In Defense of a Disputed Study of Construct Validity from South Africa

J. Philippe Rushton*
The University of Western Ontario

This analysis of a critique finds that the original study accurately showed that the items found easy or difficult by Black South African undergraduates were those found easy or difficult by their White and South Asian counterparts (r's = .90). There was no evidence of any culture-specific effect. Instead, African/non-African differences were found to be most pronounced on *g*. This was shown by item-total correlations (estimates of the item's *g* loading), which predicted the magnitude of African/non-African differences on those same items, and by a confirmatory factor analysis. The tests were equally predictive for Blacks and non-Blacks on external criteria such as course grades. The results indicate the remarkable cross-cultural generalizability of item properties across sub-Saharan Africans, South Asians, and Europeans and that these reflect *g* more than culturally specific ways of thinking.

Cronshaw, Hamilton, Onyura, and Winston (2006) have written a critique of my "race-realist" research. They raised technical objections to a study I published in this journal on South African engineering students showing that the Raven's Advanced Progressive Matrices is as valid a measure of general mental ability (GMA) for Africans as it is for non-Africans (Rushton, Skuy, & Bons, 2004). They denigrated it as the kind "discussed on White supremacist and neo-Nazi websites around the world" and as "part of a long tradition of scientific racism in psychology." They argued that it, and a "large corpus of related work conducted by Rushton and his colleagues ... should be interpreted and understood in ... broader context" (pp. 20–21). This name-calling is likely because of my conclusion that the mean IQ in sub-Saharan Africa is 70, which is 15 points below that of African Americans, and is partly heritable (Rushton & Jensen, 2005).

Rushton *et al.* (2004) tested the hypothesis that the Raven's Advanced Progressive Matrices has the same construct validity in African university students as it does in non-African students. We examined data from 306 highly select 17- to 23-year-olds in the Faculties of Engineering and the Built Environment at the University of the Witwatersrand (177 Africans, 57 East Indians, 72 Whites; 54 women, 252 men). Analyses were made of the Matrices scores, an English Comprehension test, and the Similarities subscale from the South African Wechsler Adult Intelligence Scale, end-of-year university grades, and high-school grade point average. Out of the 36 Matrices problems, the African students solved an average of 23; East Indian students, 26; and White students, 29 ($p < .001$), placing them at the 60th, 71st, and 86th percentiles, respectively, and yielding IQ equivalents of 103, 108, and 118 on the 1993 U.S. norms. The same pattern of group differences was found on the Comprehension test, the Similarities subscale, university course grades, and high-school grade-point average. The items on the Matrices "behaved" in the same way for the African students as they did for the non-African students, thereby indicating the test's internal validity. Item analyses, including a confirmatory factor analysis, showed that the African/non-African difference was most pronounced on the general factor of intelligence. Concurrent validity was demonstrated by correlating the Matrices with the other measures, both individually and in composite. For the African group, the mean $r = .28$, $p < .05$, and for the non-African group, the mean $r = .27$, $p < .05$. Although the intercepts of the regression lines for the two groups were significantly different, their slopes were not. We concluded that these results showed the Raven's Matrices are as valid for Black Africans as they are for non-Black Africans.

Cronshaw *et al.* (2006) objected to our study and concluded that the history of apartheid and racism in South Africa biased the test against Black Africans via "stereotype

*Address for correspondence: J. Philippe Rushton, Department of Psychology, The University of Western Ontario, London, ON, Canada N6A 5C2. E-mail: rushton@uwo.ca

threat." They listed five main faults with our work centering on: (1) the item difficulties; (2) the use of the item difficulties in conjunction with the item-total correlations as indicators of *g*; (3) the test of *g* using a confirmatory factor analysis; (4) the prediction of educational criteria; and (5) the use of non-equivalent groups.

Item Difficulties

Rushton *et al.* (2004) found the 36-item difficulties for Black Africans, East Indians, and Whites correlated .95. The items found difficult by one group were those found difficult by the others. We concluded that the same psychometric constructs were measured in all three groups.

Cronshaw *et al.* (2006) objected to our conclusion and carried out an alternate procedure – a "more obvious and direct comparison". They calculated a "mean item difficulty" of .64 for Black Africans, .72 for the East Indian sample, and .80 for the White sample. On this basis they concluded that the items were most biased for the Africans, the next most for the East Indians, and the least for the Whites.

My response is that Rushton *et al.* used the standard method and that Cronshaw *et al.* used one that is nonsensical. Their unorthodox procedure merely restated what was already known from the total scores – that Africans passed fewer items. Their Figures 1 and 2 provide no further clarification. Cronshaw *et al.* have misunderstood item analyses for comparing groups (Jensen, 1980).

Item Difficulty/Item-Total Correlations

Rushton *et al.* (2004) tested the Spearman–Jensen hypothesis that African/non-African differences are most pronounced on the more *g* loaded components of tests (Jensen, 1985, 1998; Spearman, 1927). We correlated the item-total score correlations (which estimate *g*), with the standardized differences between Africans and non-Africans in proportion passing each item (which estimate the group effect size), and found $r = .34$ and .42 (p's $< .05$) using the item-total correlations for the non-African group, but only .22 and .21 (NS) using the item-total correlations for the African group. (Note: 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.)

Cronshaw *et al.* (2006) emphasize that the analyses only half-supported the hypothesis. They also said they found it "rather odd" and "anomalous" for us to combine the East Indian sample with that of the Whites for this analysis. When they disaggregated the samples and re-did the analyses, they confirmed that the hypothesis was only supported for some of the comparisons.

My response is to note that a failure to reach significance in one of several analyses does not invalidate a hypothesis. Eight studies have now tested Spearman's hypothesis in African samples and the great majority of analyses (but not all) find the Black–White differences are greater on the most *g* loaded of the test components (see Rushton *et al.*, 2004 for a list of studies; Rushton, 2003, for detailed summaries of the first seven; the eighth is the one under discussion). We combined the small sample of East Indians ($N = 57$) with the small sample of Whites ($N = 72$) to reduce specificity and error variance and so increase the power of the test for making a comparison with the larger sample of Black Africans ($N = 177$).

Multi-Group Confirmatory Factor Analysis

Rushton *et al.* (2004) followed Ree and Carretta (1995) and performed a multi-group confirmatory factor analysis of the single-factor *g* solution. We did this on the 14 items (of 36) with difficulty levels between .20 and .80 for all three ethnic groups. The results indicated that the same model – single factor *g* – fit the data for all three groups.

Cronshaw *et al.* (2006) objected to our conclusion and carried out an alternate procedure – a more "straightforward analysis." They correlated all 36 item-total correlations (a measure of *g*) across the three groups, and reported they were unrelated. They concluded that in Black Africans, "the Raven's is measuring a different construct".

My response is to note that Rushton *et al.*'s procedure is the more sensitive for detecting whether *g* is common to all groups. It is standard practice to prefer items with difficulty levels of between .20 and .80, and factor analysis typically works to cancel the idiosyncratic error and specificity variance more readily than do simple correlations.

Raven's Scores and Educational Criteria

Rushton *et al.* (2004) correlated the group's Raven's scores with those on the English Comprehension Test, the Similarities subscale of the new South African Wechsler Adult Intelligence Scale, and course grades from high-school and university. The mean inter-correlation among all variables for the Africans was .23, and for the non-Africans, .27. The Raven's scores predicted a composite of the four criteria in both the Africans ($r = .28$; $p < .01$) and the non-Africans ($r = .27$; $p < .01$), with the slopes of the regression lines not significantly different for the two groups over the entire range of scores.

Cronshaw *et al.* (2006) argued that the small sample sizes did not provide enough statistical power to make adequate between-group comparisons. They wanted bias assessed using the "Thorndike model," which examines whether the magnitude of the difference between two groups in test scores is matched in the criteria. If the difference between groups is greater in the test scores than it is in the criteria, then the test can be considered biased. Cronshaw *et al.* rejected what they acknowledged most psychologists would think of as the "gold standard" Cleary model, which examines whether two groups differ in the slopes of their regression lines. If these do not differ in predicting criteria, then the test is not biased. They claimed our regression lines showed, "the bias is reported, diagrammed, and obvious to even the most cursory examination!"

My response is to repeat that we found no evidence of bias using standard procedures. We did not use the Thorndike model because criteria can often be biased too (Jensen, 1980, p. 383). For example, many faculty members at the University of the Witwatersrand (Black Africans, as well as Whites) reported to us that in the bid to hasten "Africanization," the administration was exerting pressure not to fail students and to "push them through." This was leading (they said) to higher grades for academically poorer students. Despite the "push through" philosophy, the drop out rate in the Engineering Faculty remained high, which was one of the reasons why the (Black African) Dean encouraged our study. Although the majority of new students would continue to be Black Africans, the Dean wanted a better method of selection to ensure scarce resources were not wasted.

The Use of Non-Equivalent Groups

Rushton *et al.* (2004) selected African, East Indian, and White university students taking the same course from the same instructor at the same time. Most were admitted to university after graduating from high school with courses in math and science. Seemingly, they shared similar interest patterns and vocational aspirations.

Cronshaw *et al.* (2006) objected that the groups were not equated because of "South Africa's history of segregation and apartheid . . . one of the most extreme forms of racial discrimination in the twentieth-century world". They claimed this invalidated the comparisons because it produced "stereotype threat." They queried how the subjects indicated their ethnicities, and what other information was asked of them.

My response is that "stereotype threat" was minimized in Rushton *et al.*'s study because the biographical data (including ethnic background) was collected from the Dean's office and the students were told the session was to validate the test; we were not interested in their particular scores. In any case, stereotype threat, which is a form of test anxiety, has been only poorly demonstrated as a factor in explaining average differences between Blacks and Whites in the United States (Sackett, Schmitt, Kabin, & Ellingson, 2001). To my knowledge zero studies have shown it applies in South Africa. More generally, Cronshaw *et al.*'s claim that South Africa's apartheid system exerted especially pernicious effects in lowering Black IQ scores is belied by the observation that many of the African countries showing a mean IQ = 70, such as Nigeria and Ghana, have been independent for half a century (and the Caribbean Island of Haiti for one and a half centuries), with no documented improvement in cultural achievement or IQ scores.

Conclusion

The results from Rushton *et al.*'s (2004) study of university engineering students in South Africa showed that even in these restricted range samples the items on the Matrices behaved in the same way for Africans as for non-Africans and the external validities were also the same ($r = .28$). These results dovetailed with a review by Kendall, Verster, and Von Mollendorf (1988) showing test scores were equally predictive for Africans and non-Africans (e.g., .20–.50) for both school grades and job performance. Kendall *et al.* found that many of the factors influencing scores in Africans were the same as those for Whites (e.g., coming from an urban versus rural environment; being a science rather than an arts student; having had practice on the tests; and the well-documented curvilinear relationship of test score with age). Rushton, Skuy, and Fridjhon (2002) and Rushton, Skuy, and Fridjhon (2003) also found that scores on the Advanced Matrices correlated with scores on the Standard Matrices measured 3 months earlier (.60 for Africans; .70 for non-Africans) and end-of-year exam grades 3 months later (.34 for Africans; .27 for non-Africans). Sternberg *et al.* (2001) found that GMA scores in Kenyan 12- to 15-year-olds predicted school grades with a mean $r = .40$, $p < .001$ (and after controlling for age and socio-economic status, $r = .28$, $p < .01$).

As predictors of future performance, therefore, the tests appear not to be biased against Black Africans. The large mean differences between Africans and non-Africans likely reflect GMA rather than culturally specific ways of thinking. Although it is non-arguable that intervention strategies in Africa such as the elimination of tapeworms, improved nutrition, and provision of electricity, schools, and hospitals, will raise test scores, race-realists believe these improvements will not remove the pattern of average differences in GMA between Africans and others (e.g., Lynn, 2006; Rushton & Jensen, 2005). For example, mixed-race populations such as South African "Coloreds" (their preferred term), as well as African Americans, are predicted to continue to average between the two parental types because of White admixture.

Increasing research is converging on the view that GMA, more than any other factor, predicts academic achievement, creativity, career potential, and job performance – even health-related behavior (Gottfredson, 2004; Kuncel, Hezlett, & Ones, 2004; Lubinski, 2004). The evidence is also growing that GMA test validities are internationally generalizable. Salgado, Anderson, Moscoso, Bertua, and

Fruyt (2003), for example, found GMA predicted job performance ratings .62 and training success .54 in 10 member countries of the European Community. Crucially, Salgado *et al.*'s results contradicted the view that criterion-related validity for GMA tests is moderated by differences in a country's culture, religion, language, socioeconomic level, or employment legislation. The validities were the same, or even higher, than those reported in the United States, where there is again a quite different corporate culture, mix of populations, and legislative history. Salgado *et al.* posited that GMA results are internationally generalizable in predicting varieties of work performance. The data presented in this paper are in accord with that perspective. The onus is on the critics of GMA to show how some other, alternative test, makes better predictions for some groups.

References

Cronshaw, S.F., Hamilton, L.K., Onyura, B.R. and Winston, A.S. (2006). Case for non-biased intelligence testing against Black Africans has not been made: A comment on Rushton, Skuy, and Bons (2004). *International Journal of Selection and Assessment*, **14**, 278–287.

Gottfredson, L.S. (2004) Intelligence: Is it the epidemiologists' elusive "fundamental cause" of social class inequalities in health? *Journal of Personality and Social Psychology*, **86**, 174–199.

Jensen, A.R. (1980) *Bias in mental testing*. New York: Free Press.

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. (1998) *The g factor*. Westport, CT: Praeger.

Kendall, I.M., Verster, M.A. and Von Mollendorf, J.W. (1988) Test performance of Blacks in Southern Africa. In S.H. Irvine and J.W. Berry (Eds), *Human abilities in cultural context* (pp. 299–339). Cambridge: Cambridge University Press.

Kuncel, N.R., Hezlett, S.A. and Ones, D.S. (2004) Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, **86**, 148–161.

Lubinski, D. (2004) Introduction to the special section on cognitive abilities: 100 years after Spearman's (1904) "'General intelligence,' objectively determined and measured". *Journal of Personality and Social Psychology*, **86**, 96–111.

Lynn, R. (2006) *Race differences in intelligence: An evolutionary analysis*. Augusta, GA: Washington Summit Books.

Ree, M.J. and Carretta, T.R. (1995) Group differences in aptitude factor structure on the ASVAB. *Educational and Psychological Measurement*, **55**, 268–277.

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.

Rushton, J.P. and Jensen, A.R. (2005) Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law*, **11**, 235–294.

Rushton, J.P., Skuy, M. and 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., Skuy, M. and Fridjhon, 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. and Fridjhon, P. (2003) Performance on Raven's Advanced Progressive Matrices by African, Indian, and White engineering students in South Africa. *Intelligence*, **31**, 123–137.

Sackett, P.R., Schmitt, N., Kabin, M.B. and Ellingson, J.E. (2001) High-stakes testing in employment, credentialing, and higher education: Prospects in a post-affirmative-action world. *American Psychologist*, **56**, 302–318.

Salgado, J.F., Anderson, N., Moscoso, S., Bertua, C. and Fruyt, F.D. (2003) International validity generalization of GMA and cognitive abilities: A European community meta-analysis. *Personnel Psychology*, **56**, 573–605.

Spearman, C. (1927) *The abilities of man: Their nature and measurement*. New York: Macmillan.

Sternberg, R.J., Nokes, C., Geissler, P.W., Prince, R., Okatcha, F., Bundy, D.A. and Grigorenko, E.L. (2001) The relationship between academic and practical intelligence: A case study in Kenya. *Intelligence*, **29**, 401–418.