Rushton, J.P. (2003). Race differences in g and the "Jensen Effect". In H. Nyborg (Ed), The scientific study of general intelligence: A tribute to Arthur R. Jensen (pp. 147-186). London: Elsevier.

Chapter 9

Race Differences in *g* and the "Jensen Effect"

J. Philippe Rushton

1. The Spearman-Jensen Hypothesis

Jensen (1980: 535) formally designated the view that Black-White differences were largely a matter of *g* as "Spearman's hypothesis", because Spearman (1927: 379) was the first to suggest it. Subsequently, Osborne (1980a) dubbed it the "Spearman-Jensen hypothesis" because it was Jensen who brought Spearman's hypothesis to widespread attention, and it was Jensen who did all the empirical work confirming it. More recently, to honor one of the great psychologists of our time, Rushton (1998) proposed that the term "Jensen Effect" be used whenever a significant correlation occurs between *g*-factor loadings and any variable, X; otherwise there is no name for it, only a long explanation of how the effect was calculated. Jensen Effects are not omnipresent and their absence can be as informative as their presence. For example, the "Flynn Effect" (the secular rise in IQ) is probably not a Jensen Effect because it does not appear to be on *g*.

The Black-White difference on the *g*-factor is the best known of all the Jensen Effects. The reason Jensen pursued Spearman's (1927) hypothesis was because it so exquisitely solved a problem that had long perplexed him. The average 15 to 18 IQ point difference between Blacks and Whites in the U.S. had not changed since IQ testing began almost 100 years ago. But Jensen (1969a) noted that the race differences were markedly smaller on tests of rote learning and short-term memory than they were on tests of abstract reasoning and transforming information. Moreover, culture-fair tests tended to give Blacks slightly *lower* scores than did conventional tests, as typically did non-verbal tests compared with verbal tests. Furthermore, contrary to purely cultural explanations, race differences could be observed as early as three years of age, and controlling for socioeconomic level only reduced the race differences by 4 IQ points.

Jensen (1968) initially formalized these observations in his so-called Level I-Level II theory. Level I tasks were those that required little or no mental manipulation of the input in order to arrive at the correct response whereas Level II tasks required mental manipulation. A classic example of Level I ability is Forward Digit Span in which people recall a series of digits in the same order as that in which they are presented. A

The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen

ISBN: 0-08-043793-1

clear example of a Level II task is Backward Digit Span in which people recall a series of digits in the *reverse* order to that in which they were presented. Jensen found that Black-White differences were twice as large for Backward as for Forward Digit Span.

After Jensen (1980) re-read Spearman, he realized that the Black-White differences (and his Level I-Level II formulation) were specific examples of the more general hypothesis proposed by Spearman (1927: 379), namely that the Black-White difference "was most marked in just those [tests] which are known to be saturated with *g*". It was Spearman (1904, 1927), of course, who had generated the seminal concept of *g* in the first place. The *g* factor, derived from factor analysis of the correlations among a number of tests of mental abilities, is typically the largest factor.

To test Spearman's hypothesis, Jensen developed the *method of correlated vectors*. Essentially, this method correlates the standardized Black-White mean differences on a set of cognitive tests (a vector of scores, i.e. possessing both direction and quantity), with the tests' *g* loadings (a second vector of scores). A positive and substantial correlation provides support for Spearman's hypothesis. The rationale is straightforward: if *g* is the main source of between- *and* within-group differences, then there should be a positive relationship between a test's *g*-loading and the Black-White difference on the test; the more *g*-loaded the test, the greater the Black-White difference on that test. A methodological corollary is the prediction that when the point-biserial correlations of race (Black-White) with a number of diverse cognitive tests are entered into the total matrix of correlations among all the tests, the race variable will have its largest loading on the general factor of the correlation matrix.

According to Jensen (1998: 372–373), an ideal test of Spearman's hypothesis using the method of correlated vectors, must meet several methodological requirements. These are: (1) the samples being compared must be representative of their respective populations; (2) the samples being compared must be large enough to overcome the sampling error of the correlations among tests; (3) the samples being compared must not be selected on the basis of any *g*-loaded criterion; (4) the *g* factor should be extracted from enough tests to be reliable, as would be indicated by high coefficients of congruence in independent samples from the same population; (5) any test showing psychometric test bias for the groups being compared must be excluded; (6) the tests must be sufficiently diverse to allow significant differences between their *g* loadings; (7) the scores must be corrected for reliability; (8) the *g* values must be computed separately in the different groups; (9) the scores must measure the same latent traits in the different groups (i.e. the vector of *g* loadings extracted separately from each group must show a high congruence coefficient); and (10) the hypothesis must be tested for statistical significance by both Pearson's *r* and Spearman's rank-order correlation, rho.

As also noted by Jensen, tests of Spearman's hypothesis are necessarily stringent because the degrees of freedom used for statistical rejection of the null hypothesis are based on the number of pairs of variables in the correlated vectors (e.g. 13 sub-tests from the Wechsler Scales) and not on the subject sample size. It is also worth emphasizing that Spearman's hypothesis concerns the *relative* magnitude of the group difference across various tests that differ in their *g* loadings and not the *absolute* magnitude of group differences. It is therefore conceptually independent of any secular trend in absolute test scores, viz., the Flynn Effect (discussed below).