PERGAMON

Personality and Individual Differences 26 (1999) 381–389

PERSONALITY AND
INDIVIDUAL DIFFERENCES

Secular gains in IQ not related to the *g* factor and inbreeding depression — unlike Black–White differences: A reply to Flynn

J. Philippe Rushton *

Department of Psychology, University of Western Ontario, London, Ont. N6A 5C2, Canada

Received 31 January 1998

Abstract

Earlier, Rushton (*Intelligence*, 13, 1989, 43–51) found the amount of inbreeding depression on WISC-R subtests (a genetic effect) correlates with the magnitude of the average Black–White IQ differences on the same tests. Flynn (*Personality and Individual Differences*, 26, 1999, 391–393) claimed that this tells us nothing about the origins of the race differences because the inbreeding depression scores also correlate with IQ gains made in various countries across generations (an environmental effect). Flynn's conclusion, however, rested on a very circumscribed analysis. When all the data sets are considered together, a principal components analysis shows that whereas the IQ gains over time on the WISC-R and WISC-III do cluster, suggesting they are a reliable phenomenon, these are independent of the cluster of Black–White differences, inbreeding depression scores, and *g* factor loadings. This result strongly supports the heritability of the Black–White differences and contradicts Flynn's claim that the "massive IQ gains over time" found in several countries prove that the Black–White differences in IQ are environmental in origin. The results also provide convergent and discriminant validity for the "Jensen effect" which occurs when significant correlations are found between *g*-factor loadings and other variables.

Keywords: Lynn–Flynn effect; *g*-factor; Wechsler IQ tests; Jensen effect; Race and IQ; Heritability; Construct validity

1. Introduction

Flynn (1984, 1987, 1999) has long championed the view that the "massive IQ gains over time" in the industrialized world show that the average Black–White IQ difference is

---

* To whom all correspondence should be addressed. Tel.: +1-519-661-3685; fax: +1-519-850-2302; e-mail: rushton@julian.uwo.ca

PII: S0191-8869(98)00148-2

environmental in origin. Because several countries have increased their average national IQs over the last half century by amounts comparable to the differences between Blacks and Whites (in the U.S. these gains reach a magnitude of 3 points a decade for 5 decades), Flynn hypothesized that the Black–White differences are caused by the same processes that produce secular gains (such as improvements in schooling and test taking skills). The rising test scores have become known as the "Flynn effect" after Flynn's massive documentation of the phenomenon. However, Rushton (1997b: p. 79) proposed that the rise in IQ phenomenon might be better named the "Lynn–Flynn effect" because it was actually the Lynn (1982) article in *Nature* that first identified the trend in recent times (among the Japanese).

Flynn (1999) recognizes that some of the best evidence against the environmentalist argument is the fact that while the Black–White IQ gap averages 15 points, it is more pronounced on tests of high heritability than it is on tests of low heritability. A parallel relation is found with *g*, the general factor of intelligence, where the Black–White difference is more pronounced on the more *g*-loaded tests. These tests, which best reflect the *g*-factor, also tend to be more heritable. Although tests' heritability coefficients and *g*-loadings are conceptually distinct, studies have found that they both are positively correlated with Black–White differences.

Flynn's current paper is focussed on a subset of this evidence — namely Rushton's (1989) finding that Black–White differences are greatest on IQ subtests showing the most inbreeding depression (cited by Flynn as Rushton, 1995). Inbreeding depression occurs when harmful recessive genes combine, which is more likely in the offspring of closely related parents. The estimates of inbreeding depression had been calculated for various subtests of the Wechsler Intelligence Scale for Children-Revised (WISC-R) in cousin marriages in Japan so the results are unlikely to reflect any cultural differences between Blacks and Whites in the U.S. There really is no non-genetic explanation for the correlation between inbreeding depression and Black–White differences.

Fig. 1 summarizes not only the results cited by Flynn (from Rushton, 1995, pp. 185–187) but also those based on the *g* factor (p. 187, Figure 9.1) which Flynn omitted to mention. Although Flynn neglects to say so, the Rushton (1989) inbreeding study was built directly on Jensen's (Jensen, 1985, 1987; Naglieri and Jensen, 1987) finding that Black–White differences are most pronounced on tests of high *g*. As can be seen in Fig. 1, as either the *g* loadings (panel A) or the inbreeding depression scores (panel B) increase, the differences between Blacks and Whites also increase. The inbreeding results in Fig. 1 are especially conclusive because they are contrary to the effects of any cultural influences. Culture theory predicts that differences between Blacks and Whites will be greater on those subtests most affected by the environment. Yet, the race differences are clearly greater on the more heritable subtests.

Flynn (1999) limited his criticism of the data in Fig. 1 to those on inbreeding depression in panel B while completely sidestepping the data on the *g* factor in Panel A, as well as those on the Black–White differences themselves although these are the central focus of his concern. His counter-claim is that the inbreeding depression scores are meaningless as genetic indicators because they also correlate with the (environmentally induced) secular increase in IQ. Even the empirical support for this limited interpretation is ambiguous. When Flynn examined all 11 subtests, they failed to yield a positive effect; the 6 performance subtests gave "no consistent result"; only the 5 verbal subtests showed Flynn's hypothesized relationship, and these only

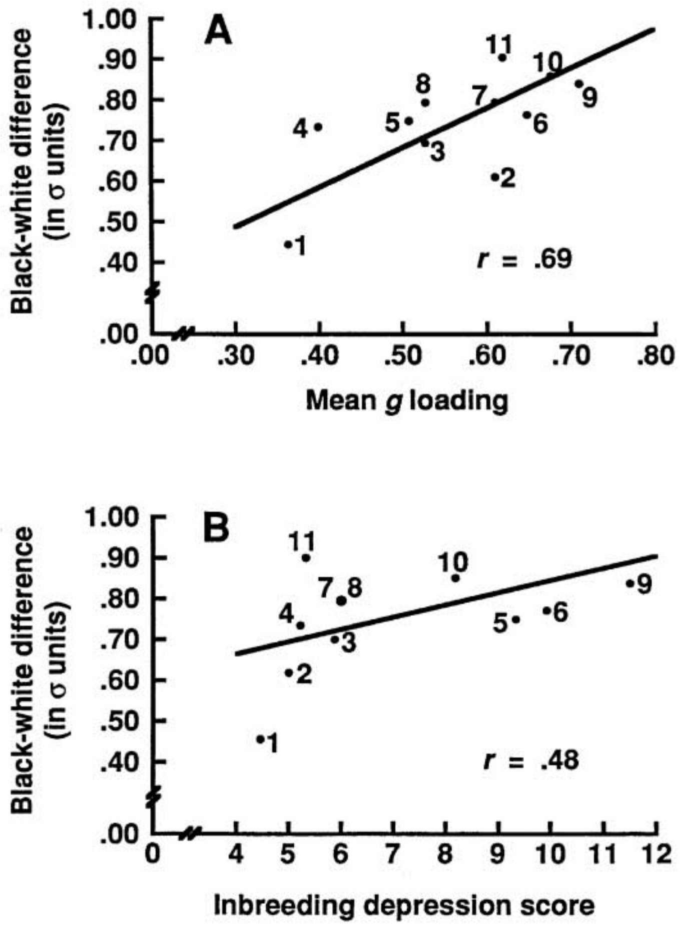

Fig. 1. Regression of Black–White differences on *g* loadings (panel A) and on inbreeding depression scores (panel B). The numbers indicate subtests from the Wechsler Intelligence Scale for Children-Revised: 1: Coding, 2: Arithmetic, 3: Picture completion, 4: Mazes, 5: Picture arrangement, 6: Similarities,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

after multiplying probabilities to gain significance (despite significance levels depending on sample size). In any case, Flynn's tactic of multiplying a number of *p* values to obtain a composite *p* is statistically wrong; the correct method for combining *p* values is provided by R.A. Fisher (1970: pp. 99–101). Nonetheless, Flynn pushed his conclusion to the limit, even to the point of suggesting "a moratorium" on methodology that goes from within-group data to between-group differences!

Flynn's procedure was far too weak for testing between alternative hypotheses, mainly because of his not analyzing most of the data. Here I present an analysis of all the available data. It examines both sets of between-group differences that Flynn discusses (IQ gains over time, Black–White differences) and their relation to two sets of within-group data (inbreeding depression scores from Japan, *g* factor loadings from the U.S.). Whereas the hereditarian

Table 1
Subtests of the Wechsler Intelligence Scale for Children-Revised (WISC-R) arranged in ascending order of inbreeding depression scores (calculated in Japan), with each subtest's reliability, Black–White difference, *g* loading (also from WISC-III), and five sets of gain scores over time

Subtest	Inbreeding depression (*N* = 1,854)	Reliability (*N* = 2,173)	Black–White difference (*N* = 4,848)	WISC-R *g* loading (*N* = 4,848)	WISC-III *g* loading (*N* = 2,200)	U.S. gain 1 (WISC to WISC-R)	U.S. gain 2 (WISC-R to WISC-III)	Germany gain	Austria gain	Scotland gain
Coding	4.45	0.72	0.45	0.37	0.37	2.20	0.70	2.80	1.10	–
Arithmetic	5.05	0.77	0.61	0.61	0.74	0.36	0.30	−0.50	−0.40	−0.90
Mazes	5.35	0.72	0.73	0.40	0.32	–	1.20	–	–	–
Block design	5.35	0.85	0.90	0.63	0.69	1.28	0.90	4.80	1.80	–
Picture	5.90	0.77	0.70	0.53	0.57	0.74	0.90	4.80	0.60	–
Comprehension	6.05	0.77	0.79	0.62	0.62	1.20	0.60	2.40	0.10	3.20
Object	6.05	0.70	0.79	0.53	0.61	1.34	1.20	2.40	1.00	–
Information	8.30	0.85	0.86	0.68	0.70	0.43	−0.30	1.00	0.20	0.70
Picture	9.40	0.73	0.75	0.52	0.47	0.93	1.90	4.80	0.90	–
Similarities	9.95	0.81	0.77	0.65	0.70	2.77	1.30	4.70	2.10	3.70
Vocabulary	11.45	0.86	0.84	0.72	0.73	0.38	0.40	2.60	1.00	0.00

Note: Inbreeding depression scores from Jensen (1983). Reliability of WISC-R from the U.S. standardization data (Wechsler, 1974). Standardized Black–White difference scores from Rushton (1995: p. 187, Tables 9.1) using a weighted average of 5 data sets from Jensen (1985, 1987, Naglieri and Jensen, 1987). WISC-R *g* loadings from Rushton (1995: p. 187, Table 9.1) using a weighted average of 10 sets of *g* loadings from Jensen (1985, 1987, Naglieri and Jensen, 1987). WISC-III *g* loadings based on the U.S. standardization data using a Schmid–Leiman hierarchical factor analysis. U.S. gain 1, U.S. gain 2, German gain, Austria gain, and Scotland gain all from Flynn (1999). The gains are scaled scores based on SD = 3 (rather than 15). They must be multiplied by 5 to get something like IQ gains over time.

model predicts Black–White differences will cluster with the genetically-loaded dimensions like inbreeding depressions and *g* loadings, the environmental interpretation predicts Black–White differences will cluster with the environmentally-loaded variables like IQ gains over time.

2. Method

Table 1 presents the data to be analyzed. The WISC-R subtests are arranged in ascending order of the inbreeding depression scores calculated in Japan. These, and the columns of data on subtest reliabilities, Black–White differences, and *g* loadings are from Rushton (1995: pp. 185–187, as cited by Flynn, 1998) and are the basis of the results shown in Fig. 1. Also in Table 1 are *g* loadings from the WISC-III standardization data (Wechsler, 1991) extracted by the Schmid–Leiman hierarchical factor analysis previously used by Jensen (1985, 1987) with the WISC-R. Table 1 also gives the five sets of gain scores from Flynn (1999). These are the "raw" data for the present paper and relations among the variables will be examined using Pearson product–moment correlations, partial correlations, and principal components analysis.

3. Results

Table 2 presents the results of the Pearson correlations carried out on the data in Table 1, with the zero-order correlations in the top half of the matrix and the first-order partial correlations (controlling for reliability) in the lower half of the matrix. Table 3 presents the results of a principal components analysis carried out on the partialled correlation matrix in Table 2. Two significant components with eigenvalues >1 were extracted. These are presented in both unrotated and varimax rotated forms.

The analyses were repeated in alternative ways and the results were found to be virtually identical. For example, Spearman correlations were carried out on the ranked data in Table 1 and the correlation between this matrix and the Pearson matrix in Table 2 was 0.96. Another alternative procedure was to correct the *g* loadings (both WISC-R and WISC-III) and Black–White differences in Table 1 for attenuation using the reliabilities in column 2 of Table 1 (by dividing the score by the square root of the reliability rather than by the method of partial correlations used above). Again, the principal components remained very similar to the ones reported in Table 3, with congruence coefficients of 0.97 and 0.92, respectively, for the resulting two unrotated components.

4. Discussion

Any conclusion that inbreeding depression scores are irrelevant to the race/IQ debate is contradicted by the complete analysis of these data. The correlational and the principal component analyses in Tables 2 and 3 show that Black–White IQ differences on the WISC-R, like the inbreeding depression scores (a purely genetic effect) and the *g* factor loadings (a largely genetic effect), are unrelated to the IQ gains over time. The results indicate that

Table 2
Pearson correlations of variables from Table 1 using subtests of the Wechsler Intelligence Scale for Children-Revised (zero-order correlations above diagonal; reliabilities partialled 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	1.00	0.50	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 *d*	0.26	–	1.00	0.69	0.53	−0.28	−0.05	0.21	0.22	0.31
WISC-R *g*	0.40	–	0.43	1.00	0.94	−0.38	−0.44	−0.18	−0.04	−0.22
WISC-III *g*	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	0.25	–	0.27	−0.33	−0.50	0.48	0.86	1.00	0.76	0.80
Austria	0.24	–	0.15	−0.32	−0.31	0.79	0.75	0.77	1.00	0.58
Scotland	0.28	–	0.56	−0.06	−0.85	0.85	0.68	0.82	0.64	1.00

Table 3
Principal components analysis and varimax rotation for the matrix of Pearson correlations of variables in Table 2 after the reliability coefficients have been partialled out of the entire matrix

Principal components				
Variables	Unrotated		Varimax Rotated Loadings	
	I	II	1	2
Inbreeding depression	0.31	0.61	0.26	**0.63**
Black–White differences	0.29	0.70	0.23	**0.72**
WISC-R *g* loadings	−0.33	0.90	−0.40	**0.87**
WISC-III *g* loadings	−0.61	0.64	−0.66	**0.59**
U.S. gains 1	0.73	−0.20	**0.75**	−0.13
U.S. gains 2	0.81	0.40	**0.77**	0.47
German gains	0.91	0.03	**0.91**	0.11
Austria gains	0.87	0.00	**0.86**	0.07
Scotland gains	0.97	0.08	**0.96**	0.17
Percent of total variance explained	48.60	25.49	48.44	25.65

different processes likely underlie the two sets of group differences, with genetic factors underlying the race differences and environmental factors underlying the generational changes.

The results provide further evidence for the crucial role played by *g* in the race–IQ debate. In his latest book, *The g Factor*, Jensen (1998b: Ch. 11) summarizes the results from 17 independent data sets on a total of nearly 45,000 Blacks and 245,000 Whites derived from 171 psychometric tests. *g* loadings consistently predict the magnitude of the Black–White difference ($r = +0.63$). This is borne out even among three-year-olds administered eight subtests of the Stanford–Binet. The rank correlation between *g* loadings and the Black–White differences is +0.71 ($p < 0.05$). Even when the *g* loading is calculated from performance on elementary cognitive tasks which correlate with IQ (such as moving the hand to press a button to turn off a light, which all children can do in less than 1 second), the correlations between the *g* loadings of these tasks and the Black–White differences range from +0.70 to +0.81.

Significant correlations occurring between *g*-factor loadings and other variables have been dubbed "The Jensen effect". Jensen effects have been found for a variety of variables, including not only scholastic and work-place performance, but also brain size, brain intracellular pH, brain glucose metabolic rate, average evoked potential, reaction time, and the heritability coefficients of various tests (including the Wechsler tests).

Jensen effects are not omnipresent and their absence can be highly informative, as in the present study. Thus the secular increase in test scores (the "Lynn–Flynn effect") is not a "Jensen effect" nor is this the first time the discriminating power of the Jensen effect has been shown. In an adoption study, no Jensen effect was found for the adopting family environment on the adoptees WISC-R IQ scores, but a clear Jensen effect was found for the biological parents' socioeconomic status (a proxy for their biological IQ) on the adoptees' IQ scores (Jensen, 1998a). These results clearly imply that environmental effects on IQ affect the non-*g* component.

Although the results of this paper remove many remaining doubts about the nature of the Black–White differences, they raise many new questions about the causes of the generational rise in IQ scores. Lynn (1996) attributed the rise in IQ to improved nutrition bringing about enhanced neurological development and larger brain size. In support, Lynn (1996) reviewed the results of six studies of identical twins born with different weights as a result of one twin receiving better nutrition from its mother's placenta than the other. In these cases the heavier twin at birth had greater intelligence in adolescence. Further, the average height, weight, and brain size of the entire population has been increasing in many countries over the last 100 years (Eveleth and Tanner, 1990). For example, male brain weights in England have shown an increase of nearly 7 grams a decade from 1860 to 1940 (Miller and Corsellis, 1977) and similar increments are found internationally (Rushton, 1997a, pp. 283–284). Given the overall correlation of 0.40 between brain size and IQ established by magnetic resonance imaging (reviewed by Rushton and Ankney, 1996; Jensen, 1998b), the rise in IQ scores seems attributable to biological factors.

An alternate hypothesis is hybrid vigor. This is the opposite of inbreeding depression and occurs because genes for intelligence are dominant in the Mendelian sense due to their adaptive value (Jensen, 1983). The more out-breeding there is, the greater is the probability that a dominant gene for high IQ will occur. Hybrid vigor, however, is a much weaker effect than inbreeding depression.

The gains in IQ over time (the Lynn–Flynn effect) are unrelated to *g* — and therefore unlike the Black–White IQ difference. While the nature of the Black–White difference is now more or less known to be due to some aspects of the size and functioning of the brain (Rushton, 1997a; Jensen, 1998b), the secular rise in IQ remains one of the unsolved psychometric mysteries. Both the nutritional and the hybrid vigor hypotheses lead to the expectation of rises in the *g*-factor. But the secular gains are little, if at all related to *g*. Perhaps, then, they are mainly due to gains in the subtest's group factors and test specificity. Practicing copying numbers and letters (coding), vocabulary, and arranging and completing pictures, likely have much non-*g* related specificity variance associated with them as well as small group factors. It is these components of variance that the results of the present study suggest are most subject to the secular rises in IQ.

References

Eveleth, P. B., and Tanner, J. M. (1990). *Worldwide Variation in Human Growth* (2nd ed.). London: Cambridge University Press.

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

Flynn, J. R. (1999). 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.

Jensen, A. R. (1983). The effects of inbreeding on mental ability factors. *Personality and Individual Differences, 4,* 71–87.

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. (1998a). Adoption data and two *g*-related hypotheses. *Intelligence, 25,* 1–6.

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

Lynn, R. (1982). IQ in Japan and the United States shows a growing disparity. *Nature, 297,* 222–223.

Lynn, R. (1996). *Dysgenics*. Westport, CT: Praeger.

Miller, A. K. H., & Corsellis, J. A. N. (1977). Evidence for a secular increase in brain weight during the past century. *Annals of Human Biology, 4,* 253–257.

Naglieri, J. A., & Jensen, A. R. (1987). Comparison of Black–White differences on the WISC-R and K-ABC: Spearman's hypothesis. *Intelligence, 11,* 21–43.

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. (1997a). Race, Evolution, and Behavior: *A Life-History Perspective* (Softcover Edition, with Afterword). New Brunswick, NJ: Transaction.

Rushton, J. P. (1997b). Review essay. (Review of the books The Bell Curve Debate, The Bell Curve Wars, Final Solutions, The Mismeasure of Man, The Nazi Connection, The Race Gallery, and The Science and Politics of Racial Research). *Society, 34,* 78–82.

Rushton, J. P., & Ankney, C. D. (1996). Brain size and cognitive ability: Correlations with age, sex, social class and race. *Psychonomic Bulletin and Review, 3,* 21–36.

Wechsler, D. (1974). *Manual for the Wechsler Intelligence Scale for Children – Revised*. New York: The Psychological Corporation.

Wechsler, D. (1991). *Wechsler Intelligence Scale for Children – Third Edition: Manual*. San Antonio, TX: The Psychological Corporation.