

Japanese Inbreeding Depression Scores: Predictors of Cognitive Differences Between Blacks and Whites

J. PHILIPPE RUSHTON
University of Western Ontario

Previous work has shown that black-white racial group differences in intelligence are largest on the most *g*-loaded subtests. Because these are the ones most genetically influenced, it is implied that the racial differences in intelligence are partly genetic in origin. To test this hypothesis, degrees of genetic influence were estimated on the WISC subtests from inbreeding depression scores calculated on cousin marriages in Japan, and correlated with the magnitude of the black-white difference. The resulting average Spearman ρ was $+ .47$ ($p < .05$; Pearson $r = .41$, $p < .05$), and after correction for subtest reliability was $+ .39$ ($p < .05$; Pearson $r = .35$, $p < .10$). Partialling out reliability rather than correcting for it reduced the average Pearson correlation to $.22$. These results suggest that the genetic contribution to racial differences in cognitive performance may be more robust across populations, languages, time periods, and measurement specifics than has been considered to date.

Inter- and intranational average differences are observable on measures of cognitive performance and educational attainment such that Mongoloids $>$ Caucasoids $>$ Negroids (Jensen, 1985; Lynn, 1987). While the nature and the causes of the differences remain much debated (for one exchange of views, see Rushton, 1988a, 1988b; Zuckerman & Brody, 1988), data suggest that, in addition to the environmental contribution, some of the between-group variance may be genetic in origin. The group differences are greatest, for example, on the most *g*-loaded subtests (Jensen, 1985, 1987a), which are the ones independently found to be the most genetically influenced (Jensen, 1987b; Vernon, in press). This paper will briefly overview the evidence on this topic and then present additional data testing the hypothesis that racial differences in cognitive performance are partly genetic.

In support of the view that the races are most differentiated on the more *g*-loaded subtests, Jensen (1985) examined 11 large-scale studies, each comprising anywhere from 6 to 13 diverse tests, and showed that a significant and substantial

This paper was completed while the author was a Fellow of the John Simon Guggenheim Memorial Foundation. Thanks are due to D.N. Jackson, A.R. Jensen, and P.A. Vernon for useful discussions.

Correspondence and requests for reprints should be sent to J.P. Rushton, Department of Psychology, University of Western Ontario, London, ON, N6A 5C2 Canada.

correlation was found in each between the test's g -loading and the mean black–white difference on the same tests. Subsequently, Naglieri and Jensen (1987; see also Jensen, 1987a) matched 86 black and 86 white children in the 4th and 5th grades for age, school, sex, and socioeconomic status (SES) and tested them with the Wechsler Intelligence Scale for Children—Revised (WISC-R) and the Kaufman Assessment Battery for Children (K-ABC) for a total of 24 subtests. They found that the standardized black–white differences on the various tests were positively correlated ($r = +.78$, $\rho = +.75$) with the tests' g -loadings.

That g -loadings correlate with estimates of genetic influence has been shown in several studies. Thus, Jensen (1983) reported a correlation of $+.81$ between the degree of inbreeding depression (defined as a lowered mean of the trait relative to the mean in a non-inbred population) from cousin marriages in Japan and the g -loadings of the 11 subtests of the WISC. This is especially interesting because inbreeding depression indicates genetic dominance, which arises when the trait confers evolutionary fitness (defined as increased survival of genes through successive generations). Subsequently, Jensen (1987b) reported rank order correlations of $+.55$ and $+.62$ between estimates of genetic influence from two twin studies and the g -loading of the Weschsler Adult Intelligence Scale (WAIS) subtests, and Vernon (in press) found a correlation of $+.60$ between the heritabilities of a variety of speed of reaction time tasks and their relationship with the g -loadings from a psychometric test of general intelligence.

Three studies reviewed by Jensen (1973, Chapter 4) provide more immediate evidence that the average black–white difference on diverse mental tests is chiefly a difference in heritable g . First an “environmentality index” (the opposite of a heritability index, based on the degree to which sibling correlations departed from the pure genetic expectation of $.5$) was computed and shown to be inversely related to the magnitude of the black–white difference on various subtests. This observation is striking because it results from a *differential* prediction: Environmental theory would have predicted a positive correlation between the strength of measured environmental effects and the magnitude of the black–white differences. Second, an unpublished doctoral thesis by Nichols (1972) was cited reporting a correlation of $+.67$ between a subtest's heritability and the degree to which it differentiated whites and blacks, thereby conceptually replicating what was found in the first study. Third, racial differences in regression to the mean from the scores of siblings were examined. Black and white children matched for high IQ showed sibling regressions approximately halfway to their respective population means rather than to the mean of the combined populations, with the opposite pattern observed with black and white children matched for low IQ. The regression line showed no significant departure from linearity throughout the range from IQ 50 to 150, a phenomenon predicted by genetic theory.

It seems odd that these pathbreaking approaches by Jensen (1973) have not

been followed up since they seem to bear directly on the issue of whether the racial differences are based in evolution as well as in culture. One possible source of artifact in the above studies, however, concerns the differential reliability of the subtests. Because a test's reliability can affect both its heritability and its loading on other variables, it could be argued that some of the observed relationships might have occurred because the variables involved are correlated with a third variable, namely test reliability. A test with low reliability, for example, would have both a low heritability and a low black-white difference score, thus producing a spurious correlation between the two. It is unclear how much this source of artifact was controlled for in the studies cited above. Nonetheless their results do seem to indicate a partly genetic causality.

The research strategy employed by Jensen (1973) to test the hypothesis that black-white differences in cognitive ability are genetic in origin involves the concept of "differential heritability," that is, that some measures are more genetically penetrated than others. Among anthropometric measures, for example, it might be expected that wrist size is more heritable than bicep size because osseous parts of the body are less susceptible to environmental modification than are muscular parts. Although the question of whether personal attributes are differentially heritable is not considered to be conclusively answered (Loehlin, Willerman, & Horn, 1988), and while it is usual to consider estimates of genetic influence to be properties of particular populations and not to be highly generalizable (Falconer, 1981), Jensen's results suggest that estimates of genetic influence may be sturdier than is often thought to be the case. In the study to be reported here, a particularly strong test is made of the generalizability of genetic estimates. Japanese children tested in the 1950s will be compared with American children tested in the 1970s on a revised version of the test given to the Japanese. This makes for a conservative test of the genetic-g hypothesis for the predicted effect has to be sufficiently robust to overcome these differences.

METHOD

Figures on inbreeding depression are used to estimate genetic influence. These were reported by Jensen (1983, Table 2) for 11 WISC subtests derived from a study by Schull and Neel (1965) who calculated them from 1854 7- to 10-year-old Japanese children administered a Japanese version of the WISC in 1958 and 1960. Since almost 50% of the sample involved a degree of inbreeding (1st cousins, 1st cousins once-removed, and 2nd cousins), it was possible to assess the degree of inbreeding depression on each subtest, expressed as the percentage decrement in the score per 10% increase in degree of inbreeding. These were calculated after statistically controlling for child's age, birth rank, and month of examination, and eight different parental variables, mostly pertaining to SES, by means of multivariate regression analysis (see Jensen, 1983). Standardized

black–white difference scores on the same tests were available from five somewhat overlapping U.S. studies, four of which were summarized by Jensen (1985, Table 5), and the fifth of which was summarized by Jensen (1987a). These samples are described below. The Japanese children were tested in the 1950s on the WISC and the American children were tested in the 1970s on the WISC-R. Also reported by Jensen (1985, Table 5) were the reliabilities of the subtests. All these data are shown in Table 1.

Sample 1

Jensen and Reynolds (1982) used a national standardization sample for the WISC-R, selected by a stratified random sampling procedure to be representative of the entire U.S. population based on the 1970 census. Black and white age-matched samples of children between the ages of 6 and 16½ years were tested. The white sample size was 1,868, and the black 305, for a total $N = 2173$.

Sample 2

Reynolds and Gutkin (1981) chose that subset of 570 individuals from Sample 1 that enabled them to make exact matches of 285 black testees with 285 white testees on four demographic variables (sex, SES, geographical region of residence, and urban vs. rural residence). As in Sample 1, the children ranged in age from 6 to 16½.

Sample 3

Sandoval (1982) took as subjects a subset of the children used to standardize the System of Multiculture Pluralistic Assessment (SOMPA), which includes the WISC-R. (These subjects are independent of the WISC-R national standardized sample.) The total sample was selected by a random-stratified (by sex, ethnicity, age, locality) sampling procedure to be representative of the 1973–1974 California elementary-school-age population aged 5 to 11. Mexican-American children were not included in the analysis. There were 332 whites and 314 blacks for a total $N = 646$.

Sample 4

Mercer (1984) examined all the black and white children used in the standardization procedures mentioned for Sample 3, randomly selected from all 5- to 11-year-olds in the California School population in 1973–1974. Altogether there were 618 whites and 619 blacks for a total sample of 1,237.

Sample 5

Naglieri and Jensen (1987) reported data from a study of 172 (86 black and 86 white) children in the 4th and 5th grades of Columbus, Ohio, matched on age, school, sex, and SES. All children were tested on the 11 subtests of the WISC-R and 13 subtests of the K-ABC.

TABLE 1
Inbreeding Depression Scores on WISC from Japan, and Black-White Difference Scores and Test Reliabilities on WISC-R from the United States

WISC Subtest	Inbreeding Depression (N = 1854)	Reliability	Black-White Standardized Differences				
			Sample 1 (N = 2173)	Sample 2 (N = 570)	Sample 3 (N = 646)	Sample 4 (N = 1237)	Sample 5 (N = 172)
Information	8.30	85	81	69	93	101	54
Similarities	9.95	81	79	53	84	82	67
Arithmetic	5.05	77	61	48	63	68	43
Vocabulary	11.45	86	88	67	78	90	53
Comprehension	6.05	77	94	80	65	65	47
Picture Completion	5.90	77	79	61	63	65	41
Picture Arrangement	9.40	73	77	65	76	79	38
Block Design	5.35	85	93	73	96	89	80
Object Assembly	6.05	70	82	64	81	82	66
Coding	4.45	72	47	39	50	46	07
Mazes	5.35	72	69	59	83	81	—

RESULTS

Spearman and Pearson correlations between the estimates of genetic penetrance and the magnitude of the black–white differences were computed both before and after correcting for reliability (by dividing the black–white difference scores by the square root of the test’s reliability, and also, in the case of the Pearson, through partial correlation). These data and results are shown in Table 2. It can be argued that Spearman’s ρ may be the more appropriate statistic since this rests only on ranks and is not dependent on possible outliers or on the scalar properties of the inbreeding estimates (Jensen, 1983, 1985).

Taking a weighted average of these correlations using the total sample sizes and testing their significance using Fisher’s test for combining probabilities (Winer, 1971), the Spearman ρ between inbreeding depression and the black–white average difference before correction is $+ .47$ ($p < .05$; Pearson’s $r = + .41$, $p < .05$) and after correction is $+ .39$ ($p < .05$; Pearson’s $r = + .35$, $p < .10$). Extremely similar figures are found if the correlations are weighted by the black n , or if indeed, unweighted averages are used. Partialling out reliability rather than correcting for it reduced the average Pearson correlation to $+ .22$. As can be seen in Table 2, reliability correlates $\rho = + .45$ and $r = + .50$ with the inbreeding depression scores, as it also did with the various black–white difference scores which all correlated with each other at about the same level of magnitude.

TABLE 2
Correlations between Inbreeding Depression Scores and Black–White Difference Scores

Black–White Difference	Inbreeding Depression		
	Spearman	Pearson	Pearson Partialled
Reliability	.45†	.50†	—
Sample 1	.47†	.40	.20
Sample 2	.43†	.25	.07
Sample 3	.38	.37	.16
Sample 4	.56*	.56*	.38
Sample 5	.32	.26	.01
Corrected Sample 1	.42†	.33	.21
Corrected Sample 2	.32	.17	.08
Corrected Sample 3	.27	.30	.18
Corrected Sample 4	.46†	.51†	.39
Corrected Sample 5	.27	.23	.03

† $p < .10$; * $p < .05$.

DISCUSSION

A range of interpretations is possible concerning the strength of the effects found in these data, ranging from "very weak" to "very strong." The most stringent conclusion would be based on the five Pearson product-moment correlations between inbreeding depression scores and the estimates of the black-white difference when partialling out reliability (range of $+ .01$ to $+ .38$, weighted mean $= + .22$). Some readers might interpret this as completely nonsignificant. This, however, could be a mistake, for if the results are occurring through pure chance, then an equal number of negative correlations should occur as positive correlations. Yet 5/5 are in the predicted direction. The binomial probability of finding this pattern by chance is less than 5 in 100. Moreover, it could be argued that attempting to control reliability with a partial correlation results in a gross "over-correction" and that despite this, a residual of "true score variance" remains in each of the five correlations, which is the reason they show the same direction on each of the five estimates. Of course it is always possible that some factor other than reliability could account for these data, but if so it is not obvious what it would be. As described above, eight different SES variables were statistically controlled in computing the inbreeding depression scores (see Jensen, 1983).

Additional support for the reality of the effects comes from the observation that the magnitudes of the predictions of the black-white differences, even after partialling out reliability, are correlated with the sample sizes. It is a truism that the potential to detect genetic effects increases with the size of the sample. If the relation between sample size and effect size was chance, this would not occur. Since the reliability of the point estimations is more important than the reliability of the subtests, which can be compensated for by larger sample sizes, it would seem reasonable to conclude that there is a relation between the magnitude of genetic influence on the trait and the magnitude of the black-white difference. It might be noted too, though, that the five samples were not completely independent of each other, which further complicates the interpretation of the "strength" (as opposed to the direction) of these effects.

The results reported here, however, are also compatible with other data showing that racial-group differences in behavior are partly genetic in origin (Rush-ton, 1988a, 1988b). As a necessary preliminary, many of the variables on which the populations differ in central tendency are found to be substantially heritable across both Caucasian and non-Caucasian populations, often to a remarkably robust degree (Rush-ton, in press; Tellegen et al., 1988). The heritability of racial differences has also been suggested more directly. The follow-up adoption study of black children raised by white parents, for example, shows that whereas at age 7 black IQ is comparable to white IQ, by age 17 black IQ, educational achievement, social deviancy, and psychopathology shows regression to the black population mean (Scarr, Weinberg, & Gargiulo, 1987).

The methodology that we have employed to “probe the genes” is undoubtedly crude. Examining chromosomes and relating the resultant metrics to behavior would be a more direct method. Recent studies examining genetic distances between various populations from mtDNA and nuclear DNA may soon have important applications for field studies of human populations. Summarizing evidence derived from these and other techniques, Stringer and Andrews (1988) show an African origin for *Homo sapiens* of about 200,000 years ago, with a divergence time of about 110,000 years ago for the Negroid–non-Negroid split and about 41,000 years ago for the Caucasoid–Mongoloid split.

Rough-hewn though our techniques may have been, they do provide another item of evidence inclining to the conclusion that genetic as well as environmental influences may contribute to the black–white difference in cognitive performance. That estimates of genetic influence calculated from inbreeding depression scores in Japanese cousin marriages on the WISC in the 1950s predict the magnitude of black–white differences in the United States on the WISC-R in the 1970s, a relationship that does not entirely disappear when disattenuating the difference scores, implies they have a greater robustness across populations, languages, time periods, and measurement specifics than has been considered to date.

REFERENCES

- Falconer, D.S. (1981). *Introduction to quantitative genetics* (2nd ed.). London: Longman.
- Jensen, A.R. (1973). *Educability and group differences*. New York: Harper & Row.
- Jensen, A.R. (1983). 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 (with commentaries and response). *Behavioral and Brain Sciences*, 8, 193–263.
- Jensen, A.R. (1987a). Further evidence for Spearman’s hypothesis concerning black–white differences on psychometric tests. *Behavioral and Brain Sciences*, 10, 512–519.
- Jensen, A.R. (1987b). The *g* beyond factor analysis. In R.R. Ronning, J.A. Gover, J.C. Conoley, & J.C. Witt (Eds.), *The influence of cognitive psychology on testing*. Hillsdale, NJ: Erlbaum.
- Jensen, A.R., & Reynolds, C.R. (1982). Race, social class and ability differences on the WISC-R. *Personality and Individual Differences*, 3, 423–438.
- Loehlin, J.C., Willerman, L., & Horn, J.M. (1988). Human behavior genetics. In M.R. Rosenzweig & L.W. Porter (Eds.), *Annual review of psychology* (Vol. 39, pp. 101–133). Palo Alto, CA: Annual Reviews.
- Lynn, R. (1987). The intelligence of the Mongoloids: A psychometric, evolutionary and neurological theory. *Personality and Individual Differences*, 6, 441–452.
- Mercer, J.R. (1984). What is a racially and culturally nondiscriminatory test? In C.R. Reynolds & R.T. Brown (Eds.), *Perspectives on bias in mental testing*. New York: Plenum.
- Naglieri, J.A., & Jensen, A.R. (1987). Comparison of black–white differences on the WISC-R and the K-ABC: Spearman’s hypothesis. *Intelligence*, 11, 21–43.
- 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.

- Reynolds, C.R., & Gutkin, T.B. (1981). Multivariate comparison of the intellectual performance of blacks and whites matched on four demographic variables. *Personality and Individual Differences*, 2, 175–180.
- Rushton, J.P. (1988a). Race differences in behaviour: A review and evolutionary analysis. *Personality and Individual Differences*, 9, 1009–1024.
- Rushton, J.P. (1988b). The reality of racial differences: A rejoinder. *Personality and Individual Differences*, 9, 1035–1040.
- Rushton, J.P. (in press). On the generalizability of genetic estimates. *Personality and Individual Differences*.
- Sandoval, J. (1982). The WISC-R factorial validity for minority groups and Spearman's hypothesis. *Journal of School Psychology*, 20, 198–204.
- Scarr, S., Weinberg, R.A., & Gargiulo, J. (1987, June). *Transracial adoption: A ten-year follow-up*. Abstract in Program of the 17th Annual Meeting of the Behavior Genetics Association, Minneapolis, MN.
- Schull, W.J., & Neel, J.V. (1965). *The effects of inbreeding on Japanese children*. New York: Harper & Row.
- Stringer, C.B., & Andrews, P. (1988). Genetic and fossil evidence for the origin of modern humans. *Science*, 239, 1263–1268.
- Tellegen, A., Lykken, D.T., Bouchard, T.J., Jr., Wilcox, K.J., Segal, N.L., & Rich, S. (1988). Personality similarity in twins reared apart and together. *Journal of Personality and Social Psychology*, 54, 1031–1039.
- Vernon, P.A. (in press). The heritability of measures of speed of information processing. *Personality and Individual Differences*.
- Winer, B.J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.
- Zuckerman, M., & Brody, N. (1988). Oysters, rabbits and people: A critique of "Race differences in behaviour" by J.P. Rushton. *Personality and Individual Differences*, 9, 1025–1033.