



Jensen Effects among African, Indian, and White engineering students in South Africa on Raven's Standard Progressive Matrices

J. Philippe Rushton^{a,*}, Mervyn Skuy^{b,1}, Peter Fridjhon^b

^a*Department of Psychology, The University of Western Ontario, London, Ontario, Canada N6A 5C2*

^b*Division of Specialized Education, The University of the Witwatersrand, Johannesburg 2050, South Africa*

Received 28 June 2001; received in revised form 29 August 2001; accepted 22 January 2002

Abstract

Low test scores are routinely observed in sub-Saharan African populations. In this paper, we explore the topic further by examining Rushton and Skuy's [Intelligence 28 (2000) 251] hypothesis that a bimodal distribution exists in the African population with a high-scoring group virtually indistinguishable from Whites, and a low-scoring group performing significantly below both Whites and the higher-scoring African group. To test this hypothesis, we sought out a potentially higher-scoring African population than has previously been studied. We administered untimed Raven's Standard Progressive Matrices (SPM) to 342 17- to 23-year-olds in the Faculties of Engineering and the Built Environment at the University of the Witwatersrand in Johannesburg (198 Africans, 86 Whites, 58 Indians; 71 women, 271 men). Out of the 60 total problems, the African students solved an average of 50, the Indian students, 53, and the White students, 56 ($P < .001$). On the 1993 US norms, Africans were at the 41st percentile, Indians at the 55th, and Whites at the 75th, with IQ equivalents of 97, 102, and 110, respectively. The African–Indian–White differences were most pronounced on those items with the highest item-total correlations, indicating a difference in g , or the general factor of intelligence. Hence, they were “Jensen Effects.” Indeed, the g loadings showed a small degree of cross-cultural generality; for example, item-total correlations calculated on the Indian students predicted the magnitude of the White–African differences. When the 60 items were aggregated into 10 “subtests,” the magnitude of the Jensen Effect was similar to that from previous studies based on

* Corresponding author. Tel.: +1-519-661-3685; fax: +1-519-850-2302.

E-mail addresses: Rushton@uwo.ca (J.P. Rushton), 135skuy@mentor.edcm.wits.ac.za (M. Skuy).

¹ Tel.: +11-27-11-716-5286; fax: +11-27-11-339-3844.

whole subtests (median $\rho=.53$). There were no sex differences. Nor did this study of African engineering students support the idea of a bimodal distribution.

© 2002 Elsevier Science Inc. All rights reserved.

Keywords: Intelligence tests; Race differences; Raven's Progressive Matrices; South Africa; African IQ; Scores

1. Introduction

For nearly 100 years, the average mean score on intelligence tests for African Americans has been about 18 points [1.2 standard deviations (S.D.)] lower than for European Americans (Herrnstein & Murray, 1994; Jensen, 1998; Rushton, 2000). Whatever their causes turn out to be, the Black–White differences are higher on tests of high-*g* than they are on tests of low-*g*, where *g* is the general factor of intelligence. Spearman (1927, p. 379) was the first to hypothesize that Black–White differences would be “most marked in just those [tests], which are known to be saturated with *g*.” This led Jensen (1980, p. 535) to formally designate it as “Spearman’s hypothesis.” Osborne (1980) then 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, Rushton (1998) proposed that, when a significant correlation occurs between *g* factor loadings and variable X, the result be termed a “Jensen Effect,” because otherwise there is no name for it, only a long explanation of how the effect was achieved.

The Black–White difference on the *g* factor is the best known of all the Jensen Effects. In his latest book, *The g Factor*, Jensen (1998, Chapter 11) summarized 17 independent data sets of nearly 45,000 Blacks and 245,000 Whites derived from 149 psychometric tests in which *g* loadings consistently predicted the magnitude of the Black–White difference ($r=.62$). This was borne out even among 3-year-olds administered eight subtests of the Stanford–Binet. The rank correlation between *g* loadings and the Black–White differences was .71 ($P<.05$). Even when the *g* loading is calculated from performance on elementary cognitive (reaction-time) 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 s), the correlations between the *g* loadings of these tasks and the Black–White differences range from +.70 to +.81.

Studies elsewhere in the world also support Spearman’s hypothesis. In the Netherlands, immigrants from the Black Caribbean (and from North Africa) score at least 1 S.D. behind the Dutch majority population. The difference turns out to be on the *g* factor (te Nijenhuis & van der Flier, 1997). In South Africa, Lynn and Owen (1994) administered the Junior Aptitude Test, a paper-and-pencil test consisting of 10 subtests (4 verbal, 6 nonverbal), to 1056 White, 1063 Indian, and 1093 Black 16-year-old high school students. They found a 2 S.D. difference between the Africans and Europeans (yielding an average African IQ of about 70) and a 1 S.D. difference between the Whites and Indians (yielding an average Indian IQ of 85). Lynn and Owen tested Spearman’s hypothesis and found the African–White differences correlated .62 ($P<.05$) with the *g* factor extracted from the African sample (although only .23 with *g* extracted from the European sample). They did not find the White–Indian differences were on the *g* factor.

It is surprising that more research on the *g* factor has not been carried out in Africa because of the low test scores obtained there. These came to widespread attention in the US when *The Bell Curve* (Herrnstein & Murray, 1994, p. 288) examined the conjecture: “[T]he African Black population has not been subjected to the historical legacy of American Black slavery and discrimination and might therefore have higher scores.” However, Black Africans turned out to have, on average, substantially lower scores than African Americans. *The Bell Curve* cited a review by Lynn (1991) of 11 studies from East, West, and Southern Africa reporting an average IQ of 70 (median = 75), 15 points (1 S.D.) lower than the mean of 85 typically found for African Americans and 30 points (2 S.D.) lower than the mean of 100 typically found for White populations. More recent analyses and reviews by Lynn (1997; Lynn & Vanhanen, 2002), and others (Rushton & Skuy, 2000) have found low scores to be ubiquitous throughout Africa.

In South Africa, university students also have low mean test scores. At the University of Venda, for example, a historically disadvantaged African university in the rural Northern Province, Grieve and Viljoen (2000) found 30 students in fourth-year law and commerce averaged a Raven’s Standard Progressive Matrices (SPM) score of 37 out of 60. By the standards of the 1993 US normative sample for 18- to 22-year-olds, this placed them at the 7th percentile with an IQ equivalent of 78 (Raven et al., 1990, p. 98; Raven, Court, & Raven, 1998, p. 77). At the University of the North, another historically disadvantaged university in the Northern Province, Zaaiman, van der Flier, and Thijs (2001) found the highest scoring African sample to date: 147 first-year math and science students who scored 52 out of 60 on the Raven’s SPM, placing them at the 50th percentile, with an IQ equivalent of 100. Their relatively high mean score may have been because they were math and science students, and also because they had been chosen for admission to the university from a pool of 700 on the basis of a math and science selection test.

Three separate studies have now supported Lynn and Owen’s (1994) finding of a Jensen Effect in South Africa. In the first, Rushton and Skuy (2000) gave 309 17- to 23-year-old first-year psychology students at the University of the Witwatersrand the Raven’s SPM without time limits. The 173 African students solved an average of 44 of the 60 problems, whereas the 136 White students solved an average of 54 of the 60 problems. These scores placed the African students at the 14th percentile and the White students at the 61st percentile, which yielded IQ equivalents of 84 and 104, respectively. Because the total score on the Raven’s is an excellent measure of *g* (Jensen, 1980, pp. 645–648), the item-total correlation is a good estimate of the item’s *g* loading. Rushton and Skuy found the item-total correlations correlated positively and significantly with the standardized differences in the proportions of Africans and Whites passing the same items, using both the African item-total correlations, $r = .39$ ($P < .01$, $N = 58$, with $\rho = .43$, $P < .01$), and the White item-total correlations, $r = .34$ ($P < .01$, $N = 46$, $\rho = .41$, $P < .01$).

In a second study, Rushton (2001) examined already published data on 154 African high school students from Soweto, Johannesburg by Skuy, Schutte, Fridjhon, and O’Carroll (2001) showing scores 1–2 S.D. below American norms on a variety of tests including the Wechsler Intelligence Scale for Children—Revised (WISC-R). Rushton calculated the mean African–White differences (using the US standardization mean of 10), and expressed them in S.D.

units, using the African S.D. Extracting the g loadings from the WISC-R national standardization data, the column vector of the subtests' g loadings correlated $r = .77$ ($P < .05$) with the column vector of the standardized African/White differences, showing the Jensen Effect. The Jensen Effect remained even after the Vocabulary subtest was excluded because English was not the students' first language ($r = .66$, $P < .05$), or when g was extracted from the African rather than from the White standardization sample ($r = .60$, $P < .05$), or if Spearman's ρ was used instead of Pearson's r ($\rho = .74$ and $.74$, respectively, $P < .005$).

In a third study, Rushton (in press) reanalyzed published data from Owen (1992) who had given the Raven's SPM without time limits to 1056 Whites, 1063 Indians, 778 mixed-race "Colored," and 1093 Black South African 14-year-olds. Out of a total of 60 items, Whites averaged 45 correct, East Indians, 42, Coloreds, 37, and Blacks 28, placing them at the 57th, 42nd, 19th, and 5th percentiles, respectively, yielding IQ equivalents of 103, 97, 87, and 75 (by the 1993 US norms; Raven et al., 1990, p. 98, 1998, p. 77). Owen found the item-total correlations correlated with the pass rate differences between the ethnic groups on the same items, which he interpreted as indicating an absence of test bias. Following Rushton and Skuy's (2000) analysis above, this might indicate the group differences are on g . To test this possibility, Rushton reanalyzed Owen's data using a purely nonparametric procedure. He found that, indeed, the more highly correlated an item was with g , the more it predicted the differences between the (now ranked) item pass rates for Africans, Whites, Indians, and Coloreds (Spearman's ρ from .35 to .85; all $P < .01$). The effects remained regardless of from which ethnic group the item g loadings were taken. All the ethnic group differences were Jensen Effects.

It is difficult to believe that the African population has a "true" mean IQ of 70. Or that African university students, who likely range 1–2 S.D. above the general population, have mean IQ scores of 78–100. Based on their data, Rushton and Skuy (2000) hypothesized that the distribution of IQ scores in Africa is bimodal, with a high-scoring group virtually indistinguishable from Whites, and a low-scoring group performing significantly below both Whites and the higher-scoring African group. Rushton and Skuy (p. 256) wrote, "The secondary peak of high scores for the Africans [p. 258, Fig. 1] suggests a possible bimodal distribution."

To examine these issues further, we follow up the finding by Zaaiman et al. (2001) that a selected group of Black African math and science students at South Africa's University of the North scored higher than most others on the Raven's SPM, with an IQ equivalent of 100. We tested African students enrolled in the Faculty of Engineering and the Built Environment at South Africa's internationally renowned University of the Witwatersrand, where university level mathematics is required. We also examined further whether any differences would be found on the g factor.

2. Method

2.1. Overview

The primary purpose was to examine performance on the Raven's SPM in an African sample normally expected to score very much above the general South African population

mean. In the US, engineers score among the very highest on tests, such as the Scholastic Aptitude Test (SAT) and the Graduate Record Examination (GRE). For example, the mean Verbal + Quantitative + Analytic scores of engineering students on the GRE is about 1800, whereas for psychology and education students it is about 1500, a difference of about 1 S.D. (Educational Testing Service, 1998). In turn, students entering university to study psychology score about 1 S.D. above the general population. Thus, engineers may be up to about 2 S.D. above the general population. First-year students from the Faculties of Engineering and the Built Environment at the University of the Witwatersrand can be regarded as being among the highest academically achieving of all African students. Participants were paid about 50 rand (at that time about US\$8).

2.2. Subjects

An initial pool of 363 subjects was reduced to 342 17- to 23-year-olds by eliminating small or ambiguous categories. Excluded were those who self-identified as “Colored” ($n=7$), “Other” ($n=8$), those who listed their age as over 23 ($n=2$), or who failed to give biographical data ($n=4$). The analyses were conducted on 198 Africans (155 men, 43 women), 58 Indians (41 men, 17 women), and 86 Whites (75 men, 11 women).

2.3. Test instruments

The Raven’s SPM is the most well known, most researched, and most widely used of all culture-reduced tests (Raven et al., 1998). It consists of 60 diagrammatic puzzles, each with a missing part that the test taker attempts to identify from several options. The 60 puzzles are divided into five sets (A, B, C, D, and E) of 12 items each. To ensure sustained interest and freedom from fatigue, each problem is boldly presented, accurately drawn, and, as far as possible, pleasing to look at. No time limit is set and all testees are allowed to complete the test. As an untimed “capacity” test, and even as a 20-min “speed” or “efficiency” test, the SPM is usually regarded as a good measure of the nonverbal component of general intelligence rather than of culturally specific information. It has been found to demonstrate reliability and validity across a wide range of populations, with retest reliabilities of .83–.93 over a 1-year interval. Internal consistency coefficients of .80 have been found across many cultural groups, including South African Blacks. The total score is also a very good measure of g , the general factor of intelligence, at least within the US (Jensen, 1980).

2.4. Procedure

The Raven’s was administered without any time limits (up to 1 h), but was typically completed within 30 min. Professor Skuy, one of the authors, and his colleagues carried out the testing in a classroom. All students appeared well motivated. The instructions requested students to wait quietly at their desks if they finished before 30 min. After 30 min, however, they could come to the front of the room, hand in their answer sheets and test booklets, and receive payment.

Table 1
Proportion of sample selecting the correct answer on items of the Raven's Standard Progressive Matrices by race

Set A			Set B			Set C			Set D			Set E							
Item	African	Indian	White	Item	African	Indian	White	Item	African	Indian	White	Item	African	Indian	White	Item	African	Indian	White
1	1.00	1.00	1.00	13	1.00	.98	1.00	25	.99	1.00	1.00	37	.97	.98	1.00	49	.88	.95	1.00
2	1.00	1.00	1.00	14	.99	.98	1.00	26	.97	1.00	1.00	38	.96	.98	.98	50	.85	.84	1.00
3	.98	1.00	1.00	15	.99	1.00	1.00	27	.98	1.00	.99	39	.92	1.00	.98	51	.87	.90	1.00
4	.99	1.00	1.00	16	.97	.97	1.00	28	.90	.91	.98	40	.96	.98	.99	52	.73	.90	.97
5	.99	1.00	1.00	17	.96	1.00	1.00	29	.94	.95	.98	41	.98	1.00	.99	53	.72	.90	.97
6	.99	1.00	1.00	18	.96	.91	.97	30	.85	.88	1.00	42	.91	1.00	.95	54	.62	.81	.93
7	.97	1.00	1.00	19	.88	.83	.94	31	.94	.95	1.00	43	.88	.93	.95	55	.58	.64	.93
8	.93	.98	.99	20	.86	.86	.98	32	.71	.84	.97	44	.88	.97	.97	56	.68	.78	.92
9	.98	1.00	1.00	21	.90	.97	.97	33	.88	.93	.98	45	.85	.93	.99	57	.55	.71	.80
10	.96	.98	.99	22	.97	.98	1.00	34	.74	.79	.94	46	.88	.97	1.00	58	.28	.38	.77
11	.92	1.00	.97	23	.93	.84	.93	35	.74	.83	.90	47	.43	.47	.70	59	.16	.21	.50
12	.80	.76	.85	24	.73	.81	.90	36	.33	.62	.71	48	.30	.38	.43	60	.23	.31	.48

3. Results

3.1. Means, S.D., and internal consistencies

All calculations are based on raw scores, with each of the 60 items scored as 0 (*incorrect*) and 1 (*correct*). Internal consistencies based on Cronbach’s α were .88 for the sample as a whole, .61 for Whites, .82 for Indians, and .87 for Africans. Table 1 shows the proportion of each of the samples that selected the correct answer on each of the 60 items. For all groups, test item Set E was the most difficult, followed by Sets C and D, while Sets A and B were the easiest. Fig. 1 shows the percentage of Africans and of Whites who attained various raw scores (with the intermediately scoring Indians being excluded for clarity). The longer tail of low scores in the African distribution and the ceiling effect for both groups are clearly visible.

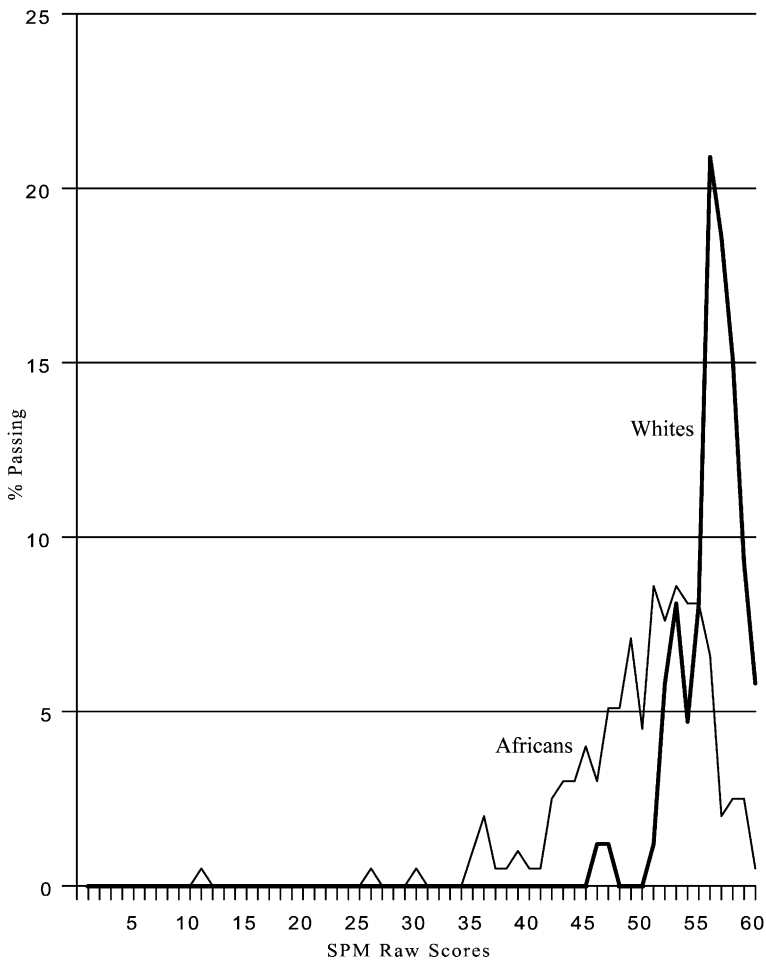


Fig. 1. Percentage of African and White 17- to 23-year-old first-year engineering students attaining various scores on the Standard Progressive Matrices Test.

The White, Indian, and African mean scores were, in order, 56, 53, and 50 out of 60 (S.D. = 2.6, 4.9, 6.4; ranges = 46–60, 37–60, 11–60). Men averaged the same scores as women (unweighted means = 52.9, 52.5; S.D. = 5.0, 3.3; ranges = 11–60, 35–60). Analysis of variance (ANOVA) with Race and Sex as factors showed a significant main effect only for Race, with no effect for Sex either as a main effect or in interaction, $F(2,342) = 24.23$, $P < .001$; $F(1,342) < 1.00$; and $F(2,342) < 1.00$. For the total score, the African–White difference was 1.00 S.D. (based on total S.D. of 6.05). The 1993 US norms for 18- to 22-year-olds shows the Whites at the 75th percentile, the Indians at the 55th percentile, and the Africans at the 41st percentile, which translate into IQ equivalents of 110, 102, and 97, respectively (Raven et al., 1990, p. 98, 1998, p. 77, Table 13).

3.2. Item analyses

Across the 60 items, the item difficulties, measured by the proportion getting the correct answer (Table 1), were very similar for Africans, Indians, and Whites ($r > .90$; $\rho > .79$, $P < .01$) suggesting that the test measured the same construct(s) in all three groups. Most of the items were too easy for these university students. Table 1 provides evidence of the ceiling effect. Relatively few of the 60 items have P -values (proportion passing) within the optimal range of .30–.70 that provides maximum discriminatory power; there are only 4 such items for the Whites, 6 for the Indians, and 7 for the Africans. Using a proportion of 70% of respondents passing as the criterion for judging an item as “too easy,” 57 of the 60 items (95%) proved too easy for the Whites, 53 or 88% for the Indians, and 50 or 83% for the Africans. No item was found to be “extremely difficult” ($P < .10$) for any of the groups.

Another index for comparing items across groups is the item-total correlation (r_{it}) for each item (Table 2). This was calculated using the point-biserial correlation (r_{pb}) of each item’s pass or fail status (0 or 1) with the total score on the test. It indicates the extent to which a particular item measures the construct that is measured by the test as a whole, as well as how well the item discriminates among the testees within each group. Since the total score on the Raven’s is a very good measure of g , the general factor of intelligence (Jensen, 1980, pp. 645–648), the item-total correlation is also an estimate of each item’s g loading.

3.3. Differences in g

To test whether African–Indian–White differences are more pronounced on the more g loaded items, we followed the same procedure as Rushton and Skuy (2000) and correlated the item-total correlations from Table 2 (the estimate of g), with the standardized differences between the ethnic groups in proportion passing each item from Table 1 (the estimate of the race effect size). The results are set out in Table 3 using, in turn, the White, Indian, and African item-total correlations. The 18 Pearson r ’s and Spearman ρ ’s ranged from $-.02$ to $.67$ with a mean of $.30$ and a median of $.31$, with 9 of the 18 correlations being individually significant. (Note that it would have been incorrect to use the item-total correlations from the combined samples because these would reflect the between-groups variance in addition to the within-groups variance and so inflate the effect.)

Table 2
Point-biserial item-total correlations for items of the Raven's Standard Progressive Matrices by race

Set A				Set B				Set C				Set D				Set E					
Item	African	Indian	White	Item	African	Indian	White	Item	African	Indian	White	Item	African	Indian	White	Item	African	Indian	White		
1	–	–	–	13	–	–	.07	–	25	.11	–	–	37	.32	.07	–	49	.44	.26	–	
2	–	–	–	14	.11	–	.07	–	26	.29	–	–	38	.40	–	.07	.27	50	.48	.52	–
3	.19	–	–	15	.43	–	–	–	27	.17	–	.04	39	.29	–	.24	51	.47	.59	–	
4	.09	–	–	16	.31	.13	–	–	28	.31	.16	.30	40	.42	.07	.00	52	.56	.44	.17	
5	.09	–	–	17	.16	–	–	–	29	.50	.50	.03	41	.36	–	.04	53	.60	.66	.42	
6	.09	–	–	18	.26	.39	–	.02	30	.45	.54	–	42	.36	–	.32	54	.48	.31	.20	
7	.39	–	–	19	.19	.53	.05	–	31	.33	.50	–	43	.40	.23	.41	55	.36	.52	.06	
8	.40	–	.07	20	.37	.26	.30	–	32	.44	.23	.27	44	.30	.17	.10	56	.55	.48	.49	
9	.36	–	–	21	.36	.06	.01	–	33	.31	.57	.39	45	.42	.37	.17	57	.56	.36	.51	
10	.34	.42	.04	22	.47	–	.07	–	34	.50	.57	.31	46	.50	.11	–	58	.46	.56	.62	
11	.36	–	.27	23	.29	.32	.27	–	35	.37	.52	.23	47	.41	.44	.50	59	.35	.34	.46	
12	.19	.34	.27	24	.39	.46	.28	–	36	.45	.35	.31	48	.34	.41	.28	60	.32	.44	.34	

Hyphen indicates that correlation could not be computed because of lack of variance on item (see Table 1).

Table 3

Correlations between item g -loadings (estimated from item-total correlations) and group differences in standardized item pass rates on the 60 items of the Raven's Standard Progressive Matrices

	Item-total correlation g -loadings					
	White ($n=37$ items)		Indian ($n=43$ items)		African ($n=57$ items)	
	r_{pb}	r_b	r_{pb}	r_b	r_{pb}	r_b
<i>Standardized pass rate differences</i>						
<i>Pearson correlations</i>						
White–Indian	.193	.148	.358 *	.277 *	.295 *	.061
Indian–African	.090	.217 [†]	.033	.037	.450**	.075
White–African	.222 [†]	.265	.318 *	.256 *	.538**	–.005
<i>Spearman correlations</i>						
White–Indian	.187	.273	.521**	.452**	.386**	.021
Indian–African	.061	.209	–.017	.004	.384**	.228 *
White–African	.245 [†]	.377 *	.398**	.320 *	.667**	.160

* $P < .05$.

** $P < .01$ (one-tailed).

[†] $P < .10$.

To test the generality of these findings, we carried out several additional analyses. Our use of the point-biserial correlation (r_{pb}) to calculate the item-total correlations (Table 2) to estimate g may have confounded g loadings with item-difficulty levels making the observed Jensen Effects due to the artifact of item difficulty rather than to g . Thus, we recalculated the values for Table 2 using the biserial (r_b) correlation (because its formula contains within it a correction to the item difficulty in the form of the ordinate, y , of the normal curve at the % pass/% fail cut-point) and related these to the standardized proportions passed in Table 1. The results of this additional analysis are shown alongside the earlier results in Table 3 and confirm the Jensen Effects, albeit in reduced form. The 18 Pearson r 's and Spearman ρ 's now ranged from $-.06$ to $.45$ with a mean of $.18$ and a median of $.21$, with 6 of the 18 correlations being individually significant.

Low reliabilities found at the item level likely contributed to the lower than usual (but still significant) Jensen Effects. Since aggregation of items into composite scores is an excellent procedure with a good track record for increasing construct validity in disparate fields of psychology (Rushton, Brainerd, & Pressley, 1983), we decided to try it here. We aggregated items six at a time to make 10 “subtests” to see if these raised the magnitude of the Jensen Effects to those usually found, e.g., when using 13 subtests from the Wechsler Scales (Jensen, 1998) and found that, indeed, this procedure did boost their size (see Table 3). When using the point-biserial item-total correlations, the aggregated White, Indian, and African g 's (each on a maximum of 10 “subtests”) predicted the aggregated White–Indian, Indian–African, and White–African standardized differences in pass rates with mean and median Spearman ρ 's of $.68$. Similarly, when using the biserial item-total correlations, the aggregated White, Indian, and African g 's predicted the standardized group differences with mean and median Spearman ρ 's of $.39$ and $.53$, respectively.

Table 4

Spearman correlations between g -loadings estimated from aggregated item-total correlations and aggregated group differences in standardized item pass rates on the 60 items of the Raven's Standard Progressive Matrices

	Item-total correlation g -loadings					
	White ($n=9$ subtests)		Indian ($n=9$ subtests)		African ($n=10$ subtests)	
	r_{pb}	r_b	r_{pb}	r_b	r_{pb}	r_b
<i>Standardized pass rate differences</i>						
White–Indian	.52 [†]	.53 [†]	.77**	.47 [†]	.47 [†]	–.35
Indian–African	.53 [†]	.72**	.60*	.76**	.76**	–.11
White–African	.68*	.83**	.96***	.83**	.83**	–.21

* $P < .05$.

** $P < .01$.

*** $P < .001$ (one-tailed).

[†] $P < .10$.

One possible explanation for why the aggregated biserial correlations for Africans (Table 4) are such outliers, which was not the case when using the point-biserial correlations (Table 3), is that the Africans departed most from the optimal P (% passing) value of .50, i.e., they had a lot of items with extreme P values or with little variance among their item P values. This view is supported by the lower correlation found between the African r_b 's and r_{pb} 's compared to the other two groups. The Spearman correlation between r_{pb} and r_b for the Whites was .84, for the Indians .92, and for the Africans .59. Regardless, in the main, the Jensen Effect comes through with a magnitude similar (median $\rho = .53$) to that found in previous studies based on whole subtests.

4. Discussion

Engineering students at the University of the Witwatersrand are one of the highest scoring African samples measured to date on the Raven's test. Out of the 60 problems, the African students solved an average of 50 correctly, with the Indians and Whites solving 53 and 56, respectively. The African engineering students' score of 50 is very similar to the score of 52 found in Zaaiman et al.'s (2001) study of highly selected African math and science students at South Africa's University of the North. Nonetheless, the African–White difference in our study amounted to 1 S.D. (based on the total sample S.D. of 6.05). By the 1993 US norms, the African students were at the 41st percentile, the Indians at the 55th, and the Whites at the 75th, yielding IQ equivalents of 97, 102, and 110, respectively.

The African students were a highly selected population. They had graduated from high school after passing standardized exams in mathematics and science, entered one of South Africa's leading universities, and been chosen for a first-year course in engineering, which includes mathematics, all on the basis of academic performance. Assuming these engineering students are 2 S.D. above the population mean, like the highly selected math and science students studied by Zaaiman et al. (2001), and 1 S.D. above the social science students studied

by Grieve and Viljoen (2000) and Rushton and Skuy (2000), the results dovetail with all the earlier work (Lynn & Vanhanen, 2002) finding that the general population in Africa currently averages a mean tested IQ of 70. On the basis of the current evidence, at least, there is no bimodal distribution of scores in the African population as hypothesized by Rushton and Skuy.

Nonetheless, the Raven's SPM was too easy for most of these engineering students. The distribution of scores was skewed to the right with a possible ceiling effect for all groups (Fig. 1). There is a lack of fine discriminative power at the upper end of the SPM distribution where the difference in raw scores between percentiles is small. For 18- to 22-year-olds, the normative group for our sample, the difference between the 61st and the 100th percentile is only six correct answers, or 6 percentile points per 1 SPM score point. Thus, the mean IQ of 110 for the White students and 102 for the Indian students are especially likely to be underestimates, as are the scores of the higher scoring African students.

A very wide range of interpretations is possible for these results. At one extreme, Nell (2000) and Olson (1986) have argued that much of the supposed "culture fairness" of the nonverbal SPM is illusory and that it requires the same Western cultural style of analytical rule following that more traditional IQ tests do, and that African languages and Black cultures are more "holistic." Nell recommended that until language proficiency, educational quality, test-wiseness, and other components of acculturation have been proved beyond any reasonable doubt to be equivalent for the groups whose scores are being compared, score differences cannot be attributed to anything other than "cultural impact." He even asserted that "the use of Western tests with clients from other cultures can be so misleading that a cultural veto should be imposed on the use of such test scores" (p. 85).

Certainly, a very important point in testing is that test takers should be sufficiently similar in cultural, educational, and social background to those on whom the test has been standardized and the test norms based (Irvine & Berry, 1988; Jensen, 1980). If the group differs markedly from the standardization sample, the use of the norms may be inappropriate. On the other hand, at least some of our data suggest that the scores are as valid for Africans as they are for Whites. Most of the items (83%) were found to be "easy" by African students (based on a 70% pass rate), so they could perform the required operations. Also, analyses of the inter-item matrices found the items "behaved" in similar ways for Africans as for non-Africans (e.g., the internal consistencies, the item difficulty levels, the item-total correlations, and the group differences being on g). If there were not some degree of cross-cultural generality, such findings as the Indian item-total correlations predicting the African–White pass-rate differences could not be observed (see Table 3).

These data confirm the magnitude of the African–White IQ gap and that the differences on the various items are positively associated with the g loading for those items. The effect appears robust and implies that g is the same in South Africa as it is in other countries, such as the US and the Netherlands. If confirmed by further research, it tells us that the main source of African–White differences across various cognitive tests is essentially the same as for the differences between individuals within each racial group, namely, g . In short, from our present vantage point, the African–White differences are Jensen Effects.

Accepting that these results reflect the current level of abstract cognitive performance of the African population, how do we account for them? And, is there anything that can be done

to raise them? Just because African–European–East Indian differences show up on the *g* factor does not preclude intervention strategies aimed at remedying cognitive deficits and narrowing the group differences, even if there is a mix of genetic and environmental factors operating (Jensen, 1998; Rushton, 2000). In the Netherlands, for example, te Nijenhuis and van der Flier (2001) reported that the immigrant/non-immigrant gap had narrowed by the second generation of immigrants. In the US, there is hope that the Black–White difference may also be narrowing (Herrnstein & Murray, 1994; Jencks & Phillips, 1998; Neisser, 1998).

One intervention technique that has boosted scores on the Raven's test in South Africa is Feuerstein's (1980) Mediated Learning Experience—using the Learning Propensity Assessment Device (LPAD) and the Instrumental Enrichment Program. Thus, Skuy and Shmukler (1987) used the LPAD to demonstrate the application of Mediated Learning Experience to raise the performance of Colored and Indian high school students, and Skuy, Hoffenberg, Visser, and Fridjhon (1990) found generalized improvements for African individuals with what they termed a “facilitative temperament.” Similarly, in an intervention study with first-year psychology students at the University of the Witwatersrand, Skuy et al. (2002) raised test scores on the Raven's SPM using the LPAD. Before intervention, the Africans scored 43 out of 60; the non-Africans (Whites, Indians, and Coloreds) averaged 52 out of 60. After the intervention, both experimental groups improved over baseline compared to control groups, with significantly greater improvement for the Africans. From a *g* perspective, the question is whether the intervention procedure increased performance through mastery of subject-specific knowledge or if it increased general problem-solving skills that would apply to other subjects and workplace training as well.

In conclusion, although the present finding of a mean IQ of 97 in selected first-year African engineering students dovetails with the earlier work finding a general population IQ mean of 70 in sub-Saharan Africa, it might still be conjectured that a test with a higher ceiling would reveal evidence of a higher scoring African group and a bimodal distribution among Africans. Further research might also provide construct validity data on the Raven's test scores of Africans to see whether they predict examination results. From the current vantage point, however, the population differences are not attributable to idiosyncratic cultural peculiarities in this or that test but to a general factor that all the ability tests measure in common. More research is obviously needed to determine what the “true” African mean IQ is, whether African/non-African differences are on the *g* factor, whether IQ scores in Africans are as predictive of grades and other criteria as they are for non-Africans, and whether intervention techniques will work to raise the IQs.

References

- Educational Testing Service (1998). *Graduate record examinations: guide to the use of scores*. Princeton, NJ: Educational Testing Service.
- Feuerstein, R. (1980). *The dynamic assessment of retarded performers*. Glenview, IL: Scott, Foresman.
- Grieve, K. W., & Viljoen, S. (2000). An exploratory study of the use of the Austin Maze in South Africa. *South African Journal of Psychology*, 30, 14–18.

- Herrnstein, R. J., & Murray, C. (1994). *The bell curve*. New York: Free Press.
- Irvine, S. H., & Berry J. W. (1988). *Human abilities in cultural context*. New York: Cambridge Univ. Press.
- Jencks, C., & Phillips M. (Eds.) (1998). *The Black–White test score gap*. Washington, DC: Brookings Institution Press.
- Jensen, A. R. (1980). *Bias in mental testing*. New York: Free Press.
- Jensen, A. R. (1998). *The g factor*. Westport, CT: Praeger.
- Lynn, R. (1991). Race differences in intelligence: a global perspective. *Mankind Quarterly*, 31, 255–296.
- Lynn, R. (1997). Geographical variation in intelligence. In H. Nyborg (Ed.), *The scientific study of human nature: essays in honor of H.J. Eysenck*. (pp. 259–281). New York: Pergamon.
- Lynn, R., & Owen, K. (1994). Spearman's hypothesis and test score differences between Whites, Indians, and Blacks in South Africa. *Journal of General Psychology*, 121, 27–36.
- Lynn, R., & Vanhanen, T. (2002). *IQ and the wealth of nations*. Westport, CT: Praeger.
- Neisser, U. (Ed.) (1998). *The rising curve: long term gains in IQ and related measures*. Washington, DC: American Psychological Association.
- Nell, V. (2000). *Cross-cultural neuropsychological assessment: theory and practice*. London: Erlbaum.
- Olson, D. R. (1986). Intelligence and literacy: the relationship between intelligence and the technologies of representation and communication. In R. J. Sternberg, & R. K. Wagner (Eds.), *Practical intelligence: nature and origins of competence in the everyday world* (pp. 338–360). New York: Cambridge Univ. Press.
- Osborne, R. T. (1980). The Spearman–Jensen hypothesis. *Behavioral and Brain Sciences*, 3, 351.
- Owen, K. (1992). The suitability of Raven's Standard Progressive Matrices for various groups in South Africa. *Personality and Individual Differences*, 13, 149–159.
- Raven, J., et al. (1990). *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research supplement No. 3: American and international norms* (2nd ed.). Oxford, England: Oxford Psychologists Press.
- Raven, J. C., Court, J. H., & Raven, J. (1998). *Manual for Raven's Standard Progressive Matrices (1998 edition)*. Oxford, England: Oxford Psychologists Press.
- Rushton, J. P. (1998). The “Jensen Effect” and the “Spearman–Jensen hypothesis” of Black–White IQ differences. *Intelligence*, 26, 217–225.
- Rushton, J. P. (2000). *Race, evolution, and behavior* (3rd ed.). Port Huron, MI: Charles Darwin Research Institute.
- Rushton, J. P. (2001). Black–White differences on the g factor in South Africa: a “Jensen Effect” on the Wechsler Intelligence Scale for Children—Revised. *Personality and Individual Differences*, 31, 1227–1232.
- Rushton, J. P., in press. Jensen Effects and African/Colored/Indian/White differences on Raven's Standard Progressive Matrices in South Africa. *Personality and Individual Differences*.
- Rushton, J. P., Brainerd, C. J., & Pressley, M. (1983). Behavioral development and construct validity: the principle of aggregation. *Psychological Bulletin*, 94, 18–38.
- Rushton, J. P., & Skuy, M. (2000). Performance on Raven's Matrices by African and White university students in South Africa. *Intelligence*, 28, 251–265.
- Skuy, M., Gewer, A., Osrin, Y., Khunou, D., Fridjhon, P., & Rushton, J. P. (2002). Effects of mediated learning experience on Raven's Matrices scores of African and non-African university students in South Africa. *Intelligence*, 30, 221–232.
- Skuy, M., Hoffenberg, S., Visser, L., & Fridjhon, P. (1990). Temperament and cognitive modifiability of academically superior Black adolescents in South Africa. *International Journal of Disability, Development, and Education*, 37, 29–44.
- Skuy, M., Schutte, E., Fridjhon, P., & O'Carroll, S. (2001). Suitability of published neuropsychological test norms for urban African secondary school students in South Africa. *Personality and Individual Differences*, 30, 1413–1425.
- Skuy, M., & Shmukler, D. (1987). Effectiveness of the learning potential assessment device with Indian and Colored adolescents in South Africa. *International Journal of Special Education*, 12, 131–149.
- Spearman, C. (1927). *The abilities of man: their nature and measurement*. New York: Macmillan.
- te Nijenhuis, J., & van der Flier, H. (1997). Comparability of GATB scores for immigrants and majority group members: some Dutch findings. *Journal of Applied Psychology*, 82, 675–687.

- te Nijenhuis, J., & van der Flier, H. (2001). Group differences in mean intelligence for the Dutch and third world immigrants. *Journal of Biosocial Science*, *33*, 469–475.
- Zaaiman, H., van der Flier, H., & Thijs, G. D. (2001). Dynamic testing in selection for an educational programme: assessing South African performance on the Raven Progressive Matrices. *International Journal of Selection and Assessment*, *9*, 258–269.