Case for Non-Biased Intelligence Testing Against Black Africans Has Not Been Made: A Comment on Rushton, Skuy, and Bons (2004)

Steven F. Cronshaw*, Leah K. Hamilton, Betty R. Onyura and Andrew S. Winston
University of Guelph

This reply reviews the conceptual, methodological, and statistical foundations of Rushton, Skuy and Bons’ article in this journal that compared Black Africans, Whites and East Indians on the Raven’s Advanced Progressive Matrices, and concluded that the Raven’s is an unbiased test. Through a technical re-analysis of both the internal and external validity criteria for test bias using data reported in the Rushton et al. paper, we demonstrate that the Raven’s Matrices test is in fact biased against Black Africans. We take issue with several additional elements of Rushton et al.’s study, including the use of non-equivalent groups in test samples. We briefly review Rushton’s racial-realist research agenda and show that the assumption of test bias is central to advancing that agenda. Industrial/organizational and occupational psychologists should critically analyze and re-evaluate the science employed in Rushton’s racial-realist research and also should better understand the ethical and social implications of accepting his reports of research findings on test bias and White–Black IQ differences as established scientific facts.

The only thing necessary for the triumph of evil is for good men to do nothing – (Edmund Burke).

In the September 2004 issue of this journal, Rushton, Skuy, and Bons (2004) published an article that investigated the construct validity of Raven’s Advanced Progressive Matrices test for Black African and non-Black African engineering students in South Africa. They report the finding that there were significant mean group differences among Black, East Indian, and White South African students: The White and East Indian students had higher mean scores than the Black students. After reporting results of test bias analysis using criteria of internal and external validity, they conclude that “the scores on the Raven’s Matrices are as valid for Africans as they are for non-Africans” (p. 220) and “[t]here was no evidence of test bias” (p. 227). Rushton et al. (2004, p. 228) claim that their study thus demonstrates that Black African and non-Black African differences reflect the g factor of intelligence rather than any culturally specific ways of thinking. The authors then suggest their results indicate that the Raven’s and other g loaded tests are appropriate for use in selection among highly educated Black Africans.

This paper presents a review of the conceptual, methodological, and statistical foundations of the Rushton et al. (2004) study, while pointing to a number of shortcomings in their paper. We offer different interpretations of their findings and draw the opposite conclusion to that made by Rushton et al. (2004). That is, their results clearly demonstrate that the Raven’s is biased against Black Africans.

Throughout this reply we will refer to the three study groups investigated by Rushton et al. (2004) as Black African, White, and East Indian. In cases where they (sometimes inexplicably) collapsed analyses across groups, we will refer to study groups as Black African and non-Black African (i.e., White and East Indian).


Rushton et al. (2004) made substantial errors in their analysis and subsequent interpretation of their results. In addition, they failed to report results that would contradict their
conclusion that the Raven’s test is unbiased against Black Africans. We will discuss our concerns with the technical aspects of the Rushton et al. (2004) article under four headings. These concerns span the gamut of issues explored by Rushton et al. (2004) in their analyses of Raven’s scores.

**Internal Validity Criteria for Test Bias: Comparison of Item Difficulties**

Rushton et al. (2004) report the item difficulties for 36 Raven’s items across Black African, East Indian, and White groups. They compute the correlation between the item difficulties as .95 which is statistically significant at $p < .001$ (which we take to mean the correlation between the item difficulties for the Black African sample with the item correlations of the combined White and East Indian samples), concluding that the difficulties across the 36 Raven’s items “were similar for all groups” (p. 224). This finding, according to the authors, “suggests that the test measures the same construct in all three groups” (p. 224). We made the more obvious and direct comparison between item difficulties by computing the mean item difficulty for each of the three ethnic samples from the data presented in Rushton et al.’s Table 1. These mean item difficulties were, respectively: .64 for the Black African sample; .72 for the East Indian sample; and .80 for the White sample. The item difficulties reported by Rushton et al. are patently not “similar for all groups” in that the mean item difficulties are higher for Black Africans than for the other two samples. Notably, the mean difference between the Black African and White samples is the greatest. The discrepancies between item difficulties for Black Africans as compared with the White group are clearly shown in the scattergram of item difficulties presented in Figure 1. Interestingly, the item difficulties for the East Indian group, when plotted against the item difficulties for the White group in Figure 2, show a somewhat more regular pattern (i.e., the scatterplot points tend to fall on a parallel with the 45° line), as

<table>
<thead>
<tr>
<th>Differences in item difficulties for</th>
<th>Black African ($n = 36$)</th>
<th>East Indian ($n = 34$)</th>
<th>White ($n = 32$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White–Black African</td>
<td>$-.24 (-.31^*)$</td>
<td>$.21 (.18)$</td>
<td>$.42** (.45**$)</td>
</tr>
<tr>
<td>White–East Indian</td>
<td>$-.38^* (-.42**$)</td>
<td>$.01 (.08)$</td>
<td>$.13 (.17)$</td>
</tr>
<tr>
<td>East Indian–Black African</td>
<td>$.03 (.05)</td>
<td>$.26 (.20)</td>
<td>$.42** (.44**$)</td>
</tr>
</tbody>
</table>

**Table 1. Correlations of differences in item difficulties with item discrimination indices in pairwise comparisons of ethnic groups**

Notes: First correlation for each entry is Pearson’s product moment; second correlation in parentheses is Spearman’s $\rho$. $^* p < .05$, $^{**} p < .01$. Bold entries in this table are further discussed in the text.

Figure 1. Scattergram of Raven’s item difficulties for Black African vs. White groups.

Figure 2. Scattergram of Raven’s item difficulties for East Indian vs. White groups.
compared with the generally “off-parallel” pattern when the Black group is plotted against the White group. The pattern of item difficulties in the Black–White comparison in Figure 1 show that discrepancies in item difficulties between groups are most pronounced at and below the mid-range of item difficulties, an effect that is not found for the White–East Indian comparison in Figure 2. The meaning of this discrepancy in findings between Figures 1 and 2 will be further discussed at the end of the next section when we look at the second internal validity problem in connection with Rushton et al.’s study.

**Internal Validity Criteria for Test Bias:**

**Comparison of Group Differences in Item Difficulties with Item-Total Correlations for the Three Ethnic Groups**

Rushton et al. (2004) tested to see whether Black African and non-Black African differences in Raven’s item difficulties were “more pronounced on the more g loaded items” (p. 234). On the one hand, when they performed the analysis with the non-Black African sample, they found moderate but significant correlations between the magnitude of these differences and item-total correlations. On the other hand, when they performed the analysis with the Black African sample, they found non-significant correlations between the magnitude of the Black African/non-Black African differences and item-total correlations. These findings must be considered non-supportive of Rushton et al.’s hypothesis, given that the “g-loaded” items in both the Black African and non-Black African groups should be positively correlated with Black African/non-Black African differences in item difficulty if the Raven’s test is indeed unbiased against the Black African group. However, beyond reporting “only” non-significant correlations for the Black African group, Rushton et al. do not comment further on this matter.

We found it rather odd that Rushton et al. (2004) reported their aggregated findings for the Black African and non-Black African groups (i.e., they combined the data for the East Indian and White samples into a single aggregate score for non-Black Africans). This was particularly anomalous given that the first two tables of their paper (i.e., item difficulties in Table 1 and item-total correlations in Table 2) reported detailed information for all three samples (Black African, East Indian, and White). To investigate this further, we reanalyzed their data from Tables 1 and 2 for all three samples, using the procedure recommended in a related paper by Rushton and Skuy (2000) on performance by Black and White African students on the Raven’s Matrices. In that paper, Rushton and Skuy state that “[i]f a test measures the same ability in the African and the White groups, then items that best measure ability within each group (i.e., those items with the largest item-total correlations) should also discriminate most between the groups” (p. 260). We take this to mean that, in line with the Spearman hypothesis (Jensen, 1998); the most “g-loaded” items – those with the highest item-total correlations – will be associated with the largest differences in item difficulties between the White and Black African groups, where these differences will consistently favor the White group. As the results of our re-analysis show in our Table 1, Rushton et al.’s prediction is borne out when g loadings are estimated for the White sample and then correlated with the White/Black African differences in item difficulties, but not when g loadings are estimated for the Black African group and then correlated with the White/Black African differences in item difficulties. In fact, the tendency is for the opposite effect to occur for the Black African sample: The higher “g” loadings, when determined on the basis of the test data for the Black African group, tend to be associated with smaller differences in item difficulties between the White and Black African samples. If the Raven’s matrices test was measuring the same construct (g?) in the White and Black African groups, one would expect similar positive and significant correlations between the g loadings of the Raven’s matrices items for those respective groups and the between-group differences in item difficulties that apply within the respective groups.

Our internal consistency results are not consistent with the position advocated by Rushton et al. (2004) that the Raven’s matrices test, as a measure of g, is unbiased when applied to the Black African group. We further are left to wonder why Rushton et al. chose to aggregate the data for the Black African and East Indian groups for the purposes of their analyses, so yielding results (even if less than convincing in their own right) that are more supportive of their non-bias hypothesis.

The results reported in Table 1, when compared with those of Figure 1, raise further questions about the internal validity of the Raven’s for Black Africans. As already seen, the difference between item difficulties are accentuated at and below the mid-range of the joint distribution represented in Figure 1 with disproportionately higher item difficulties for the Black Africans as compared with Whites within this range. If we read Figure 1 in tandem with Table 1, we see a finding that is counter to that expected under classical measurement theory or CMT (Nunnally &

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.</td>
<td>.69*** (.75***).</td>
<td>0.00 (.14).</td>
<td></td>
</tr>
<tr>
<td>2.</td>
<td>.27 (.25).</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Notes: First correlation for each entry is Pearson’s product moment; second correlation in parentheses is Spearman’s ρ.

*** p < .001.

---

International Journal of Selection and Assessment

© 2006 The Authors
Journal compilation © Blackwell Publishing Ltd. 2006
Bernstein, 1994). CMT states that higher item discrimination indices in the mid-range of item difficulty will occur if the intelligence test is construct valid for the given group. We see supportive evidence for the validity of the Raven’s when the test is administered to the White group (i.e., items toward the mid-range of item difficulty in Figure 1 are more highly correlated with the total Raven’s score), but non-supportive, in fact disconfirmatory, evidence for the validity of the Raven’s test, when it is administered only in the Black African group (i.e., item discriminations are especially poor for the Black African group at and below the mid-range of item difficulty). It appears that CMT works in support of the validity of the Raven’s for the white group, but the application of scientific measurement principles again demonstrates the invalidity of the Raven’s for measuring g in the Black African group.

Internal Validity Criteria for Test Bias: Comparison of Group Differences in Factor Loadings for the Three Ethnic Groups

Rushton et al. (2004) present results here that initially appear to be compelling in their support of the non-bias hypothesis. They follow a procedure previously reported by Ree and Carretta (1995) who report confirmatory factor analysis results for ethnic differences in factor loadings across subtests of the Armed Services Vocational Test Battery (ASVAB). Rushton et al. (2004) investigate the factor invariance of Raven’s items, rather than subtests, across racial groups; a methodological extension that appears appropriate. Using confirmatory factor analysis, Ree and Carretta (1995) found excellent comparative fit of the single-factor g model across U.S. White, Black, and Hispanic groups, as did Rushton et al. (2004) in their CFI comparison across the Black, East Asian, and White African groups. In other words, in neither study are there appreciable differences in the proportion of common variance accounted for by g across the ethnic/racial groups. However, when Ree and Carretta (1995) reported the results of univariate tests across the factor loadings of the 10 ASVAB subtests on hierarchical g, “[a]ll of the loadings on g were found to be significantly different for Whites versus Blacks” (p. 273). Although Rushton et al. (2004) report factor loadings by the three ethnic groups in their Table 3, and the differences are at least as pronounced as those reported by Ree and Carretta (1995), they do not report univariate tests between factor loadings across the race groups as did Ree and Carretta. We are therefore left to wonder if the factor structure of cognitive ability is indeed “equivalent for Africans and non-Africans” (p. 225) as Rushton et al. report.

We conducted a straightforward analysis of the data reported by Rushton et al. in their Table 2 to address group differences in g loadings across the three ethnic groups. When the Raven’s item discrimination indices from their Table 2 are correlated between the three ethnic groups, they are found to be completely unrelated ($r = .00$, Table 2) when the Black African and White groups are compared. A positive correlation should be found if g is being measured equivalently by the Raven’s across these two groups. The zero correlation between the White and Black African item discrimination indices is not surprising given the evidence in the previous two sections showing that the Raven’s is invalid for measuring “g” in the Black African group, but is valid for measuring g in the White group. Apparently the Raven’s is measuring a different construct in the White and Black African groups. The case for test bias of the Raven’s against Black Africans continues to get stronger.

External Validity Criteria for Test Bias: Comparison of the Relationship of the Raven’s Scores with Criteria of Educational Success

In this area, we review Rushton et al.’s analyses under three separate subsections: differential validity, predictive bias using the Thorndike model, and predictive bias using the Cleary model.

Differential Validity

Rushton et al. (2004) conclude that the Raven’s is not differentially valid across the three ethnic samples (i.e., the Raven’s correlates with academic success equally well across the three samples of test takers). This conclusion is consistent with non-bias in g loaded tests. However, the power to find significantly different validity coefficients for the Black African vs. the non-Black African group is low given the sample sizes and statistical tests employed by Rushton et al. (e.g., Aguinis, Boik, & Pierce, 2001 have done considerable work on estimating the power of between-groups comparisons using correlational statistics). Hence, Rushton et al.’s conclusion that the same pattern of correlations between the various test scores and the two sets of grades existed for both the Black African and non-Black African students, indicating a “lack of bias” (p. 226), cannot be made with any confidence with respect to the Raven’s.

Predictive Bias as Assessed by the Thorndike Model

Rushton et al. do not report having assessed bias in the Raven’s using the Thorndike model. The Thorndike model (Thorndike, 1971) states that a test is biased if the difference between the mean test scores of two groups (here Black and White African) is greater than the difference between their mean ratings on criterion performance. The Thorndike model has been rejected by many researchers (e.g., Jensen, 1980) as a standard for assessing test bias. However, other researchers have argued that the Thorndike model is in fact appropriate for evaluating test
bias (see Chung-Yan & Cronshaw, 2002). Rushton et al. (2004, p. 224) report that the Raven’s total score difference between Black Africans and non-Black Africans was 2.02 SDs based on a total (pooled?) SD of 5.98. Unfortunately, they do not report a comparative statistic for the two educational criteria they considered. This is an unfortunate exclusion from their article because it precludes drawing any conclusions about test bias using the Thorndike model.

**Predictive Bias as Assessed by the Cleary Model**

Psychologists have been much more accepting of the regression-based Cleary model for test bias and in fact propose it as the gold standard for determination of test bias. For example, the 4th Edition of the *Principles for the Validation and Use of Personnel Selection Procedures* published by the Society for Industrial and Organizational Psychology (2003) state:

> Predictive bias is found when for a given subgroup, consistent nonzero errors of prediction are made for members of the subgroup...Although other definitions of bias have been introduced, such models have been criticized and found wanting on grounds such as lack of internal consistency... (p. 32).

Interestingly, when the Cleary model is applied “in the cognitive ability domain” the *Principles* state that slope differences in regression lines for White and African American subgroups are “rarely found.” They further state that “...while intercept differences are not uncommon, they typically take the form of overprediction of minority group performance” (p. 32). The thinly concealed subscript to this statement is that cognitive ability testing in fact disadvantages Whites as compared with Black Americans! Such statements no doubt would seem fanciful and even ridiculous to many non-experts, but most I/O Psychologists apparently take them as fact.

Cronshaw, Caiger, Dupont-Chalaoui, and Chung-Yan (2001) flatly rejected the view widely accepted in the testing community that cognitive ability tests have been definitely demonstrated as being unbiased using the Cleary rule. They calculated a mean level of statistical power of .14 for a p < .05 for 44 previously reported minority/non-minority sample comparisons where job performance was regressed against cognitive ability tests scores. Interestingly, they found that 12% of the minority regression slopes were significantly different – A finding that would be expected if slope-based differences were prevalent across the samples. These results are consistent with statistical simulations demonstrating the low power of tests for regression slope differences under typical testing conditions (Aguinis & Pierce, 1999). Not surprisingly then, Rushton et al. (2004) do not find significant differences for the regression line slopes for the Black African and non-Black African samples when they regress a composite of high school and university course grades against a composite of three g-loaded tests, including the Raven’s, illustrated in their Figure 4 (unfortunately, they do not report regression results for the Raven’s in a Black/White African comparison). They report significant intercept differences between the two groups, with the regression line for the Black African group lying above the non-Black African group at the lower test score ranges and converging with the non-Black African group at the higher test score ranges. A common regression line for the two groups would therefore underpredict the test scores of Black Africans relative to non-Black Africans throughout much of the test score range and this tendency becomes more pronounced as the test scores decrease. In fact, Rushton et al.’s findings contradict the assertion in the *Principles* that minority group performance is overpredicted by cognitive ability tests. Even more astonishingly, Rushton et al. (2004) conclude that their predictive validity findings are consistent with non-biased tests – even when the bias is reported, diagrammed, and obvious to even the most cursory examination!

**Alternative Explanations of Group Differences on the Raven’s Advanced Matrices**

There are at least two other problems with the Rushton et al. (2004) study that do not directly involve statistical considerations. These are discussed below. Interestingly, they provide some additional understanding into the possible source of the test bias we found on our re-examination of the Rushton et al. findings.

**Use of Non-Equivalent Groups in Test Samples**

Rushton et al. (2004) define the Black African sample as a “highly select” sample, given that they are students in a university engineering program. This description of the sample gives readers the impression that the Black South African sample group is substantially equivalent (on relevant social characteristics) to the comparison samples of White and East Indian South Africans and, additionally, with reference to comparison samples of engineering students in the United States. However, the three sample groups differ greatly on many social and economic characteristics. South Africa has a well-documented record of legal (and de facto) racial discrimination. Indeed, South Africa’s history of segregation and apartheid has been described as one of the most extreme forms of racial discrimination in the 20th-century world (Beinhart & Dubow, 1995).
Under the White-dominated South African Nationalist party, the Bantu Education Act (47) was introduced. This 1953 Act established a Black Education Department, in the Department of Native Affairs, which compiled an education curriculum that was specifically designed to constrain the skill levels of Black Africans so as to “fit them better” for their designated low-level roles within the apartheid economy (Butler, 2004). The result of this type of government-sanctioned racial discrimination was the implementation of an education system for Black South Africans that was greatly inferior to that of White South Africans.

The discrepancies in educational opportunity among White and Black South Africans were glaring (opportunities for Asian groups fell somewhere in between). There were large differences in the level of education among teachers in schools for White South Africans and teachers in schools for Black South Africans. In 1978, all White teachers had at least 10 years of schooling and one-third of them had degrees whereas only 16.09% of Black teachers had more than 10 years of education. Only 2.45% of Black African teachers held degrees (Davenport, 1987). Teacher–pupil ratios in primary schools averaged 1:18 in White schools, 1:24 in Asian schools, 1:27 in Coloured schools, and 1:39 in Black schools (see Kallaway, 1984; United States Library of Congress, 2003).

Only in 1996, was education rationalized into a single education system. However, the apartheid inheritance lives on in the privileges accorded to institutions that once serviced White South Africans exclusively (Butler, 2004). Educational discrepancies among ethnic groups still exist to date. Teachers remain under-qualified and facilities remain uneven (Butler, 2004). It is widely acknowledged that it will take many more years, and significant resources, before there is any semblance of equality in the South African education system.

It is reasonable to assume, given these facts, that the Black South Africans in the Rushton et al. (2004) sample have an educational (and socio-economic) background that is considerably inferior to that of their White counterparts. In order to meet the country’s growing demand for more engineers, affirmative action policies have been introduced at universities, including the University of Witwatersrand (where Rushton collected his data; University of the Witswatersrand, Johannesburg, 2005a), in order to increase the representation of Black South Africans who make up the bulk of the country’s population (Boroughs, 1999). However, having a second-rate apartheid education, coupled with having learned English as a second language, Black South African students are largely ill-prepared for the demands of the program (Boroughs, 1999). In an attempt to remedy educational disadvantages imposed on Blacks in the apartheid era, many South African universities (including the University of Witwatersrand’s Faculty of Engineering and the Built Environment) have implemented supplementary mediation-learning programs (University of the Witswatersrand, Johannesburg, 2005b) in a bid to offset the negative effects of Black South Africans’ vastly inferior primary and high school education and the negative effect of coming from disadvantaged backgrounds (Boroughs, 1999).

Differences in educational opportunity provide a viable hypothesis for the predictive test bias against Black Africans found on re-examination of the Rushton et al. results reported above. There is clear evidence of under-prediction of Black African criterion scores (i.e., of high school and university educational success) at the lower cognitive test scores and this bias becomes increasingly prominent as we move progressively lower in the test score range. This is exactly what we would expect if we assume that (1) higher cognitive test scores are at least partly a function of more and better educational opportunity; (2) most, but not necessarily all, of the Black Africans in the sample received inferior educational opportunities to the individuals in the non-Black African comparison group; and (3) the individuals in the Black African sample have undertaken independent efforts to compensate for the poor quality in their formal education (e.g., though self-study). The net effect of these three factors would produce the predictive test bias so evident in the Rushton et al. paper (even if this test bias remains unrecognized by them). In support of points (1) and (2) above, a recent study on group differences in Cognitive Ability Test (CAT) performance stratified its South African sample so that the comparison groups of both White and Black participants had English as their first language, and were in attendance at an advantaged school. In that study, CAT scores were found to be largely equivalent, thus demonstrating the importance of educational quality in influencing CAT performance (Shuttleworth-Edwards, Kemp, Rust, Muirhead, Hartman, & Radloff, 2004). The Shuttleworth-Edwards et al. (2004) paper tells a powerful and distressing story about Black African disadvantage as one of the delayed legacies of the South African apartheid period. We believe that this educational disadvantage historically experienced by Black South Africans has compromised the ability to validly and equivalently measure g through use of the Raven’s. Clearly, this is not the story intended by Rushton et al. who, along with the larger I/O psychology profession, are firm in denying the existence of cognitive test bias. However, this is an explanation that is very plausible in light of Rushton et al. (2004) results.

**Stereotype Threat and the Time Allotment for Test Administration**

Considering South Africa’s political history, and the extent to which affirmative action programs were used to increase Black South African students’ admission to university programs (Boroughs, 1999) – including the Faculties of Engineering and the Built Environment at the University of
Witwatersrand – it is likely that the Black participants in Rushton et al.’s sample were aware of the negative stereotypes that exist surrounding the performance of Blacks on scholastic or intellectual tasks. Extant research has demonstrated that fearing one’s performance on a test of intelligence will confirm an unfavourable stereotype about one’s own group can impair performance on the test (e.g., Steele & Aronson, 1995). This phenomenon, labelled stereotype threat, can cause Blacks to underperform on cognitive ability tests, compared with Whites, particularly when participants are told that a test is diagnostic of ability (Steele & Aronson, 1995). In fact, previous studies have shown that simply asking participants to indicate their ethnicity at the outset of a testing session is enough to induce stereotype threat (e.g., Shih, Pittinsky, & Ambady, 1999). Regrettably, Rushton et al. make no reference as to how they obtained participants’ demographic information or what participants were told before they took the test. This omission is unfortunate because it is possible that participants’ demographic information was obtained at the outset of the testing session, or that participants were told that the test was diagnostic of their ability. The effect of inducing stereotype threat would be to lower mean CAT scores in the Black African group as compared with the non-Black African groups, without a corresponding effect of lowering the criterion scores for the Black African group. The results would again be predictive test bias of the type found by Rushton et al. (2004).

Rushton et al. (2004) acknowledge that one potential threat to the validity of their study is the 30 min time limit that they imposed on their participants for completing the Raven’s Advanced Progressive Matrices. The authors mention that this time limit may have lowered the scores of the Black African students if we can assume that, due to less exposure to Westernized testing practices in the school system, Black Africans were overrepresented among those who tended to work more slowly. In order to understand the possible impact of this validity threat, we will outline some information presented by the publishers of the Raven’s Matrices (Raven, Raven, & Court, 1998). There are two different approaches to administering the Raven’s Advanced Progressive Matrices: untimed and timed. If it is timed, the test provides a measure of intellectual efficiency, a useful measure when assessing suitability for tasks requiring fast and accurate judgements (Raven et al., 1998). Although intellectual efficiency is a closely related construct to general intelligence, the two are not synonymous, and thus should not be treated as such. Furthermore, Raven et al. (1998) state that a 40 min time limit is standard for measuring intellectual efficiency. However, Rushton et al. (2004) imposed a 30 min time limit, ostensibly because of class time restrictions. Because the researchers did not administer the test according to standardized test guidelines, it is inappropriate for them to compare participants’ scores with norms established under appropriate standardized testing conditions.

As stated by Raven et al. (1998), because Set II of the Advanced Progressive Matrices has 36 items arranged in ascending order of difficulty, not all test takers will necessarily attempt all test items before stopping. If members of the Black African group tended to work more slowly and carefully than the non-Black African groups, it is likely that they did not have a chance to answer the problems nearing the end of the test. They may also have been forced to rush through these problems because of time restraints. Additionally, it is possible that Black African students’ fear of confirming negative stereotypes about their group’s performance influenced the speed at which they worked and, as a result, they worked slower than the comparison White and East Indian groups. This becomes particularly problematic when examining the Rushton et al. (2004) standardized factor solution for the hypothesis of general intelligence (g) factor model. The items they selected for their factor model (items 18, 20, 24, 25, 26, 27, 28, 29, 30, 21, 32, 22, 34, 35) all fell at the end of the test. These items represent the most difficult problems in the test. Many of the Black African students did not answer one or more of these problems that were answered by members of the other groups, due to the unstandardized time limit imposed. Rushton et al.’s unorthodox administration of the Raven’s test makes their results difficult to interpret from the perspective of conventional test theory and further muddies the interpretation of their work.

A Brief Overview of Rushton’s Related Research Work

The Rushton et al. (2004) paper cannot be viewed as being independent of Rushton’s related work on race and intelligence. Indeed, the paper is part of a broader Black–White hereditarian research program by Ruston and his colleagues on the relationship between intelligence and a number of individual difference and societal level variables. As the most recent example, Rushton and Jensen (2003a) review 10 categories of evidence to argue in support of their hereditarian position, one category of which draws heavily on Rushton’s research into mean race–IQ differences between Whites and sub-Saharan Africans. This race–IQ research in turn overlaps extensively in authorship and approach with the Rushton et al. (2004) article that we are critiquing here (the other papers include Rushton & Skuy, 2000; Rushton, Skuy, & Fridjohn, 2002, 2003; Skuy, Gewer, Osrin, Khunou, Fridjhon, & Rushton, 2002). Rushton and Jensen further argue in their 2005 review paper that:

... because the tests show similar patterns of internal item consistency and predictive validity for all groups, and because the same differences are found on relatively culture-free tests, many psychometricians have concluded that the tests are valid measures of racial differences, at
least among people sharing the culture of the authors of the test (Rushton & Jensen, 2005a, p. 241).

Leaving aside the last clause of this sentence (which appears to us to invalidate the argument made by Rushton and Jensen in the rest of the sentence for using IQ tests in the assessment of sub-Saharan Africans), the position taken by these authors is clear. The case of race differences in IQ will be invalidated if IQ tests do not meet the psychometric criteria set out by Rushton and Jensen (2005a). This paper has shown that at least one published set of results by Rushton and his colleagues fail to meet the scientific burden of proof that they themselves set for their work. We encourage other researchers to critically and empirically re-examine the validation evidence provided by Rushton and his colleagues in all their related work on mean race–IQ differences between Whites and sub-Saharan Africans. These re-analyses and re-evaluations are badly needed given the implications of their findings, and their interpretations of those findings, for science and practice in psychology.

Rushton, Jensen, and other “racial realists” (their self-label) state that sub-Saharan Black Africans have a mean IQ of 70 (Rushton & Jensen, 2005a). Combined with their hereditarian argument that 80% of race differences in IQ are genetically based (Rushton & Jensen, 2005b), the public policy implications for disadvantaged groups, especially Blacks, and developing societies in the third world, are dire to say the least. Rushton links lower Black IQs to smaller brains, a genetically based tendency to commit crime due to higher testosterone levels, a genetically based tendency to have many children and invest little care in them (“r vs. K selection theory”), and poor genetic potential for “social organization” (e.g., Rushton, 1995). The Rushton et al. (2004) paper is but one contribution to a large corpus of related work conducted by Rushton and his colleagues. Its results and implications should be interpreted and understood in that broader context.

We believe it is fair to point out that Rushton has not shied away from bringing his racial-realist theories into the public sphere. For example, Rushton has been a featured speaker at the biennial conferences of an organization known as the American Renaissance in 1996, 1998, 2000, and 2002, where he presented his analysis of racial differences in crime and intelligence and his theory of the biological basis for nationalism and ethnocentrism, as well as his take on South African students and test bias. Rushton has also posted a statement summarizing his views of “Race as a biological concept” on the American Renaissance website (www.amren.com/Rushton.htm). The American Renaissance Online magazine and archives consist of articles about the dangers of Black crime, the rape of White women by Black men, the general mistreatment and denial of rights of White people, the dangers of non-White immigration to America, and the immutability of racial differences. In a number of American Renaissance conference speeches and magazine articles, the solution to this problem is made clear: the forced separation of the races in North America, either by dividing the country or removing non-Whites (e.g., McCulloch, 1995). At the 1996 conference of American Renaissance, Rushton ended his talk by stating “any country with large black populations will have large problems,” and received a standing ovation (Brown, 1996, p. 8). Further, at the 2002 conference, Rushton presented data on Black–White IQ differences in South African university students. Rushton’s books are sold, his papers are reproduced, and his work discussed on White supremacist and neo-Nazi websites around the world. These include openly Nazi organizations such as Stormfront (www.stormfront.org) or the National Renaissance (www.natvan.com) as well as the more sanitized but viciously antisemitic website of former Ku Klux Klan and National Socialist White People’s Party leader David Duke (www.davidduke.org, see Rose, 1992). On the discussion forums of these groups, the alleged lack of bias in the Raven’s Progressive Matrices is presented a key link in “proving” White racial superiority. The use of psychological research by racist groups has a long history and shows no sign of ending (see Guthrie, 1998; Jackson, 2005; Richards, 1997; Tucker, 1994, 2002; Winston, 1998, 2004). Rushton is certainly not the first psychologist whose work has been of value to neo-Nazi groups, and this phenomenon is part of along tradition of scientific racism in psychology. The reader is invited to investigate these questions further and in-depth by consulting the reference works listed in this paragraph.

We think it is also fair to note that Rushton’s empirical work on race, as well as his interpretations of his and others’ work in this area, have been subjected to very severe scientific criticism from geneticists, evolutionary biologists, psychologists, sociologists, and anthropologists, but it is not the purpose of this article to review these works. These critiques have made the case that Rushton’s work on intelligence testing and race contains fundamental errors, inappropriate conceptualization of “race,” inappropriate statistical comparisons, misuse of sources, and flaws of logic of a very serious nature (e.g., Brace, 1996; Cain & Vanderwolf, 1990; Cernovsky & Litman, 1993; Gabor & Roberts, 1990; Graves, 2002; Lieberman, 2001; Lynn, 1989; Peters, 1991, 1993, 1995a–c; Tooby & Cosmides, 1989; Tucker, 1994, 2002; Wahlsten, 1994; Weizmann, Weiner, Wiesenthal, & Ziegler, 1990, 1991; Winston, 1997; Zuckerman, 1990; Zuckerman & Brody, 1988). Most of these observations are made by writers who emphasize the importance of the scientific disciplines of biology and genetics. Importantly, many of the criticisms made about the broader corpus of Rushton’s work are entirely relevant to the Rushton et al. (2004) paper that we have reviewed here.

Why should any of this matter to the readers of the International Journal of Selection and Assessment? After all, Rushton is entitled to his point of view and has the right in a free society to express his opinions. In further
elucidating our concern, we refer to the first paragraph of the Aims and Scope of this journal:

*International Journal of Selection and Assessment (IJSA) is a quarterly scientific journal that publishes original articles related to all aspects of personnel selection, staffing, and assessment in organizations.*

A scientific journal assumes the responsibility to present research that is conducted and interpreted impartially and objectively, insofar as that is possible. We have noted our concerns that Rushton et al. (2004) fell considerably short of this standard in the scientific rigor of their research paper. As well, we ask I/O and occupational psychologists to consider two other issues when reading and using the article by Rushton et al. (2004): (1) The article, with the defects in analysis and interpretation that we have noted above, should not be read in isolation of other work by Rushton and others that promotes a larger racial-realist research agenda and (2) the results reported by Rushton et al. (2004), along with their interpretation of those results, have extremely important ethical and social policy implications. With these two considerations in mind, we believe that I/O and occupational psychologists should subject the test bias conclusions drawn by Rushton et al. (2004), along with their scientific justification for these conclusions, to close critical scrutiny.

**Conclusion**

If the analysis of research data and the interpretation of research results come into question on scientific grounds (as we have shown here with reference to Rushton et al., 2004 research into test bias of the Raven’s in South Africa), then a tacit agreement by I/O Psychologists with Rushton’s conclusions, by neglecting to sufficiently analyze, critique, and challenge such work, comprises a serious ethical violation because of the underlying societal and economic injustice that can result from application of his work. There is a strong likelihood that tacit acceptance of Rushton’s work on test bias (including that of Rushton et al., 2004) as scientifically valid will promote further disadvantage among Black citizens the world over. As a case in point, Rushton et al. (2004) encourage future researchers to use their conclusions about non-bias in the Raven’s as a rationale for imposing a test cut-off score as a university admission criterion for South African Engineering Students. On the other hand, if IQ tests including the Raven’s are in fact shown to be psychometrically biased against Blacks, then much of the racial-realist agenda that promotes the inherent inferiority of Black mental capability becomes unhinged. The question is a vitally important one. We should devote considerably more attention to an informed study of it.

**References**


Cronshaw, S.E., Caiger, L., Dupont-Chalauvi, V. and Chung-Yan, G. (2001), Why cognitive ability tests may yet prove to be biased using the Cleary rule. Paper presented at the Annual Conference of the Canadian Psychological Association, Quebec City, Quebec.


