flynn, 1987; National Center for Education Statistics, 1993). IQ and literacy scores for Whites over this period were correlated at $r = .95$ ($r^2 = .91$; $F_{1,5} = 48.12$; $p = .001$; Fig. 1). IQ and literacy scores for the Black population of the United States were available for the period 1980–2000 (Dickens & Flynn, 2006). IQ and literacy scores for the Black population over the period 1980–2000 were correlated at $r = .90$ ($r^2 = .82$; $F_{1,3} = 13.87$; $p = .0337$). Camp-

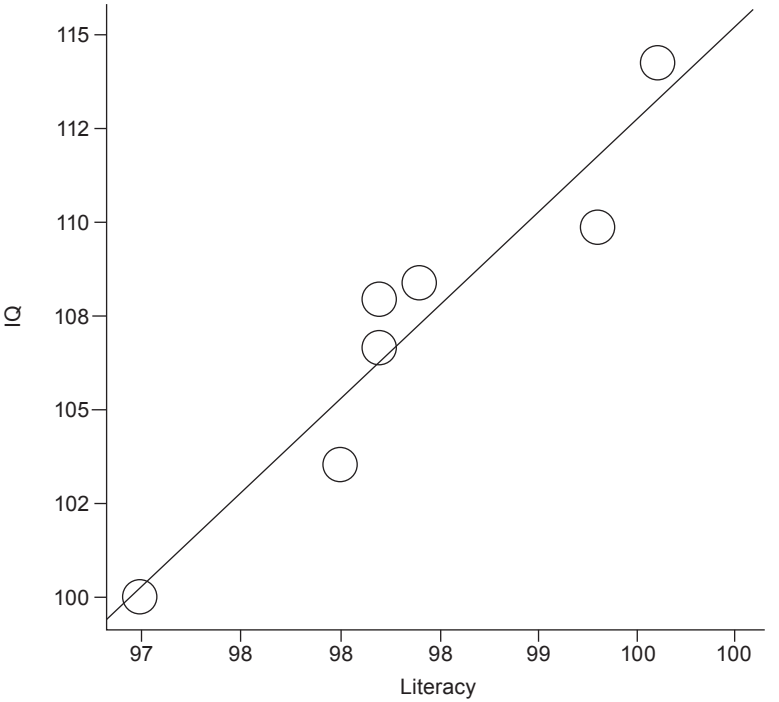

FIG. 1. Mean IQs against self-reported literacy for the White European-American population of the USA. The literacy data were obtained from the National Center for Education Statistics (1993) data on historical trends in illiteracy in the USA. This dataset gives the percentage of persons 14 years old and older who were illiterate (unable to read or write in any language) by race and nativity: 1870 to 1979. The IQ test scores for White Americans for the period 1932 to 1978 were derived from Table 7 of Flynn (1987).

bell, Hombo, and Mazzeo (2000) presented a report on the academic progress of students over three decades which included data on reading scores of White, Black, and Hispanic 17-yr.-olds. Using these data, the racial gap in reading scores of 17-yr.-olds was computed. The racial gap in mean IQ for the same years was computed from the data presented by Dickens and Flynn (2006). Regression of the White–Black IQ gap on the White–Black reading score gap yielded a correlation of $r = .79$ ($r^2 = .62$; $F_{1,6} = 9.73$; $p = .02$). The United States population data on IQ and literacy suggest that the secular changes in two measures are highly related. However, because this relationship is based only on two linearly increasing trends over time, this evidence must be supported by other kinds of data. In fact, these analyses suggest that a strong association between IQ and literacy skills is a general property of *all* IQ data.

IQ and literacy skills in Denmark and Norway.—Danish literacy skill es-

timates for the 17- to 25-year-old male population from 1960 to 1995 (Coulombe & Tremblay, 2006) were correlated with IQ data for Danish 18-year-old military draftees, from 1960 until 2004 (Teasdale & Owen, 2005). Literacy skills in Danish men increased from 1960 until 1995 when they reversed downward; mean IQs followed that trend. From 1960 until 1998, mean IQ among Danish military men showed linear increases and then, from 1998, mean IQ declined continuously until the data supply ended in 2004. Linear regression of mean IQ against time over the 35-yr. period 1960–1995 produced a correlation coefficient r of .99 ($r^2 = .98$; $F_{1,17} = 772.54$; $p < .0001$). Linear regression of average literacy scores for Danish males over the same period produced a correlation coefficient r of .78 ($r^2 = .60$; $F_{1,10} = 15.09$; $p = .003$). The relationship between the Danish military men's mean IQ and literacy skills was best fitted by a quadratic function with a correlation coefficient r of .98 ($r^2 = .96$; $F_{2,9} = 111.28$; $p < .0001$). This curvilinearity may have been caused by measurement error, the distribution asymmetries that are known to exist, or it could indicate that IQs are being driven by another factor in addition to literacy skills, a possibility which is supported by data presented below. It is also possible that there is a critical literacy level beyond which IQ is unaffected by literacy.

IQ data for 18- to 20-year-old Norwegian male conscripts were available for 1954 to 2002 (Sundet, Barlaug, & Torjussen, 2004) while literacy skills estimates for 17- to 25-year-old Norwegian males were available from 1960 to 1995 (Coulombe & Tremblay, 2006). The shape of the IQ and literacy functions over time showed striking similarities. There were two phases, an initial increase in both measures followed by a decrease. Linear regression of IQ against time over the entire 35-yr. period resulted in a correlation coefficient of .91 ($r^2 = .82$; $F_{1,6} = 27.56$; $p = .0019$), while the linear regression for literacy skills against time over the same period resulted in an r value of .85 ($r^2 = .72$; $F_{1,6} = 15.45$; $p = .0077$). Linear regression of IQ against literacy over this period yielded an r value of .98 ($r^2 = .96$; $F_{1,6} = 164.89$; $p < .0001$).

Cross-sectional analyses of IQ and literacy skills.—This analysis investigated the relationship between national average IQ (Lynn & Vanhanen, 2002) and average literacy scores for 81 countries in the time period 1995–2000 (United Nations Development Program, 2006). Best fit to the data was obtained with a quadratic function yielding $r = .83$ ($r^2 = .69$; $F_{2,83} = 94.22$; $p < .0001$; Fig. 2).

Changes in National Intelligence Test Score Distributions Associated With Improvements in Literacy Skills

Data from a variety of sources have confirmed all three predictions, although the findings are not always consistent. In the Norwegian study, the mean increase in IQ below the median from 1957–1959 to 1993–2002

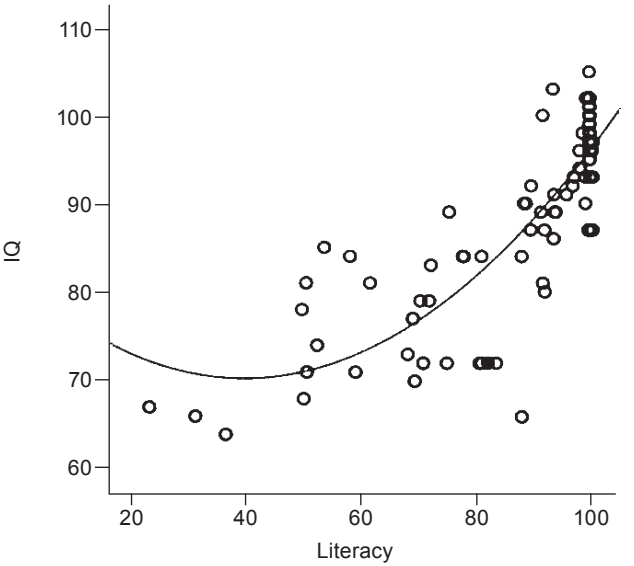

FIG. 2. Mean IQs against literacy scores for 81 countries. The literacy dataset was obtained from the United Nations Human Development Report (2006). These scores are provided by national governments and international bodies, the results of surveys and statistical analyses based on large samples. The IQ data were extracted from Lynn and Vanhanen (2002). Only the 81 countries from this dataset where intelligence test scores had been measured, not estimated, are included.

cohorts was about three times the increase in IQ for the above median scores (Sundet, *et al.*, 2004). As the mean score on a measure of word similarities increased from 1957, when the mean score was 27.26, to 1993, at 33.48, the standard deviation of the intelligence test score distribution decreased by 26% from 11.95 to 9.08, a statistically significant change ($F_{14749,16452} = 1.73$; $p < .000001$). Over this period the skewness of the distribution also changed from +.06 to −0.37 as predicted, which was significant using Taylor's (1947) significance test for the difference in skewness of two independent distributions ($p < .01$).

In the Danish study, the standard deviation of the ability test score distribution decreased by 22% as the mean test score increased to its maximum over the period 1959–1998 (Teasdale & Owen, 2005), a statistically significant change. However, IQ increases at the bottom end of the distribution were not observed on all subscales. As predicted, skewness shifted from +0.05 in 1959, when the mean ability score was 4.67, to −0.49 in 1998 when the mean reached its maximum of 5.94. This difference was significant using Taylor's (1947) significance test for the difference in skewness of two independent distributions ($p < .01$).

As predicted, in both populations: (i) mean IQ gains were associated

with a much-reduced proportion of low IQ scorers; (ii) modest increases only occurred among high IQ scorers; (iii) there was a statistically significant lowering of IQ variance; and (iv) there was a statistically significant increase in negative skewness.

Some of the strongest evidence that IQ improvements were lowest at the top end and highest in the middle and lower parts of the IQ distribution is apparent in the data presented by Dickens and Flynn (2006; see their Fig. 1). The data in the figure indicate that the lowest IQ gains are in the top 15% of the distribution while the highest gains are in the bottom 2%, providing evidence of the asymmetry predicted by the literacy hypothesis. Similar trends are observable in the WISC data presented by Flynn (2007; Fig. 5, p. 137) for the period 1950–1970, but not for 1970–2000. In a recent analysis of British IQ gains on the Standard Progressive Matrices from 1938 to 2008, for all ages from 9.5 to 15.5 yr., the top half showed a significantly lower gain than the bottom half (Flynn, 2009). Data from the Army Alpha and Beta tests also showed that the more difficult tests had a prominent mode at zero and positive skewness while easier tests showed a negative skew (Yerkes, 1921, cited by Gould, 1981).

Similar Patterning in Flynn Effect and White–Black Differences across Subscales Indicates a Similar Origin

It has been acknowledged for some time that secular and racial IQ differences vary in magnitude for different subscales and kinds of test. Full Scale IQ provides only a crude summary of the overall performance differences that occur between cohorts or populations. The literacy hypothesis predicts that verbal subscales will show larger Flynn effect sizes and racial IQ differences than do the nonverbal subscales. Flynn (2007, Fig. 1 and Table 1) presents data on United States WISC IQ gains from 1947 to 2002. The Similarities subtest showed an increase of 24 points while the Vocabulary subtest gained four points and Information and Arithmetic only two points. Table 1 shows Flynn effect sizes for individual subscales of the WISC–IV which have been mapped onto the Gf–Gc configuration proposed by the C–H–C theory of cognitive abilities. The Similarities and Vocabulary subscales are measures of the narrow abilities of Language Development and Lexical Knowledge within the domain of crystallized intelligence (Gc). Similarities produced the largest Flynn effect size, 1.59 standard deviation units, over the period 1947 to 2001, providing support for the literacy theory. Another test of Gc, Information, produced the low effect size of 0.09 but, according to the C–H–C theory, this subtest represents General Information, not Language Development or Lexical Knowledge, and so the low Flynn effect size is not an issue for the literacy hypothesis. WISC measures of Perceptual Reasoning, Working Memory, and Processing Speed produced Flynn effect sizes over a wide range (0.07–

TABLE 1

Flynn Effect Sizes and White–Black Differences For Subtests of the WISC–IV Measuring Broad C–H–C Abilities of the Gf-Gc Configuration

Broad C–H–C Abilities	WISC Subtest	Narrow Abilities	Flynn Effect 1947–2001 *SD* Units*
Crystallized intelligence, Gc	Similarities	Language development Lexical knowledge	1.59
	Vocabulary	Language development Lexical knowledge	0.29
	Comprehension	Language development General information	0.73
	Information	General information	0.09
Quantitative knowledge, Gq	Arithmetic	Mathematical achievement Quantitative reasoning	0.77
Visual processing, Gv	Block Design	Spatial relations Visualization	1.06
	Picture Completion	Flexibility of closure General information	0.78
	Picture Arrangement	General information Induction	1.43
	Object Assembly	Closure speed	1.16
Short-term memory, Gsm	Digit Span†	Memory span	0.07
Processing speed, Gs	Coding	Rate of test taking Perceptual speed	1.20

*Source: Flynn (2007), Table 1, pp. 180-181. †The data for this subtest are for the period 1972–2002.

1.20 standard deviations) with the lowest effect sizes in Working Memory as predicted. The pattern of effects in U.S. IQ data collected over a 50-year period shown in Table 1 provides strong support for the literacy theory but also suggests that Visual Processing (Gv) and Processing Speed (Gs) have also contributed to the Flynn effect.

Another relevant source of data is provided by the study by Wicherts, Dolan, Hessen, Oosterveld, van Baal, Boonsma, *et al.* (2004) which investigated the effect sizes for Flynn effect in five different intelligence tests: the Dutch WAIS for cohorts in 1967/8 and 1998/9 (four verbal and seven nonverbal subscales); Børn Prien's Prøve for cohorts of Danish draftees in 1988 and 1998 (two verbal and two nonverbal subscales); the Dutch DAT for Dutch high school students in 1984 and 1994/5 (four verbal and three nonverbal subscales); the Revised Amsterdam Kinder Intelligence Test for Dutch children 1981/2 and 1992/3 (three verbal and three nonverbal subscales in two different datasets); and the National Intelligence Test with Estonian children in 1934/6 and 1997/8 (five verbal and five nonverbal subscales). In total, there were 38 subscales within these five intelligence tests. However, for the Revised Amsterdam Kinder Intelligence Test, effect sizes were available for two different samples, a standardization sample and a

sample of twins, giving a total of 44 subscale effect sizes for this analysis, of which 22 were verbal subscales and 22 were nonverbal subscales. The subscales within each cohort study were ranked and, for convenience, dichotomized, according to their Flynn effect sizes, as reported by Wicherts, *et al.* (2004), and categorized as "above the median" or "at or below the median." Fourteen of 22 verbal subscales had an above-median effect size, while 15 of 22 nonverbal subscales had at, or below, median effect sizes ($\chi^2 = 3.279$, $p < .05$, one-tailed), consistent with the literacy hypothesis.

The next analysis explored whether the pattern of secular IQ gains is replicated in racial IQ differences. Jensen (1998, p. 386) presented a figure showing the relationship between the mean White–Black differences in IQ and the *g* loading for subtests of the WISC–R and K-ABC tests. The subtests that showed large, modest, or small racial IQ differences were analyzed in similar fashion to the Flynn effect data. As predicted, verbal subscales contributed disproportionately to racial IQ differences in comparison to the nonverbal subscales, with all four verbal subscales at or above the median, and five of seven nonverbal subscales below the median. A Fisher Exact Test for these data yielded $p = .0225$.

In the next analysis the pattern of White–Black differences in subscale scores is directly compared to the pattern of differences obtained in a typical Flynn effect study. Jensen's data (1998, p. 386) on White–Black differences for 11 subscales of the WISC are cross-tabulated with the Flynn effect sizes computed by Wicherts, *et al.* (2004) for 11 subscales of the Dutch WAIS from cohorts in 1967–1968 and 1998–1999. The 11 subscale scores of the WISC and the WAIS are loaded on three factors: verbal comprehension (V), memory (M), and perceptual/organizational (P). The effect sizes for these 11 scales from Jensen's (1998) study of racial IQ differences and Wicherts, *et al*. (2004) study of secular IQ gains were independently ranked from 1 to 11. In Table 2, the 11 subtest scores for the two datasets are cross-tabulated to enable the association between the two independent rankings to be computed using a chi-squared statistic.

In Table 2, Cell A contains four subscales with above-median effect sizes for both the White–Black and Flynn effect datasets. The literacy theory predicts that Cell A should be occupied by subscales that represent verbal ability. In fact, as predicted, three of the four Verbal subscales (Similarities, Comprehension, and Vocabulary) are in Cell A, representing the largest effects in both secular and racial IQ differences. Cell D, on the other hand, contains four subscales with below-median values for both Flynn effect and White–Black differences. The literacy theory predicts that subscales that represent nonverbal abilities should occupy Cell D. In fact, as predicted, all three of the Memory/Indistractibility subscales (Picture Arrangement, Digit Symbol, and Digit Span) are members of this

TABLE 2

Cross-tabulation For 11 Subscales According to Magnitude of Flynn Effect Sizes For Dutch WAIS (Wicherts, *et al.*, 2004) and United States White–Black Differences in WISC (Jensen, 1998)

Measure	Size of the Flynn Effect on the Dutch WAIS (Wicherts, *et al.*, 2004)		Total
	Above the Median Effect Size	At or Below the Median Effect Size	
White–Black differences in mean IQ on the WISC (Jensen, 1998)			
At or above the median	CELL A *Similarities (V)* *Comprehension (V)* *Vocabulary (V)* Block Design (P)	CELL B Object Assembly (P) *Information (V)*	
	4	2	6
Below the median	CELL C Picture Completion (P)	CELL D *Arithmetic (M)* Picture Arrangement (P) *Digit Span (M)* *Digit Symbol (P & M)*	
	1	4	5
Total	5	6	11

Note.—Factor For each subscale is indicated in parentheses: Verbal Comprehension (V), Perceptual/Organizational (P), and Memory/Indistractibility (M).

cell, although the Fisher Exact Test was not significant ($p = .069$). That three Verbal subscales are located in the predicted cell (A) and three Memory subscales are in the diagonally opposite cell (D) has an exact probability of .00024.[6] The patterning of findings in Tables 1 and 2 is fully consistent. The data provide strong support for the literacy hypothesis of IQ: secular and racial IQ differences are both an artifact of literacy, language, and lexical skill variations over time and race.

Misinterpretations of Intelligence Test Scores

Herrnstein and Murray's (1994) claims about IQ differences, ethnicity, and social issues rested on data obtained from the United States Department of Labor's National Longitudinal Survey of Youth. The National Longitudinal Survey of Youth data included an aptitude test developed by the Department of Defense to measure the ability of potential recruits to learn to perform military jobs, the Armed Forces Qualifications Test. The Armed Forces Qualifications Test was designed as a test of educability, that is, how well high school graduates would do in the armed forces (Fischer, Hout, Jankowski, Lucas, Swidler, & Voss, 1996). To evaluate the

[6]No prediction was made concerning the Perceptual/Organizational (P) subscales which were not included in the analysis.

nature of the Armed Forces Qualifications Test in more detail, the literature was searched for datasets containing test estimates for populations or groups for which literacy scores could also be ascertained. Two Armed Forces Qualifications Test datasets were identified. The first was generated in a study to assess functional literacy required in military jobs (Sticht, Caylor, Kern, & Fox, 1972). The authors studied men occupying three different jobs (supply specialists, repairmen, and cooks) and obtained data on the relationship between their Armed Forces Qualifications Test scores and reading ability on the Survey of Reading Achievement, Junior High Level, California Test Bureau (Tiegs & Clark, 1959). The three occupational groups had been pre-selected for their reading ability (high, medium, or low) producing a total of nine independent groups, ranging in size from nine to 34, covering the entire range of reading ability. The data were fitted by a linear function with an r of .96 ($r^2 = .96$; $F_{2,6} = 38.2$; $p = .0004$). The data show that the Armed Forces Qualifications Test scores of the nine groups of soldiers could be almost perfectly predicted from their literacy scores.

Kilburn, Hanser, and Klerman (1998) compared the percentage scores on National Longitudinal Survey of Youth (1980) and National Educational Longitudinal Study (1992) for high school seniors in each Armed Forces Qualifications Test category for the Black, Hispanic, and White populations. The total percentage scores for the top four categories for 1980 and 1992 were computed for the three ethnic groups (see Table 5.2 of Kilburn, *et al*., 1998, p. 27). The Black and Hispanic populations both showed significant improvements in their Armed Forces Qualifications Test scores between 1980 and 1992 while the White population showed a slight decrement. Reading scores for 17-year-olds for the same ethnic groups and dates were obtained from the National Assessment of Academic Progress (see Fig. 2.1, Campbell, *et al.*, 2000, p. 33). The data were best fitted by a quadratic function giving an r value of .997 ($r^2 = .99$; $F_{2,3} = 250.7$; $p = .0005$). This correlation of .997 supports of the literacy theory since the six pairs of data points were from six independent population samples evaluated by two separate groups of investigators. On the basis of the studies summarized here, there can be little doubt that the Armed Forces Qualifications Test is a measure of literacy.

Discussion

The above findings lend support to the hypothesis that IQ differences across time, race, and nationality are associated with differences in literacy skills and not with genuine differences in intelligence. The evidence shows that, when the average literacy of a population is relatively high, so is the population's mean IQ (and vice versa). If literacy scores reach a plateau or begin to decline, mean IQs also plateau or begin to decline. Improvements in IQ are unequally distributed, with the greatest gains oc-

curring in the lower half of the IQ distribution. This finding follows from the fact that improvements in literacy skills would differentially increase IQ in the lower half than the upper half of the distribution. Flynn (2007) discussed but dismissed the possibility that the variance of IQ has diminished over time, yet some of the strongest evidence that IQ improvements were lowest at the top end and biggest in the middle and lower parts of the IQ distribution is apparent in Dickens and Flynn (2006; see their Fig. 1) and in Flynn (2009). However a similar prediction is made by the nutrition hypothesis and further analyses must examine whether the nutrition or literacy hypothesis provides the best explanation of these data. In light of his discussion of secular nutritional changes and their relationship to social class, Flynn (2009) dismisses the nutrition hypothesis as untenable. In addition, some predictions are specific only to the literacy hypothesis and are not predicted by the nutrition hypothesis. These concern the patterns of secular and racial differences, which strongly favor the literacy hypothesis over the nutrition hypothesis. The Flynn effect and racial IQ differences follow a similar pattern, with verbal subscales showing the largest effects, memory subscales the smallest effects, and perceptual subscales showing no consistent pattern. However, the results shown in Table 1 indicate that literacy is only one factor, and not the only factor, responsible for the Flynn effect. While Crystallized Intelligence (Gc) is a major contributor to the Flynn effect, Visual Processing (Gv) and Processing Speed (Gs) are also contributing to the Flynn effect.

These conclusions are supported by the studies reviewed in the Introduction showing that racial differences in test performance are eliminated when previously absent knowledge or skills are taught to the participants. The present conclusions on the role of verbal abilities confirm Bridgeman and Buttram's (1975) suggestion that a significant proportion of observed racial IQ differences may be attributable to different usage of verbal strategies rather than to reasoning deficits. In regards to the Flynn effect, support is provided by Nettelbeck and Wilson (2004) and Wicherts, *et al.* (2004) who showed that intelligence increases could not be assumed in the Flynn effect due to measurement problems across time. Both secular and racial IQ differences evidence larger changes in verbal than nonverbal subscales of intelligence.

These analyses in support of the literacy hypothesis are the first steps toward identification of a larger web of associations that may exist between literacy and intelligence. Further work using a variety of statistical techniques, including multigroup structural equation modeling, could be helpful in addressing the purported role of literacy in the covariance and mean structure of literacy and IQ tests. Such modeling would address the structure of individual differences, group differences, and the relation-

ship between the two. The study by Dolan, Colom, Abad, Wicherts, Hessen, and van de Sluis (2006) concerning the relationship between intelligence test scores and a measure of educational attainment in males and females provides a useful model for such research. Experimental studies are planned in which a literacy intervention will be applied within a group of young people who have relatively low literacy skills to assess the effect of enhancing literacy skills on IQs in comparison to a waiting-list control group.

These findings based on population scores cannot be extrapolated to individual test scores and correlation does not indicate causation. Before definite conclusions can be reached, evidence is needed from experimental studies with individual participants or studies using structural equation modeling. However, the consistent, replicable association between literacy scores and IQ distributions across different populations and the common patterning of subscale effects with secular and racial differences at least suggest the possibility of a causal relationship between literacy and intelligence test performance. This study has shown that population literacy changes with population IQ, but all that one can say unequivocally is that an as-yet-unidentified variable is causing them both to change. Only experimental studies on the effects of literacy interventions will enable the causality hypothesis to be properly tested.

The findings suggest that using intelligence test performance to make inferences about hereditary theories of human intelligence may be misguided. The racial theory of intelligence claims that the observed racial IQ differences are real genetic differences between Caucasian, Asian, African, African-American, and African-Caribbean people (Herrnstein & Murray, 1994; Jensen, 1998; Vernon, *et al.,* 2000; Lynn & Vanhanen, 2002; Rushton & Jensen, 2005). However, these claims must be considered suspect in light of the evidence that *a large proportion of both secular and racial differences may be produced by differences in literacy ability* via verbal subscales. The Flynn effect and racial IQ differences may have a similar environmental basis. In addition, in light of the present findings showing extremely high correlations ($r = .96–.99$) between the Armed Forces Qualifications Test and literacy scores, it must be concluded that Herrnstein and Murray's (1994) use of the Armed Forces Qualifications Test as a measure of "IQ," and the conclusions that followed, were in error.

REFERENCES

Agassiz, E. C. (1850) The diversity of origin of the human races. *Christian Examiner*, 49, 110-145.

Alfonso, V. C., Flanagan, D. P., & Radwan, S. (2005) The impact of the Cattell—Horn—Carroll theory on test development and interpretation of cognitive and academic abilities. In D. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues.* (2nd ed.) New York: Guilford. Pp. 185-202.

Benson, N. (2008) Cattell–Horn–Carroll cognitive abilities and reading achievement. *Journal of Psychoeducational Assessment*, 26, 27-41.

Bridgeman, B., & Buttram, J. (1975) Race differences on nonverbal analogy test performance as a function of verbal strategy training. *Journal of Educational Psychology*, 67, 586-590.

Brody, N. (1992) *Intelligence*. San Diego, CA: Academic Press.

Campbell, J. R., Hombo, C. M., & Mazzeo, J. (2000) *Trends in academic progress: three decades of student performance, NCES 2000-469*. Washington, DC: U.S. Department of Education, Office of Educational Research and Improvement, National Center for Education Statistics, NAEP 1999.

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

Colom, R., Lluis-Font, J. M., & Andrés-Pueyo, A. (2005) The generational intelligence gains are caused by decreasing variance in the lower half of the distribution: supporting evidence for the nutrition hypothesis. *Intelligence*, 33, 83-91.

Coulombe, S., & Tremblay, J. F. (2006) Literacy and growth. *Topics in Macroeconomics,* 6, 1-32.

Dickens, W. T., & Flynn, J. R. (2006) Black Americans reduce the racial IQ gap: evidence from standardization samples. *Psychological Science,* 17, 913-920.

Dolan, C. V., Colom, R., Abad, F. J., Wicherts, J. M., Hessen, D. J., & van de Sluis, S. (2006) Multi-group covariance and mean structure modeling of the relationship between the WAIS–III common factors and sex and educational attainment in Spain. *Intelligence*, 34, 193-210.

Eberhardt, J. L. (2005) Imaging race. *American Psychologist*, 60, 181-190.

Fischer, C. S., Hout, M., Jankowski, M. J., Lucas, S. R., Swidler, A., & Voss, K. (1996) *Inequality by design: cracking the bell curve myth.* Princeton, NJ: Princeton Univer. Press.

Flanagan, D. P., & Ortiz, S. O. (2004) CHC theory: a taxonomy of cognitive abilities. In T. Steurt Watson & C. H. Skinner (Eds.), *Encyclopedia of school psychology*. New York: Kluwer Academic. Pp. 139-141.

Flynn, J. R. (1984) The mean IQ of Americans: massive gains 1932 to 1978. *Psychological Bulletin,* 95, 29-51.

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

Flynn, J. R. (2007) *What is intelligence?* Cambridge, UK: Cambridge Univer. Press.

Flynn, J. R. (2009) Requiem for nutrition as the cause of IQ gains: Raven's gains in Britain 1938 to 2008. *Economics and Human Biology,* 7, 18-27.

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

Gardner, H. (1999) *Intelligence reframed: multiple intelligences for the 21st century*. New York: Basic Books.

Gelade, G. A. (2008) The geography of IQ. *Intelligence,* 36, 495-501.

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

Greenfield, P. M. (1998) The cultural evolution of IQ. In U. Neisser (Ed.), *The rising curve: long-term gains in IQ and related measures.* Washington, DC: American Psychological Association. Pp. 81-123.

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

Horn, J. L., & Noll, J. (1997) Human cognitive capabilities: Gf–Gc theory. In D. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues*. New York: Guilford. Pp. 53-91.

Huang, M-H., & Hauser, R. M. (1998) Trends in Black–White test score differentials: II. The WORDSUM Vocabulary Test. In U. Neisser (Ed.), *The rising curve: long-term gains in IQ and related measures.* Washington, DC: American Psychological Association. Pp. 303-332.

Hunt, E., & Wittmann, W. (2006) National intelligence and national prosperity. *Intelligence*, 36, 1-9.

Jensen, A. R. (1998) *The* g *factor*. Westport, CT: Praeger.

Kilburn, M. R., Hanser, L. M., & Klerman, J. A. (1998) *Estimating AFQT scores for National Educational Longitudinal Study (NELS) respondents.* Santa Monica, CA: RAND Distribution Services.

Kirsch, I. S., Jungeblut, A., Jenkins, L., & Kolstad, A. (1993) *Adult literacy in America: a first look at the results of the National Adult Literacy Survey*. Princeton, NJ: Educational Testing Service.

Lynn, R. (1998) In support of nutrition theory. In U. Neisser (Ed.), *The rising curve: long-term gains in IQ and related measures.* Washington, DC: American Psychological Association. Pp. 207-215.

Lynn, R. (2006) *Race differences in intelligence: an evolutionary analysis*. Augusta, GA: Washington Summit.

Lynn, R. (2008) *The global bell curve*. Augusta, GA: Washington Summit.

Lynn, R., & Mikk, J. (2009) National IQs predict educational attainment in math, reading and science across 56 nations. *Intelligence*, 37, 305-310.

Lynn, R., & Vanhanen, T. (2002) *IQ and the wealth of nations*. Westport, CT: Praeger.

Lynn, R., & Vanhanen, T. (2006) *IQ and global inequality*. Augusta, GA: Washington Summit.

Martorell, R. (1998) Nutrition and the worldwide rise in IQs. In U. Neisser (Ed.), *The rising curve: long-term gains in IQ and related measures.* Washington, DC: American Psychological Association. Pp. 183-206.

McGrew, K. S. (2005) The Cattell–Horn–Carroll theory of cognitive abilities: past, present, and future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary intellectual assessment: theories, tests, and issues.* (2nd ed.) New York: Guilford. Pp. 136-182.

McGrew, K. S., & Flanagan, D. P. (1998) *The Intelligence Test Desk Reference (ITDR): Gf–Gc cross-battery assessment.* New York: The Guilford Press.

Mingroni, M. A. (2004) The secular rise in IQ: giving heterosis a closer look. *Intelligence,* 32, 65-83.

Murray, C. (2006) Changes over time in the Black–White difference on mental tests: evidence from the children of the 1979 cohort of the National Longitudinal Survey of Youth. *Intelligence*, 34, 527-540.

National Assessment of Adult Literacy. (2004) *A first look at the literacy of America's adults in the 21st century.* Washington, DC: U.S. Department of Education, Institute of Education Sciences, NCES 2006-470.

National Center for Education Statistics. (1993) *120 years of American education: a statistical portrait.* (T. Snyder, Ed.) Washington, DC: U.S. Department of Education, Institute of Education Sciences, NCES 1993.

Neisser, U. (1998) *The rising curve: long-term gains in IQ and related measures.* Washington, DC: American Psychological Association.

Nettelbeck, T., & Wilson, C. (2004) The Flynn effect: smarter not faster. *Intelligence,* 32, 85-93.

Nott, J. C., & Gliddon, J. R. (1854) *Types of mankind*. Philadelphia: Lippincott, Grambo.

Perie, M., & Moran, R. (2005) *NAEP 2004 trends in academic progress: three decades of student performance in reading and mathematics*. Washington, DC: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, NCES 2005-464.

Rindermann, H. (2007a) Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence*, 36, 127-142.

Rindermann, H. (2007b) The *g* factor of international cognitive ability comparisons: the homogeneity of results in PISA, TIMSS, PIRLS, and IQ tests across nations. *European Journal of Personality,* 21, 667-706.

Rushton, J. P., & Jensen, A. R. (2005) Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law,* 11, 235-294.

Rushton, J. P., & Jensen, A. R. (2006) The totality of available evidence shows the race IQ gap still remains. *Psychological Science,* 17, 921-922.

Rushton, J. P., & Rushton, E. W. (2003) Brain size, IQ, and racial group differences: evidence from musculoskeletal traits. *Intelligence,* 31, 139-155.

Shaw, P., Greenstein, D., Lerch, J., Clasen, L., Lenroot, R., Gogtay, N., Evans, A., Rapoport, J., & Giedd, J. (2006) Intellectual ability and cortical development in children and adolescents. *Nature,* 440, 676-679.

Sigman, M., & Whaley, S. E. (1998) The role of nutrition in the development of intelligence. In U. Neisser (Ed.), *The rising curve: long-term gains in IQ and related measures.* Washington, DC: American Psychological Association. Pp. 155-182.

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

Spearman, C. S. (1927) *The abilities of man*. New York: Macmillan.

Sternberg, R. J. (1980) Sketch of a componential subtheory of human intelligence. *Behavioral & Brain Sciences*, 3, 573-584.

Sternberg, R. J. (1999) *Successful intelligence.* New York: Plume.

Sticht, T. G., Caylor, J. S., Kern, R. P., & Fox, L. C. (1972) Project REALISTIC: determination of adult functional literacy skill levels. *Reading Research Quarterly*, 7, 424-465.

Sundet, J. M., Barlaug, D. G., & Torjussen, T. M. (2004) The end of the Flynn effect? A study of secular trends in mean intelligence test scores of Norwegian conscripts during half a century. *Intelligence,* 32, 349-362.

Taylor, E. K. (1947) Tables for the determination of the significance of skewness and of the significance of the difference in the skewness of two independent distributions. *Psychometrika*, 12, 111-125.

Teasdale, T. W., & Owen, D. R. (2005) A long-term rise and recent decline in intelligence test performance: the Flynn effect in reverse. *Personality and Individual Differences,* 39, 837-843.

Tiegs, E. W., & Clark, W. W. (1959) *Survey of Reading Achievement: junior high level (grades 7–9).* Monterey, CA: California Test Bureau.

United Nations Development Program. (2006) *United Nations Human Development Report: beyond scarcity: power, poverty and global water crisis*. Basingstoke, UK: Palgrave Macmillan.

Vernon, P. A., Wickett, J. A., Bazana, G., & Stelmack, R. M. (2000) The neuropsychology and psychophysiology of human intelligence. In R. J. Sternberg (Ed.), *Handbook of intelligence.* Cambridge, UK: Cambridge Univer. Press. Pp. 245-264.

Wechsler, D. (1949) *The Wechsler Intelligence Scale for Children.* New York: Psychological Corp.

Wechsler, D. (1955) *The Wechsler Adult Intelligence Scale.* New York: Psychological Corp.

Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, G. C. M., Boomsma, D. I., & Span, M. M. (2004) Are intelligence tests measurement invariant over time? Investigating the nature of the Flynn effect. *Intelligence*, 32, 509-537.

Williams, W. M. (1998) Are we raising smarter children today? School- and home-related influences on IQ. In U. Neisser (Ed.), *The rising curve: long-term gains in IQ and related measures.* Washington, DC: American Psychological Association. Pp. 125-154.

Yerkes, R. M. (Ed.) (1921) Psychological examining in the United States Army. *Memoirs of the National Academy of Sciences*, 15, 1-8.

Accepted April 22, 2010.

This article has been cited by:

1. Vittorio Daniele, Nicola Ostuni. 2013. The burden of disease and the IQ of nations. *Learning and Individual Differences* **28**, 109-118. [CrossRef]
2. Vittorio Daniele. 2013. Does the intelligence of populations determine the wealth of nations?. *The Journal of Socio-Economics* **46**, 27-37. [CrossRef]
3. Marlene Sotelo-Dynega, Samuel O. Ortiz, Dawn P. Flanagan, William F. Chaplin. 2013. ENGLISH LANGUAGE PROFICIENCY AND TEST PERFORMANCE: AN EVALUATION OF BILINGUAL STUDENTS WITH THE WOODCOCK-JOHNSON III TESTS OF COGNITIVE ABILITIES. *Psychology in the Schools* **50**:10.1002/pits.2013.50.issue-8, 781-797. [CrossRef]
4. Kees-Jan Kan, Rogier A. Kievit, Conor Dolan, Han van der Maas. 2011. On the interpretation of the CHC factor Gc. *Intelligence* **39**, 292-302. [CrossRef]
5. Christopher Hassall, Thomas N. Sherratt. 2011. Statistical inference and spatial patterns in correlates of IQ. *Intelligence* **39**, 303-310. [CrossRef]