Intelligence 38 (2010) 213–219

Contents lists available at ScienceDirect

Intelligence

Editorial

The rise and fall of the Flynn Effect as a reason to expect a narrowing of the Black–White IQ gap ☆

J. Philippe Rushton [a,*], Arthur R. Jensen [b]

[a] Department of Psychology, University of Western Ontario, London, Ontario, Canada N6A 5C2
[b] Department of Education, University of California, Berkeley, CA 94308, United States

ARTICLE INFO

Article history:
Received 31 May 2009
Received in revised form 5 December 2009
Accepted 7 December 2009
Available online 6 January 2010

ABSTRACT

In this Editorial we correct the false claim that *g* loadings and inbreeding depression scores correlate with the secular gains in IQ. This claim has been used to render the logic of heritable *g* a "red herring" and an "absurdity" as an explanation of Black–White differences because secular gains are environmental in origin. In point of fact, while *g* loadings and inbreeding depression scores on the 11 subtests of the Wechsler Intelligence Scale for Children correlate *significantly positively* with Black–White differences (0.61 and 0.48, $P<0.001$), they correlate *significantly negatively* (or not at all) with the secular gains (mean $r=-0.33$, $P<0.001$; and 0.13, *ns*, respectively). Moreover, heritabilities calculated from twins also correlate with the *g* loadings ($r=0.99$, $P<0.001$ for the estimated true correlation), providing biological evidence for a true genetic *g*, as opposed to a mere statistical *g*. While the secular gains are on *g*-loaded tests (such as the Wechsler), they are negatively correlated with the most *g*-loaded *components* of those tests. Also, the tests lose their *g* loadedness over time with training, retesting, and familiarity. In an analysis of mathematics and reading scores from tests such as the NAEP and Coleman Report over the last 54 years, we show that there has been no narrowing of the gap in either IQ scores or in educational achievement. From 1954 to 2008, Black 17-year-olds have consistently scored at about the level of White 14-year-olds, yielding IQ equivalents of 85 for 1954, 82 for 1965, 70 for 1975, and 81 for 2008. We conclude that predictions about the Black–White IQ gap narrowing as a result of the secular rise are unsupported. The (mostly heritable) cause of the one is not the (mostly environmental) cause of the other. The Flynn Effect (the secular rise in IQ) is *not* a Jensen Effect (because it does not occur on *g*).

1. Introduction

Ever since the "Flynn Effect" came to light, the "massive gains" in IQ scores over time have been proposed as a reason to expect the 15- to 20-point gap between Blacks and Whites to gradually disappear (Flynn, 1984, 1987a, 1999b). Rather than interpreting the secular gain of 3 IQ points a decade as evidence that people become familiar with test material over time, requiring periodic updates to the test, Flynn took it to mean that "real" intelligence levels have increased, at least in abstract reasoning. Flynn points out that the secular gains are on *g*-loaded tests such as the Raven and Wechsler, which Jensen (1998) described as almost pure measures of *g*, and which factor analyses show involve no significant factors beyond *g*. Furthermore, Flynn (2008) calculated that in 2002, the Black mean IQ was 4 points higher than the White mean in 1947–48.

Contra Flynn, however, Jensen (1998) also pointed out that increased test sophistication and other factors lead to

☆ Portions of this Editorial were presented at the Symposium on Group Differences, 10th Annual Meeting of the International Society for Intelligence Research, Madrid, Spain.

* Corresponding author. Tel.: +1 519 661 3685.
E-mail addresses: rushton@uwo.ca (J.P. Rushton), Nesnejanda@aol.com (A.R. Jensen).

0160-2896/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2009.12.002

enhanced test taking skills and higher scores. Moreover, Jensen disentangled IQ gains from psychometric *g* gains and so predicted no significant real-world effects in terms of intelligence. He noted that tests lose their *g* loadedness over time with training, retesting, and familiarity (see te Nijenhuis, van Vianen, & van der Flier, 2007).

Three recent books present a strong environmental perspective on Black–White differences. All of them assert that the Black–White IQ gap has narrowed. They are: Nisbett's (2009) *Intelligence and How to Get It*, Flynn's (2007) *What is Intelligence?*, and his (2008) *Where Have All the Liberals Gone?* Of the three books, Nisbett's is the most comprehensive and builds upon the other two. In a technical Appendix, "The Case for a Purely Environmental Basis for Black/White Differences in IQ," the author critiques our position (the default hypothesis of behavior genetics) that both individual and group differences are the result of both nature and nurture (Jensen, 1969, 1973, 1998; Rushton, 1995, Rushton & Jensen, 2005), along with many conclusions from *The Bell Curve* (Herrnstein & Murray, 1994). We have replied to the arguments in Nisbett's book in detail (Rushton & Jensen, 2010).

In this editorial, we clarify the relation between *g* loadings, heritabilities, Black–White differences, and the secular rise in IQ. We dispute a claim made by Flynn and Nisbett that *g* loadings and inbreeding depression scores correlate as highly with the secular gains as they do with Black–White differences. Because secular gains are environmental in origin, the claim is said to render heritable *g* an "absurdity" as evidence for a genetic component in race differences.

In reviewing the history of the false claim about heritable *g* and the secular gains, we find we have eliminated the Flynn Effect as a reason to expect Black–White differences to narrow. Furthermore, we present analyses that demonstrate that over the last 54 years there has been no narrowing of the Black–White gap in either IQ or in educational achievement.

2. Black–White differences are greater on the more heritable and *g*-loaded tests

If population group differences are greater on the more *g*-loaded and more heritable subtests, it implies they have a genetic origin (Jensen, 1973, 1998). Strong inference is possible (Platt, 1964): (1) Genetic theory predicts a positive association between heritability and group differences; (2) culture theory predicts a positive association between *environmentality* and group differences; (3) nature + nurture models predict both genetic *and* environmental contributions to group differences; while (4) culture-only theories predict a zero relationship between heritability and group differences.

Jensen (1998) developed the *method of correlated vectors* (MCV) to determine whether there is an association between a column of quantified elements (such as a test's *g* loading or its heritability) and any parallel column of independently derived scores (such as mean differences between groups). Using that method, he (1998, pp. 369–379) summarized 17 independent data sets of nearly 45,000 Blacks and 245,000 Whites derived from 149 psychometric tests and found the *g* loadings consistently predicted the magnitude of the mean Black–White differences ($r = 0.63$, $P < 0.001$). This was true even among three-year-olds administered eight subtests of the Stanford–Binet; the rank-order correlation between the *g* loadings and the Black–White differences was 0.71 ($P < 0.05$; Peoples, Fagan, & Drotar, 1995).

The term "Jensen Effect" has been used to designate significant correlations between *g* loadings and other variables, and they have been found for many other group differences. In Hawaii, *g* loadings from 15 cognitive tests correlated with the mean differences between East Asians and Whites, favoring East Asians (Nagoshi, Johnson, DeFries, Wilson, & Vandenberg, 1984). In South Africa, *g* loadings on the items of the Raven Matrices predicted mean differences (on the items) between White, South Asian, and Black university students (Rushton, Skuy, & Bons, 2004; Rushton, Skuy, & Fridjohn, 2002, 2003). In Serbia, item *g* loadings from the Raven Matrices correlated with mean differences between the Roma (Gypsies, a people of South Asian origin) and Whites. In Zimbabwe, *g* accounted for 77% of the difference between African and White 12- to 14-year-olds in a reanalysis of WISC-R data originally published by Zindi (1994) (Rushton & Jensen, 2003).

The method of correlated vectors has also demonstrated a relation between test *heritabilities* and mean Black–White differences. Nichols (1972) found the heritabilities of 13 tests correlated 0.67 ($P < 0.05$) with the mean Black–White differences. Jensen (1973) reported *environmentalities* (calculated as the degree to which sibling correlations departed from the pure genetic expectation of 0.50) on 16 tests had an inverse relation of −0.70 ($P < 0.01$) with mean Black–White differences. Rushton (1989) found inbreeding depression scores on 11 subtests of the Wechsler Intelligence Scale for Children (WISC) correlated 0.48 ($P < 0.05$) with the mean Black–White differences. Inbreeding depression, a purely genetic effect, occurs when offspring receive two copies of the same harmful recessive gene from each of their closely related parents (see Jensen, 1998, pp. 189–196). The inbreeding depression had been calculated by Schull and Neel (1965) from 1854 cousin marriages in Japan on the WISC and showed an overall 7.5 point decrement (0.50 SD) in the offspring, with each subtest showing a greater or lesser amount. There is no non-genetic explanation for why Black–White differences in the US should be more pronounced on those subtests showing the most inbreeding depression among the Japanese in Japan (Jensen also demonstrated inbreeding depression effects on the Raven Matrices in India; Agrawal, Sinha, & Jensen, 1984).

Criticisms have been made of Jensen's method of correlated vectors. For example, Dolan, Roorda, and Wicherts (2004) and Ashton and Lee (2005) argued that it lacked specificity so that Jensen Effects might occur even when differences are not on *g*. They advocated the use of more powerful statistics such as multi-group confirmatory factor analysis (MGCFA). However, this criticism misses the point because there is no absolute claim that *g* effects have been proven; only that what is observed is what would have been expected if an underlying *g* did in fact exist (see Bartholomew, 2004, for the logic of *g* inferences). Further, several studies have corroborated the results on *g* and group differences using MGCFA with Black–White differences in the US (Wicherts et al., 2004), Black–White differences in South Africa (Rushton et al., 2004), and Roma–White differences in Serbia (Rushton, Cvorovic, and Bons, 2007).

There can be little doubt that components of heritable *g* correlate with mean Black–White differences on the same tests. The relation was found again by Rushton, Bons, Vernon, and Cvorovic (2007) using twins, including 152 pairs of twins from the Minnesota Study of Twins Reared Apart (MISTRA). Heritabilities calculated for 36 diagrammatic puzzles from the Raven Colored Matrices, and 58 from the Standard Matrices, correlated a mean 0.40 ($P<0.05$) with the pass rate differences (on those items) between the Roma in Serbia, and Whites, South Asians, Coloreds, and Blacks in South Africa. Subsequently, Wicherts and Johnson (2009) criticized this study for using "unreliable" item-level analyses, even though the items found relatively difficult (or easy) by twins in North America were the ones found relatively difficult (or easy) by the Roma in Serbia, and by Whites, South Asians, Coloreds, and Blacks in South Africa (mean $r=0.87$). However, Rushton and Jensen (2010) corroborated the results after organizing the items into more reliable *parcels*, each containing six or more items. As the heritability of the parcels increased, so did the mean group differences (mean $r=0.74$; $P<0.01$).

A Jensen Effect for heritability has also been found, with the *g* loadings from various subtests correlating with the heritabilities of these same subtests (Jensen, 1998). A Jensen Effect for heritability provides biological evidence for a true genetic *g*, as opposed to the mere statistical reality of *g*. It makes problematic theories of intelligence that do not include a general factor as an underlying biological variable, but only explain the positive manifold, such as the model proposed by Dickens and Flynn (2001), and the mutualism model by van der Maas, Dolan, Grasman, Wicherts, Huizenga, and Raijmakers (2006).

Recent Jensen Effects for heritability come from two studies conducted in the Netherlands (Kan, Haring, Dolan, & van der Maas, 2009; van Bloois, Geujes, te Nijenhuis, & de Pater, 2009). In a psychometric meta-analysis on 1512 twin pairs, van Bloois et al. (2009) found a value of +1.01 for the estimated true correlation between *g* and heritability. In a re-analysis of the Raven Matrices data by Rushton, Bons, et al. (2007), we correlated the 36 item heritabilities on the Colored Matrices (e.g., from twins reared together) and the 58 on the Standard Matrices (e.g., from the Minnesota Study of Twins Reared Apart), with the item *g* loadings (e.g., from the item-total scores) and found a mean *r* of 0.47 ($P<0.01$). Correcting the correlations raised the value from 0.55 to 1.00 (depending on whether using the test's alpha coefficient or the item's test–retest correlation). Arranging the items into parcels also raised the original value (The item-level data are available on-line at the journal; Rushton, Bons, et al., 2007).

3. Do *g* and inbreeding depression scores also correlate with the secular trends?

The pervasiveness and potency of heritable *g* came to widespread attention with the publication of *The g Factor* (Jensen, 1998), *The Bell Curve* (Herrnstein & Murray, 1994), and *Race, Evolution, and Behavior* (Rushton, 1995). Thus, Herrnstein and Murray (1994) made *g* pivotal to their thesis that intelligence was the basis for social stratification in America. Rushton (1995) made *g* central to his theory that race differences in IQ had evolved as part of a coordinated life history of 60 different traits.

Fig. 1, taken from Rushton (1995), shows the regression of Black–White differences on *g* factor loadings and inbreeding depression scores from the 10 sets of WISC *g* loadings and 5 sets of Black–White differences ($N=4848$) previously summarized by Jensen (1985, 1987). As the *g* loadings and inbreeding depression scores increase, so do mean Black–White differences. These findings led Rushton to infer a genetic origin for the race differences.

Flynn (1999a, p. 373) offered "Evidence against Rushton" by examining the relation between the inbreeding depression scores and the five sets of gain scores on the same 11 WISC subtests. In his first analysis, Flynn found inbreeding depression correlated between −0.08 and +0.18 (mean 0.08) with the total gains on the WISC. When he examined their relation to the six Performance subtests, he found these too averaged a non-significant −0.05. However, when Flynn looked at the relation between the inbreeding depression scores and the gain scores for the five Verbal subtests, he found they correlated 0.52. This was not significant either with an $N=5$. However, its numerical value, and the fact that a correlation of 0.30 or higher was found in all five samples, enabled Flynn (1999a) to offer it as rebuttal.

In his reply to Flynn, Rushton (1999) analyzed *all* the data on the 11 WISC subtests from Rushton (1995) and Flynn (1999a). Table 1 presents the zero-order correlations in the top half of the matrix and the first-order partial correlations (after controlling for reliability) in the lower half of the matrix. As can be seen, inbreeding depression correlated significantly positively with

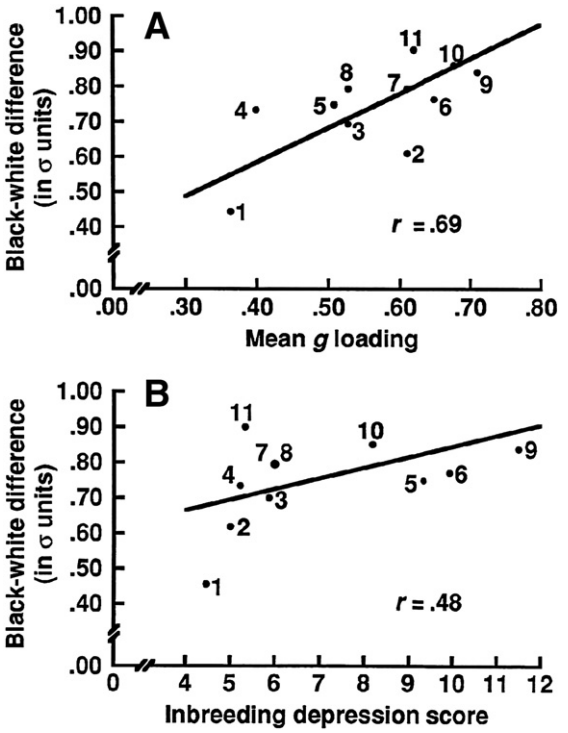

Fig. 1. Regression of Black–White differences on *g* loadings (Panel A) and on inbreeding depression scores (Panel B). The numbers indicate subtests from the Wechsler Intelligence Scale for Children-Revised: 1, Coding; 2, Arithmetic; 3, Picture completion; 4, Mazes; 5, Picture arrangement; 6, Similarities; 7, Comprehension; 8, Object assembly; 9, Vocabulary; 10, Information; 11, Block design.
From Rushton (1995: p. 188, Figure 9.1).

Table 1
Pearson correlations of variables using subtests of the Wechsler Intelligence Scale for Children-Revised (zero-order correlations above diagonal; reliabilities partialed out below diagonal).

	Inbreeding depression scores	Reliabilities	Black–White differences	WISC-R *g* loadings	WISC-III *g* loadings	U.S. gains 1	U.S. gains 2	German gains	Austria gains	Scotland gains
Inbreeding depression scores	1.00	.50	0.48	0.61	0.39	−0.07	0.07	0.22	0.29	0.13
Reliabilities	–	1.00	0.60	0.84	0.73	−0.27	−0.54	0.00	0.16	−0.23
Black–White differences	0.26	–	1.00	0.69	0.53	−0.28	−0.05	0.21	0.22	0.31
WISC-R *g* loadings	0.40	–	0.43	1.00	0.94	−0.38	−0.44	−0.18	−0.04	−0.22
WISC-III *g* loadings	0.05	–	0.17	0.87	1.00	−0.35	−0.48	−0.34	−0.09	−0.73
U.S. gains 1	0.07	–	−0.16	−0.30	−0.24	1.00	0.46	0.46	0.70	0.86
U.S. gains 2	0.47	–	0.41	0.03	−0.14	0.39	1.00	0.73	0.54	0.68
German gains	0.25	–	0.27	−0.33	−0.50	0.48	0.86	1.00	0.76	0.80
Austria gains	0.24	–	0.15	−0.32	−0.31	0.79	0.75	0.77	1.00	0.58
Scotland gains	0.28	–	0.56	−0.06	−0.85	0.85	0.68	0.82	0.64	1.00

the Black–White differences ($r = 0.48$; $P < 0.05$) but not with the gain scores (mean $r = 0.13$; range = −0.07 to 0.29). Similarly, the *g* loadings correlated *significantly positively* with the Black–White differences (0.53, 0.69) but *significantly negatively* with the gain scores (mean $r = -0.33$; range = −0.04 to −0.73; $P < 0.00001$, Fisher, 1970, pp. 99–101).

Rushton (1999) also conducted a principal components analysis of the partialed correlation matrix and extracted two significant components with eigenvalues > 1. Table 2 presents these in both unrotated and varimax rotated forms. The relevant findings are: (1) the IQ gains on the WISC-R and WISC-III form a cluster, showing that the secular trend in overall scores is a reliable phenomenon; but (2) this cluster is *independent* of the cluster formed by Black–White differences, inbreeding depression scores (a purely genetic effect), and *g* factor loadings (a largely genetic effect). This analysis shows that the secular increase in IQ and the mean Black–White differences in IQ behave in entirely different ways. The secular increase is unrelated to *g* and other heritable measures, while the magnitude of the Black–White difference is related to heritable *g* and inbreeding depression.

Table 2
Principal components analysis and varimax rotation for Pearson correlations of inbreeding depression scores, Black–White differences, g loadings, and gains over time on the Wechsler Intelligence Scales for Children with reliability partialed out.

Variables	Principal components			
	Unrotated loadings		Varimax rotated loadings	
	I	II	1	2
Inbreeding depression scores from Japan (WISC-R)	0.31	0.61	0.26	**0.63**
Black–White differences from the U.S. (WISC-R)	0.29	0.70	0.23	**0.72**
WISC-R g loadings from the U.S.	-0.33	0.90	-0.40	**0.87**
WISC-III g loadings from the U.S.	-0.61	0.64	-0.66	**0.59**
U.S. gains 1 (WISC to WISC-R)	0.73	-0.20	**0.75**	-0.13
U.S. gains 2 (WISC-R to WISC-III)	0.81	0.40	**0.77**	0.47
German gains (WISC to WISC-R)	0.91	0.03	**0.91**	0.11
Austria gains (WISC to WISC-R)	0.87	0.00	**0.86**	0.07
Scotland gains (WISC to WISC-R)	0.97	0.08	**0.96**	0.17
% of total variance explained	48.6	25.49	48.44	25.65

Note. From "Secular gains in IQ not related to the *g* factor and inbreeding depression—unlike Black–White differences: A reply to Flynn," by J. P. Rushton, 1999, *Personality and Individual Differences*, *26*, 381–389. Copyright 1999 by Elsevier Science. Reprinted with permission of publisher.

In order to provide a new "counterweight to Rushton's analysis," Flynn (2000, p. 214) collaborated with William Dickens. They: (1) discarded the WISC Maze subtest, thereby reducing the number of subtests from 11 to 10 (no reason given); (2) discarded the gain scores and Black–White differences on the WISC-III on the grounds that most of the data were on the WISC; (3) averaged the five sets of gain scores on the grounds that five gain indicators were too many for Rushton's factor analysis to be fair (though Rushton had used an equal number of variables to extract *g*); and (4) calculated a new *g* loading for each of the Wechsler subtests by correlating it with the Raven Matrices and retaining *some* of the results.

Flynn (2000) argued that it was necessary to calculate this highly selective "alternative" *g* because the Matrices, an excellent measure of "fluid" *g*, showed the greatest secular gains while Rushton had measured "crystallized" *g* (though Rushton, in fact, used the standard method to extract *g* from the Wechsler tests and Flynn's new *g* correlated not at all with the WISC *g*, although it too had shown substantial secular gains). Flynn (2000) reported a series of non-significant correlations (with $N = 10$): (1) 0.50 between *g* and secular gains, reversing Rushton's highly significant *negative* −0.33; (2) 0.28 between inbreeding depression and secular gains, up from Rushton's near zero 0.13; (3) 0.50 between *g* and Black–White differences, down from Rushton's significant 0.61; and (4) 0.29 between inbreeding depression and Black–White differences, down from Rushton's significant 0.43.

Flynn (2000) acknowledged that "The data contained herein are not robust" (p. 212) and that none of his new correlations were significant with $N = 10$. Nonetheless, he claimed they cast doubt on the relation between heritable *g* and Black–White differences because the logic of heritable *g* led to the "absurd" conclusion that the secular gains were also heritable. Subsequently, both he, and especially Nisbett, dismissed heritable *g* as a "red herring" for the race-IQ debate (2009, pp. 216–218).

Also contra Flynn and Nisbett, a *negative* correlation between *g* and secular gains has been found in other countries. For example, a negative correlation of −0.40 was found between *g* and the secular rise in Estonia over a 60-year period from 1934 to 1998 with 12- to 14-year-olds on the Estonian National Intelligence Test (Must, Must, & Raudik, 2003). Although not all studies confirm the negative correlation, a recent meta-analysis of 17 studies ($N = 12{,}732$) has provided a remarkably exact corroboration of Rushton's (1999) finding, with a rho of −0.33 ($P < 0.00001$) between *g* and the secular gains (te Nijenhuis & van der Flier, 2009).

Independent procedures also demonstrated that Black–White differences are qualitatively different from cohort differences. Studies using multi-group confirmatory factor analyses (MGCFA) have found that measurement invariance is often *present* in data on Black–White differences, indicating that the test scores have *similar* meanings for both groups (Dolan, 2000; Dolan & Hamaker, 2001). On the other hand, measurement invariance is typically *absent* in data on cohort differences, indicating the test scores have *different* meanings for these groups (Wicherts et al., 2004).

Interestingly, in his most recent book, Flynn (2008) has apparently changed his mind about the relation between *g* and Black–White differences. While he still maintains the race differences are mostly environmental in origin, he now agrees with Rushton and Jensen (2005) and disagrees with Nisbett (2009), as well as his own former opinion (2000):

> There are two messages. The first is familiar: You cannot dismiss black gains on whites just because they do not tally with the *g* loadings of subtests. But the second is new and unexpected. The brute fact that black gains on whites do not tally with *g* loadings tells us something about causes. The causes of the black gains are like hearing aids. They do cut the cognitive gap but they are not eliminating the root causes. And conversely, if the root causes are somehow eliminated, we can be confident that the IQ gap and the *g* gap will both disappear (p. 85).

4. Is the IQ gap narrowing?

Rushton and Jensen (2005, 2010) maintain that the IQ gap between Blacks and Whites has remained at least 15- to 20-points (1.1 standard deviations) since the time of World War I (1917) when mass testing first began (Roth, Bevier, Bobko, Switzer, & Tyler, 2001; Shuey, 1966). On the other hand, Flynn (1987b, 1999b) argued that the mean difference has decreased from the Army Alpha of World War I (1917), to the Army General Classification Test of World War II (1946), to the Armed Forces Qualification Test of the Vietnam era (1968). More recently, Dickens and Flynn (2006) claimed that Blacks had closed the IQ gap by 5.5 points (35%) between 1970 and 1992. Over the same time period, Nisbett (2009) claimed that Blacks had narrowed the gap in educational achievement by a commensurate 35% on the National Assessment of Educational Progress (NAEP) tests. Nisbett also argued that educational interventions such as the Milwaukee project, the Abecedarian project, and the Infant Health and Development Program implied that the gap could be eliminated altogether.

To the contrary, we find there is little or no evidence of narrowing. The evidence presented in its favor rests mainly on insufficient sampling and selective reporting. For example, Rushton and Jensen (2006) calculated that the mean Black gain on the IQ tests discussed by Dickens and Flynn (2006) was only 2.1 points (14%) because these authors, for a variety of proffered methodological reasons, had excluded several tests showing small, nil, and negative gains, and also because they had used a projected trend line that exaggerated the gain. Nor was there any evidence of narrowing on other IQ tests over the 1970 to 1992 time period (Murray, 2006, 2007).

Nisbett's (2009) claim of a 35% Black improvement on the NAEP tests is also greatly exaggerated. Gottfredson (2005) estimated these gains were only about 20% and had ceased completely by 1990. In fact, her appraisal, as well as one by Herrnstein and Murray (1994) of a 20% Black gain may have been over-optimistic (Herrnstein and Murray, 1994, actually reported the results were mixed, with other tests showing an *increasing* distance between Blacks and Whites).

To get a more complete picture, we calculated the mean of the mathematics and reading scores from the NAEP long-term assessment tests from 1975 to 2008 for the White, Black, and Hispanic 9-, 13-, and 17-year-olds. Fig. 2 plots the scores for White, Hispanic, and Black 17-year-olds, plus those for White 13-year-olds. As can be seen, Black 17-year-olds have not closed the gap on Hispanic 17-year-olds (for many of whom English is a second language), and barely closed it on

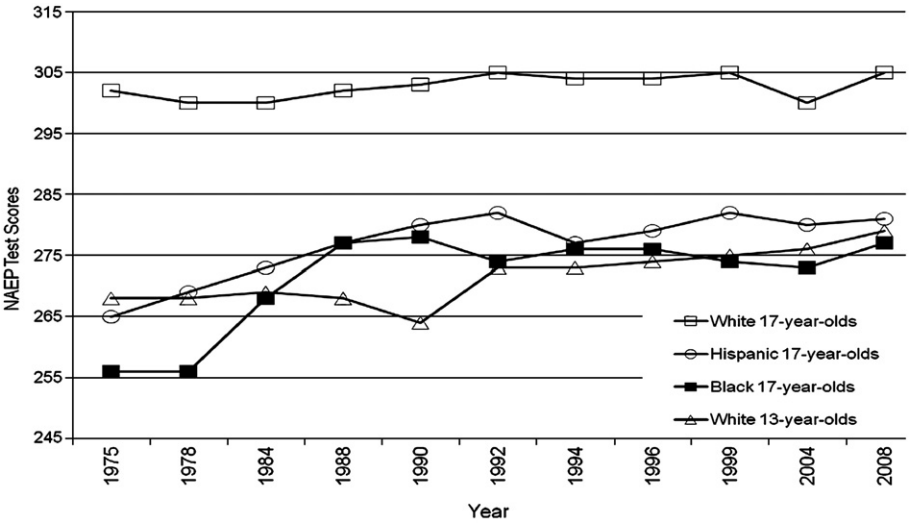

Fig. 2. NAEP scores from 1975 to 2008 for White 13-year-olds and White, Hispanic, and Black 17-year-olds. Data are from Rampey, Dion, and Donahue (2009: pp. 14–17, 34–37, Figures 4, 5, 10, and 11).

White 13-year-olds. Black 17-year-olds lag White 17-year-olds by over three years. The comparison of Black 13-year-olds with Hispanic 13-year-olds and White 9-year-olds shows similar results. Note that these data are from nationally representative samples of over 26,000 students; the NAEP tests are often referred to as "The Nation's Report Card."

The 3+ year education gap between Blacks and Whites did not begin with the 1975 NAEP tests. It was found from 1954 to 1965 in the State of Georgia with data on reading and mathematics from about 1500 White and 800 Black students using the California Achievement Test (Osborne, 1961, 1967). Both Blacks and Whites improved their scores with age, and showed the now familiar secular rise in scores. However, by grade 10 (age 16), the Black–White achievement gap remained consistently at about three years. In Virginia, Garrett (1964) carried out a study of reading ability in 2000 Black and White students and found the mean difference of three years by grade 7 (age 13). Both Garrett and Osborne's studies were dismissed as due to "convenience samples" and the *result* of the school segregation legally mandated at the time in the South (rather than as a *cause* of segregation, as the system apologists declared).

The Coleman Report (1966) authorized by the Civil Rights Act of 1964 and carried out under the auspices of the U.S. Department of Health, Education and Welfare, confirmed Osborne and Garrett's observations. In a nationally representative survey of nearly 600,000 schoolchildren and 60,000 teachers from 4000 schools throughout the US, including from the metropolitan northeast and California, mean Black achievement scores averaged 1.6 years behind that of Whites in grade 6 (at age 12); 2.4 years in grade 9 (age 15); and 3.3 years in grade 12 (age 18). The Report also found that Blacks lagged American Indians, despite this population scoring lower than Blacks on most socioeconomic indicators. It surprisingly found that the educational resources devoted to Blacks and Whites were nearly equal, even in the South, and that none of the expected financial or educational "inputs" could be correlated with any "outputs." The main determinant of children's test scores was not the amount of money spent on schools, but the parents' socioeconomic status. Going to good or bad schools, by itself, apparently had little influence on the students' performance on standardized tests.

Coleman et al. (1966) did find, however, that Black students who attended middle-class majority White schools achieved higher than other Blacks. They surmised this was due to peer attitudes in such schools and recommended that Black students be assigned to schools where there was a majority of middle-class attitudes, a recommendation that earned Coleman the moniker, "the sociologist who inspired busing." Across much of the U.S., forced integration through court-ordered busing transferred tens of thousands of White and Black students to each other's schools. By 1975, Coleman had to publish that school busing led to "White flight" as parents moved their children to private schools and ever more distant suburbs.

In order to re-examine the Black–White differences over the last 54 years, we calculate mean Black IQs from the formula IQ = MA/CA × 100, with the White mean set at 100. From the 1954 Georgia study (Osborne, 1967, p. 385), the mean IQ for Black 8th graders (14-year-olds) was 86 (12/14 × 100), and in 1965, 81 (11.3/14 × 100). From the 1966 Coleman Report, the mean IQ for Black 12-year-olds was 87 (10.4/12 × 100); for 15-year-olds, 84 (12.6/15 × 100); and for 18-year-olds, 82 (14.7/18 × 100). From the 1975 NAEP tests, the mean IQ for Black 13-year-olds was 70 (9/13 × 100), and for 17-year-olds, 71 (12/17 × 100); from the 2008 NAEP tests, for Black 13-year-olds, 85 (11/13 × 100); and for 17-year-olds, 77 (13/17 × 100). These results indicate *no* Black gain in either mean IQ or in educational achievement for over 50 years.

A much stronger dose of skepticism is required than either Flynn or Nisbett have demonstrated in regard to the power of educational interventions. As Jensen (1969) pointed out long ago, when it comes to what can be done to increase IQ and school achievement scores, sadly, the answer is still "not much."

5. Conclusion

Heritable *g* is at the core of the debate over how much the mean Black–White gap in IQ and school achievement is due to the genes rather than to the environment, and therefore, how much it can be expected to narrow. While *g* and genetic estimates correlate *significantly positively* with Black–White differences 0.61 and 0.48 ($P<0.001$), they correlate *significantly negatively* (or not at all) with the secular gains ($r=-0.33$; $P<0.001$) and 0.13 (*ns*). Similarly, *g* loadings and heritabilities from the items of the Raven Matrices correlate significantly positively with each other and with Black–White differences (mean $r=0.74$, $P<0.01$). Although the secular gains are on *g*-loaded tests (such as the Wechsler), they are *negatively* correlated with the most *g*-loaded *components* of those tests. Tests lose their *g* loadedness over time as the result of training, retesting, and familiarity (te Nijenhuis et al., 2007).

Some issues, however, remain to be resolved. For example, Lynn (2009) found a secular rise in the Developmental Quotients of infants in the first two years of life, which he suggested was due to improved pre-natal and early post-natal nutrition. He supported his conjecture by pointing to equivalent gains in birth weight, stature, and brain size, and the correlation of these variables with later IQ. If it becomes possible to disentangle environmental factors that do affect *g*, from the environmental factors that do not affect *g*, the negative correlation between *g* and secular gains may increase from −0.33 to nearer −1.00.

Predictions about the Black–White IQ gap narrowing due to the secular rise is based on faith rather than evidence. There is no more reason to expect Black–White differences in IQ to narrow as a result of the secular rise in IQ than to expect male–female differences in height to narrow as a result of the secular rise in height. The (mostly heritable) cause of the one is not the (mostly environmental) cause of the other. From the present perspective, the Flynn Effect (the secular rise in IQ) is *not* a Jensen Effect (because it does not occur on *g*).

References

Agrawal, N., Sinha, S. N., & Jensen, A. R. (1984). Effects of inbreeding on Raven's matrices. *Behavior Genetics*, *14*, 579−585.
Ashton, M., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence*, *33*, 431−444.
Bartholomew, D. J. (2004). *Measuring intelligence: Facts and fallacies.* Cambridge University Press.
Coleman, J. S. (1975). Recent trends in school integration. *Educational Researcher*, *4*, 3−12.

Coleman, J. S., Campbell, E. Q., Hobson, C. J., McPartland, J., Mood, A. M., Weinfeld, F. D., & York, R. L. (1966). *Equality of educational opportunity*. Washington, DC: U.S. Department of Health, Education, and Welfare.
Dickens, W. T., & Flynn, J. R. (2001). Heritability estimates versus large environmental effects: The IQ paradox resolved. *Psychological Review*, *108*, 346—369.
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. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, *35*, 21—50.
Dolan, C. V., & Hamaker, E. L. (2001). Investigating Black–White differences in psychometric IQ: Multi-group confirmatory factor analyses of the WISC-R and K-ABC and a critique of the method of correlated vectors. In F. Columbus (Ed.), *Advances in Psychology Research*, *vol. 6*. (pp. 31—59) Huntington, NY: Nova Science Publishers.
Dolan, C. V., Roorda, W., & Wicherts, J. M. (2004). Two failures of Spearman's hypothesis: The GATB in Holland and the JAT in South Africa. *Intelligence*, *32*, 155—173.
Fisher, R. A. (1970). *Statistical methods for research workers*, 14th ed. New York: Hafner Press.
Flynn, J. R. (1984). The mean IQ of Americans: Massive gains 1932 to 1978. *Psychological Bulletin*, *95*, 29—51.
Flynn, J. R. (1987a). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171—191.
Flynn, J. R. (1987b). Race and IQ: Jensen's case refuted. In S. Modgil & C. Modgil (Eds.), *Arthur Jensen: Consensus and controversy* (pp. 221—232). Lewes, England: Falmer Press.
Flynn, J. R. (1999a). Evidence against Rushton: The genetic loading of WISC-R subtests and the causes of between-group IQ differences. *Personality and Individual Differences*, *26*, 373—379.
Flynn, J. R. (1999b). Searching for justice: The discovery of IQ gains over time. *American Psychologist*, *54*, 5—20.
Flynn, J. R. (2000). IQ gains, WISC subtests and fluid *g*: *g* theory and the relevance of Spearman's hypothesis to race. In G. R. Bock, J. A. Goode, & K. Webb (Eds.), *The nature of intelligence: The Novartis Foundation symposium* (pp. 202—227). New York: Wiley.
Flynn, J. R. (2007). *What is intelligence? Beyond the Flynn Effect.* New York: Cambridge University Press.
Flynn, J. R. (2008). *Where have all the liberals gone? Race, class, and ideals in America.* New York: Cambridge University Press.
Garrett, H. E. (1964). IQ and school achievement of Negro and White children of comparable age and school status. *Mankind Quarterly*, *5*, 45—49.
Gottfredson, L. S. (2005). What if the hereditarian hypothesis is true? *Psychology, Public Policy, and Law*, *11*, 311—319.
Herrnstein, R. J., & Murray, C. (1994). *The bell curve.* New York, NY: Free Press.
Jensen, A. R. (1969). How much can we boost IQ and scholastic achievement? *Harvard Educational Review*, *39*, 1—123.
Jensen, A. R. (1973). *Educability and group differences.* London: Methuen.
Jensen, A. R. (1985). The nature of the black–white difference on various psychometric tests: Spearman's hypothesis. *Behavioral and Brain Sciences*, *8*, 193—263.
Jensen, A. R. (1987). Further evidence for Spearman's hypothesis concerning the black–white differences on psychometric tests. *Behavioral and Brain Sciences*, *10*, 512—519.
Jensen, A. R. (1998). *The g factor.* Westport, CT: Praeger.
Kan, K. -J., Haring, S., Dolan, C., & van der Maas, H. (2009, December 19). Evaluation of the relation between heritability (h^2) and g-loadings in psychometric intelligence tests. *Paper presented at the Symposium on Group Differences, 10th Annual Meeting of the International Society for Intelligence Research, Madrid, Spain*.
Lynn, R. (2009). What has caused the Flynn effect? Secular increases in the Development Quotients of infants. *Intelligence*, *37*, 16—24.
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.
Murray, C. (2007). The magnitude and components of change in the black–white IQ difference from 1920 to 1991: A birth cohort analysis of the Woodcock–Johnson standardizations. *Intelligence*, *35*, 305—318.
Must, O., Must, A., & Raudik, V. (2003). The secular rise in IQs: In Estonia, the Flynn effect is not a Jensen effect. *Intelligence*, *31*, 461—471.
Nagoshi, C. T., Johnson, R. C., DeFries, J. C., Wilson, J. R., & Vandenberg, S. G. (1984). Group differences and first principal-component loadings in the Hawaii Family Study of Cognition: A test of the generality of "Spearman's hypothesis.. *Personality and Individual Differences*, *5*, 751—753.
Nichols, P. L. (1972). *The effects of heredity and environment on intelligence test performance in 4– and 7–year–old white and Negro sibling pairs.* Unpublished doctoral dissertation, University of Minnesota, Minneapolis.
Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count.* New York: Norton.
Osborne, R. T. (1961). School achievement of White and Negro children of the same mental and chronological ages. *Mankind Quarterly*, *2*, 26—29.
Osborne, R. T. (1967). Racial differences in mental growth and school achievement. In R. E. Kuttner (Ed.), *Race and modern science* (pp. 381—406). New York: Social Science Press.
Peoples, C. E., Fagan, J. F., III, & Drotar, D. (1995). The influence of race on 3-year-old children's performance on the Stanford–Binet: Fourth edition. *Intelligence*, *21*, 69—82.
Platt, J. R. (1964). Strong inference. *Science*, *146*, 347—353.
Rampey, B. D., Dion, G. S., & Donahue, P. L. (2009). *NAEP 2008 trends in academic progress (NCES 2009-479)*. Washington, DC: National Center for Education Statistics, U. S. Department of Education.
Roth, P. L., Bevier, C. A., Bobko, P., Switzer, F. S., III, & Tyler, P. (2001). Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, *54*, 297—330.
Rushton, J. P. (1989). Japanese inbreeding depression scores: Predictors of cognitive differences between Blacks and Whites. *Intelligence*, *13*, 43—51.
Rushton, J. P. (1995). *Race, evolution, and behavior: A life history perspective.* New Brunswick, NJ: Transaction.
Rushton, J. P. (1999). Secular gains in IQ not related to the g factor and inbreeding depression—unlike Black–White differences: A reply to Flynn. *Personality and Individual Differences*, *26*, 381—389.
Rushton, J. P., & Jensen, A. R. (2003). African–White IQ differences from Zimbabwe on the Wechsler Intelligence Scale for Children-Revised are mainly on the *g* factor. *Personality and Individual Differences*, *34*, 177—183.
Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on group differences in cognitive ability. *Psychology, Public Policy, and the Law*, *11*, 235—294.
Rushton, J. P., & Jensen, A. R. (2006). The totality of available evidence shows race-IQ gap still remains. *Psychological Science*, *17*, 921—922.
Rushton, J. P., & Jensen, A. R. (2010). Race and IQ: A theory-based review of the research in Richard Nisbett's *Intelligence and How to Get It*. *The Open Psychology Journal*, *3*, 9—35.
Rushton, J. P., Skuy, M., & Fridjohn, P. (2002). Jensen Effects among African, Indian, and White engineering students in South Africa on Raven's Standard Progressive Matrices. *Intelligence*, *30*, 409—423.
Rushton, J. P., Skuy, M., & Fridjhon, P. (2003). Performance on Raven's Advanced Progressive Matrices by African, Indian, and White engineering students in South Africa. *Intelligence*, *31*, 123—137.
Rushton, J. P., Skuy, M., & Bons, T. A. (2004). Construct validity of Raven's Advanced Progressive Matrices for African and non-African engineering students in South Africa. *International Journal of Selection and Assessment*, *12*, 220—229.
Rushton, J. P., Bons, T. A., Vernon, P. A., & Cvorovic, J. (2007). Genetic and environmental contributions to population group differences on the Raven's Progressive Matrices estimated from twins reared together and apart. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *274*, 1773—1777.
Rushton, J. P., Cvorovic, J., & Bons, T. A. (2007). General mental ability in South Asians: Data from three Roma (Gypsy) Communities in Serbia. *Intelligence*, *35*, 1—12.
Schull, W. J., & Neel, J. V. (1965). *The effects of inbreeding on Japanese children.* New York: Harper & Row.
Shuey, A. M. (1966). *The testing of Negro intelligence*, 2nd ed. New York: Social Science Press.
te Nijenhuis, J., & van der Flier, H. (2009). *Is the Flynn Effect on g?* Manuscript under review: A meta-analysis.
te Nijenhuis, J., van Vianen, A. E. M., & van der Flier, H. (2007). Score gains on *g*-loaded tests: No *g*. *Intelligence*, *35*, 283—300.
van Bloois, R. M., Geutjes, L. -L., te Nijenhuis, J., & de Pater, I. E. (2009, December 19). *g* loadings and their true score correlations with heritability coefficients, giftedness, and mental retardation: Three psychometric meta-analyses. *Paper presented at the Symposium on Group Differences, 10th Annual Meeting of the International Society for Intelligence Research, Madrid, Spain*.
van der Maas, H. L. J., Dolan, C. V., Grasman, R. P. P. P., Wicherts, J. M., Huizenga, H. M., & Raijmakers, M. E. J. (2006). A dynamical model of general intelligence: The positive manifold of intelligence by mutualism. *Psychological Review*, *113*, 842—861.
Wicherts, J. M., & Johnson, W. (2009). Group differences in the heritability of items and test scores. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *276*, 2675—2683.
Wicherts, J. M., Dolan, C. V., Hessen, D. J., Oosterveld, P., van Baal, 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.
Zindi, F. (1994). Differences in performance. *The Psychologist*, *7*, 549—552.