Construct Validity of Raven's Advanced Progressive Matrices for African and Non-African Engineering Students in South Africa

J. Philippe Rushton*
The University of Western Ontario

Mervyn Skuy
The University of the Witwatersrand

Trudy Ann Bons
The University of Western Ontario

We test 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 by examining 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, 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 US 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. The results imply that scores on the Raven's Matrices are as valid for Africans as they are for non-Africans.

For almost 100 years, individual and group differences in general mental ability have engendered debate as to how important they are as predictors of work performance (Ones & Viswesvaran, 2002; Viswesvaran & Ones, 2002). As test scores are the best predictor of success in Western society (Schmidt & Hunter, 1998), group differences pose a similar conundrum for industrial–organizational psychologists as they do for school psychologists and social policy analysts (Gottfredson, 2002, 2003). As the trend to the global economy continues, mean group differences in cognitive performance are likely to become more salient, both within and across countries. Questions arise as to how valid the measures are for various populations.

Reviews of the literature in the US and around the world typically find that East Asians and their descendants average an IQ of about 106, Europeans and their descendants about 100, and Africans and their descendants about 85. The lowest average IQ scores are reported for

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

sub-Saharan Africa, about 70 (Jensen, 1998; Lynn & Vanhanen, 2002; Rushton, 2000). These group differences were widely debated after publication of *The Bell Curve* with its original analysis of 11,878 youths (including 3022 Blacks) from the 12-year US National Longitudinal Survey of Youth (Herrnstein & Murray, 1994). It found that most 17-year olds with high scores on the Armed Forces Qualification Test, regardless of ethnic background, went on to occupational success by their late 20s and early 30s, while those with low scores were more inclined to welfare dependency. The study also found that the average IQ for 'African' Americans was lower than those for 'Latino,' 'White,' 'Asian,' and 'Jewish' Americans (85, 89, 103, 106, and 113, respectively, pp. 273–278). Currently, the 1.1 standard deviation difference in average IQ between Blacks and Whites in the US is not in itself a matter of empirical dispute. A meta-analytic review by Roth, Bevier, Bobko, Switzer III, and Tyler (2001) showed it also holds for college and university application tests such as the Scholastic Aptitude Test (SAT; $N = 2.4$ million) and the Graduate Record Examination (GRE; $N = 2.3$ million), as well as for tests for job applicants in corporate settings ($N = .5$ million), and in the military ($N = .4$ million).

Within the US a consensus has arisen that the tests are not biased in any psychometric sense, at least among people *sharing the culture of the authors of the test* (Jensen, 1980; Wigdor & Garner, 1982). For example, several large-scale studies of military samples have shown the construct equivalence of tests among ethnic/race groups. Ree and Carretta (1995) examined a nationally representative sample of young Black, White, and Hispanic men and women who took the Armed Services Vocational Aptitude Battery (ASVAB; $N = 9173$). The ASVAB, which is used to select applicants for all military enlistments and assign them to first jobs, consists of 10 separately scored subtests. Despite the especially wide variety of subtests, Ree and Carretta found the hierarchical factor structure of ASVAB subtest scores was virtually identical across the three ethnic groups. Similarly, Carretta and Ree (1995) examined the more specialized and diverse Air Force Officer Qualifying Test (AFOQT), a multiple-aptitude battery that had been given to 269,968 applicants (212,238 Whites, 32,798 Blacks, 12,647 Hispanics, 9460 Asian Americans, and 2551 Native Americans). The *g* factor accounted for the greatest amount of variance in all groups and its loadings differed little by ethnicity. The American Psychological Association's Task Force on intelligence concurred: 'Considered as predictors of future performance, the tests do not seem to be biased against African Americans' (Neisser, Boodoo, Bouchard Jr., Boykin, Brody, Ceci, Halpern, Loehlin, Perloff, Sternberg, & Urbina, 1996, p. 93).

It is an empirical question how far this conclusion can be universally generalized. It would be useful to have more construct validity data available from other regions of the world especially now that so much new research is being carried out in non-Western countries and some controversial findings and interpretations are being reported. For example, in their book *IQ and the Wealth of Nations*, Lynn and Vanhanen (2002) found that the world average IQ is about 90, and that the mean sub-Saharan African IQ is about 70.

Early reports of the low mean IQ scores reported for sub-Saharan Africans were especially disputed (Kamin, 1995; Nell, 2000; Van de Vijver, 1997). Yet, Lynn and Vanhanen's (2002) review of the literature found this average IQ of 70 in over two-dozen studies from West, Central, East, and Southern Africa. For example, in Nigeria, Fahrmeier (1975) collected data on 375 6- to 13-year olds in a study of the effects of schooling on cognitive development. The children's mean score on the Colored Progressive Matrices was 12 out of 36, which is at the 4th percentile for 9.5-year-olds on US norms, or an IQ equivalent of about 75 (Raven *et al.*, 1990, pp. 97–98). In Ghana, a study by Glewwe and Jacoby (1992) tested a representative sample of 1736 11- to 20-year old primary school graduates and found they averaged 19 out of 36 items correct on the Colored Progressive Matrices – an IQ of less than 70. In a South African study, Owen (1992) administered the Standard Progressive Matrices to 1093 African high-school students and found they solved an average of 28 out of 60 problems, which is near the 10th percentile.

There can be little doubt about the replicability of the mean African IQ of 70, or the impartiality of the investigators, for studies continue to report low scores. Two recent studies by Robert Sternberg are cases in point. In Kenya, Sternberg, Nokes, Geissler, Prince, Okatcha, Bundy, and Grigorenko (2001) had 85 12- to 15-year olds complete the Colored Progressive Matrices; they averaged 23.5 correct out of 36 (IQ equivalent about 70). In Tanzania, Sternberg, Grigorenko, Ngrosho, Tantufuye, Mbise, Nokes, Jukes, and Bundy (2002) gave the Wisconsin Card Sorting Task to 358 11- to 13-year olds who received a Perseverative Error score of 18.53. Although procedural differences may make the normative comparison problematic, as it stands, this score is equivalent to the fifth percentile on American norms for 12-year olds ($IQ = 75$). After training on how to sort attributes, the children's scores went up to 16.5 (lower scores meant fewer errors), but this was still only at the ninth percentile on American norms ($IQ < 80$). In South Africa, Skuy, Schutte, Fridjhon, and O'Carroll (2001) reported mean scores one to two standard deviations below US norms on a wide array of tests administered individually to 154 high-school students. Included were the Rey Auditory Verbal Learning Test, the Stroop Color Word Test, the Wisconsin Card Sorting Test, the Bender Gestalt Visual Motor Integration Test, the Rey Osterreith Complex Figure Test, the Trail Making Test, the Spatial Memory Task, various Drawing Tasks, and the Wechsler Intelligence Scale for Children-Revised (WISC-R). On the WISC-R, African students averaged -1.81 standard deviations below American norms (-1.58 SDs with the vocabulary subtest excluded).

Seven studies of university students in Africa show a median IQ of 84 (range = 77–103). Assuming that African university students are one standard deviation (15 IQ points) above the population mean, the finding of a median IQ of 84 corroborates the general population mean of 70. In South Africa, one study of 63 undergraduates at Fort Hare University, the University of Zululand, the University of the North, and the Medical University of South Africa, who were given the Wechsler Adult Intelligence Scale-Revised (WAIS-R), obtained a full-scale IQ of 77, 1.5SDs below US norms (Avenant, 1988; cited by Nell, 2000, pp. 26–28). Similarly, Grieve and Viljoen (2000) reported an average score of 37 out of 60 on the Standard Progressive Matrices (IQ equivalent of 78 on US norms) for 30 students in fourth-year law and commerce at the University of Venda in South Africa's Northern Province. Zaaiman, van der Flier, and Thijs (2001) tested 147 first-year mathematics and science students at the University of the North and found an average score of 52 out of 60 on the Standard Progressive Matrices (equivalent to an IQ of 100). Four studies by Rushton and colleagues on several hundred undergraduates at Johannesburg's University of the Witwatersrand using the Raven's Matrices found an average IQ of 84 for psychology students (Rushton & Skuy, 2000; Skuy, Gewer, Osrin, Khunou, Fridjhon, & Rushton, 2002) and of 100 for engineering students (Rushton, Skuy, & Fridjhon, 2002, 2003).

Critics, however, claim that Western-developed IQ tests are not valid for sub-Saharan Africans because of the large cultural differences that are also found (Kamin, 1995; Nell, 2000; Van de Vijver, 1997). Thus, Nell (2000, pp. 35–46) provided an 'environmentalist-cultural-relativist' hypothesis based on Marx, Vygotsky, and Luria (e.g., Luria, 1979) to suggest that the tests have different meanings for Africans than they do for non-Africans, and that Africans obtain lower mean scores because they are, on average, less testwise, less interested, more anxious, or give up sooner on items they find difficult than do non-Africans. For these reasons, Nell (2000) contends that standardized tests should not be used in South Africa.

Psychometric test biases are of two main types: those *internal* and those *external* to the test. With respect to internal bias, the finding that item difficulties on the Progressive Matrices correlated for Africans, East Indians (those whose ancestors originated from the Indian subcontinent), Whites, and other ethnic groups in South Africa with mean $r = .90$ ($p < .001$) among thousands of high-school students (Owen, 1992), and hundreds of university students (Rushton & Skuy, 2000; Rushton *et al.*, 2003, 2003; Skuy *et al.*, 2002), as well as the finding that the item–total score correlations are very similar across all groups, suggests that the items are measuring the same constructs in all groups. The only reliable example of internal bias so far found in this extensive literature is the rather obvious linguistic one of the Vocabulary components of tests like the Wechsler for groups that do not have English as their first language (e.g., in Skuy *et al.*, 2001). Even then the language factor accounts for only about .5 of a standard deviation, out of the overall 2.0 standard deviation difference between Africans and Whites.

The claim of internal bias is also contradicted by studies showing that African–White IQ differences are mainly on *g*, the general factor of intelligence. Jensen (1980, p. 535) formally designated the view that Black–White differences were largely a matter of *g* as 'Spearman's hypothesis,' because Spearman (1927, p. 379) was the first to suggest it. Osborne (1980) 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. Subsequently, Rushton (1998) proposed that the term 'Jensen Effect' be used whenever a significant correlation occurs between *g*-factor loadings and any variable, *X*; otherwise there is no name for it, only a long explanation of how the effect was calculated. Within the US, Jensen's (1985; see also 1998) analysis of 17 independent data sets of nearly 45,000 Blacks and 245,000 Whites derived from 171 psychometric tests found that the *g* loadings consistently predicted the relative magnitude of the Black–White differences ($r = .63$; Spearman $\rho = .71$, $p < .05$). Similar results have been obtained comparing the majority Dutch population with immigrants from the Third World who now comprise 6% of the Dutch population (te Nijenhuis & van der Flier, 1997; te Nijenhuis, Evers & Mur, 2000; te Nijenhuis, Tolboom, Resing, & Bleichrodt, 2004).

Sub-Saharan African–White differences are also found to be mainly on *g* (i.e., to be 'Jensen Effects'). Studies have been carried out using the Junior Aptitude Test, the Wechsler Intelligence Test-Revised, and the Progressive Matrices. Thus, an analysis of 10 subtests of the Junior Aptitude Test for over 3000 African, East Indian, and White high-school students found the African–White differences were mainly on the *g* factor (Lynn & Owen, 1994). Similarly, analyses of 10 subtests of the WISC-R for 154 high-school students in South Africa, as well as for 204 12- to 14-year olds in Zimbabwe, found the African–White differences were mainly on *g* (Rushton, 2001; Rushton & Jensen, 2003). Analyses of the items on the Progressive Matrices in 4000 high-school students and in several hundred engineering and psychology students in South Africa, found the four-way African–Colored–East Indian–White mean differences were all on *g* (Owen, 1992; Rushton, 2002; Rushton & Skuy, 2000; Rushton *et al.*, 2002, 2003). These results suggest that the items measured the same underlying construct in both African and non-African groups.

With respect to the claim of external bias, Kendall, Verster, and Von Mollendorf's (1988) review showed that test scores are about equally predictive for both Africans and non-Africans (e.g., .20–.50 for both school grades in students and job performance in employees). They also showed that many of the factors influencing scores in Africans are 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). Sternberg *et al.*'s (2001) study of Kenyan 12- to 15-year olds found that IQ scores predicted school grades with a mean $r = .40$, $p < .001$ (and continued to do so after controlling for age and socio-economic status, $r = .28$, $p < .01$). Similarly, Rushton *et al.*'s (2002, 2003) study of engineering students found that both the African and the non-African students' Advanced Matrices scores correlated with their scores on the Standard Matrices measured three months earlier (.60 for Africans; .70 for non-Africans) and their end-of-year exam grades measured 3 months later (.34 for Africans; .27 for non-Africans).

The present study provides a more definitive test of whether test bias exists in the South African context by using a wider array of criteria and a fuller set of analyses than in our previous research. The construct validity of IQ scores is examined in a highly select group of African, East Indian, and White engineering students on the more difficult Raven's Advanced Progressive Matrices. The existence of *internal* bias is assessed by examining whether the item structure and *g* loading patterns are different for the three racial groups, and whether African/non-African group differences are specific (as predicted by the relativist hypothesis) or pronounced on the *g* factor (as hypothesized by Spearman, 1927; see Jensen, 1998). The existence of *external* bias would come from a demonstration that scores on the advanced matrices do not predict equally well for the different groups the other criteria in this study, such as scores on the English Comprehension Test, the Similarities subscale from the South African Wechsler Intelligence Scale, the end-of-year university grades, and the high-school grade point average.

Method

Subjects

An initial pool of 392 first-year students from the Faculties of Engineering and the Built Environment at the University of the Witwatersrand took the Advanced Progressive Matrices. This was reduced to 306 17- to 23-year olds by excluding those without examination scores ($n = 27$), or high-school grades ($n = 16$), or biographical data ($n = 14$), or who had listed their age as over 23 ($n = 14$), or who had self–identified as 'Colored' (the accepted designation for South Africa's mixed-race population) or 'Other' for who the sample sizes were too small for separate analysis ($n = 15$). Included in the analyses are 177 Africans (142 men, 35 women), 57 East Indians (45 men, 12 women), and 72 Whites (65 men, 7 women).

Test Instrument

The most researched and widely used of all culture-reduced tests are the Raven's Colored Progressive Matrices, the Raven's Standard Progressive Matrices, and the Raven's Advanced Progressive Matrices (Raven, 2000). The most difficult of these is the Advanced Matrices, which was constructed for those of higher intellectual ability such as students for advanced scientific or technical studies (Raven, Raven, & Court, 1998, pp. 1–2). The reliability and validity of the test has been established across a range of populations, including African Americans and other non-Whites. The total score on the Raven's has proven to be a good measure of *g*, the general factor of intelligence, at least within the US (Jensen, 1980; Raven *et al.*, 1998; Vernon, 1983).

The Advanced Matrices are published in two sets. Set I contains 12 diagrammatic puzzles, each with a missing part that one must attempt to identify from several options. It is typically used for practice and to reduce anxiety. Students in this study took home the Set I puzzles and the correct answers to allow them to practice for a week prior to testing. Set II has 36 puzzles identical in presentation to those in the practice set. The problems are presented in a bold, accurately drawn, and pleasant looking format to maintain interest and minimize fatigue. In accordance with manual guidelines, a time limit of 30 min was given for completing the Advanced Matrices because of constraints on class time. Additionally, participants completed a 10-min English Comprehension Test as well as a 10-min Similarities Test from the new South African Wechsler Adult Intelligence Scale. High-school grades and end-of-year university course grades for these students were obtained from the Dean's Office.

Procedure

One of the authors of the study (M.S.), and his colleagues administered the Advanced Progressive Matrices, the English Comprehension Test, and the Similarities Test during a regularly scheduled class period of 50 min duration. All students appeared to be well motivated. The instructions requested students to wait quietly at their desks if they finished before 50 min. As they left the room at the end of the 50-min class period, they handed in their answer sheets and test booklets, and received payment of 50 rand (at that time about U.S. $8) each. The IQ and ability tests were administered in March of 2002 and the course evaluations gathered in December 2002. (In the Southern Hemisphere, December is the end of the academic year and the beginning of the long summer break.) The high-school grades from the Dean's office were from December 2001, the year before the students entered university.

Results

For the Raven's Advanced Progressive Matrices, all calculations are based on raw scores, with each of the 36 items scored as 0 (incorrect) or 1 (correct). Internal consistencies based on Cronbach's α were .86 for the sample as a whole ($n = 306$), .86 for the Africans ($n = 177$),

.79 for the East Indians ($n = 57$), and .75 for the Whites ($n = 72$). For some analyses to be reported, the smaller samples of East Indians and Whites will be grouped together as 'non-Africans' to facilitate tests of the validity of the lower scores of Africans.

The mean scores on the Raven's for the African, East Indian, and White subjects were 23, 26, and 29, respectively (SDs = 6.1, 4.8, and 4.1; ranges = 3–35, 11–34, 16–35). Men averaged the same scores as women (means = 25, 24; SDs = 6.2, 4.3; ranges = 2–35, 14–35), although they showed more variability. Analysis of variance (ANOVA) with race and sex as factors showed a significant effect only for race (three groups), with no effect for sex, either as a main effect or in interaction, $F(2, 300) = 8.34$, $p < 0.01$; $F(2, 300) < 1.00$; and $F(2, 300) < 1.00$. For the total score, the African/non-African difference was 2.02SDs (based on total SD of 5.98). Using the 1993 US norms for 18- to 22-year olds the Africans, East Indians, and Whites fall at the 60th, 71st, and 86th percentiles, respectively, yielding IQ equivalents of 103, 108, and 118 (Raven *et al.*, 1998).

Figure 1 shows the percentage distribution for the Africans and for the non-Africans (East Indians and Whites combined) who attained various raw scores. It shows a substantial overlap of the two groups at the high end and an elongated tail at the low end of the African distribution. (Subsequent analyses, not reported here, which re-examined all variables after eliminating the outliers on the lower half of the distribution did not change the results.)

Item Analyses and Differences in *g*

Table 1 shows the proportion of Africans, East Indians, and Whites who selected the correct answer for each of the 36 items. For all groups, the test items increased in difficulty as the test progressed (mean $r = .92$, $p < .001$). The item difficulties, measured by the proportion getting the correct answer (Table 1), were similar for all groups ($r > .95$, $p < .001$). This suggests that the test measures the same construct in all three groups. However, even on this Advanced test, 17 items (47%) proved too easy for the Africans, 20 (56%) did so for the East Indians, and 27 (75%) did so for the Whites. Across all groups, only 38 of the 108 items (35%) had an optimal pass rate for maximum discriminatory power of between 30% and 70%. Only six of the items were too difficult, with a pass rate between 0% and 29% (four of them for Africans).

Another index of item similarity across groups is the item–total correlation (r_{it}) calculated using the biserial (r_b) correlation (see Table 2) 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 same construct measured by the test as a whole, as well as how well the item discriminates between testees within each group. As 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 also provides an estimate of each item's *g* loading.

To test whether African/non-African differences are more pronounced on the more *g* loaded items, we first followed the procedure used in the studies by Rushton and Skuy (2000) and Rushton *et al.* (2002, 2003), and correlated the item–total correlations shown in Table 2 (which estimate *g*), with the standardized differences between Africans and non-Africans in proportion passing each item shown in Table 1 (which estimates the race effect size). The correlation between the *g* loadings and the

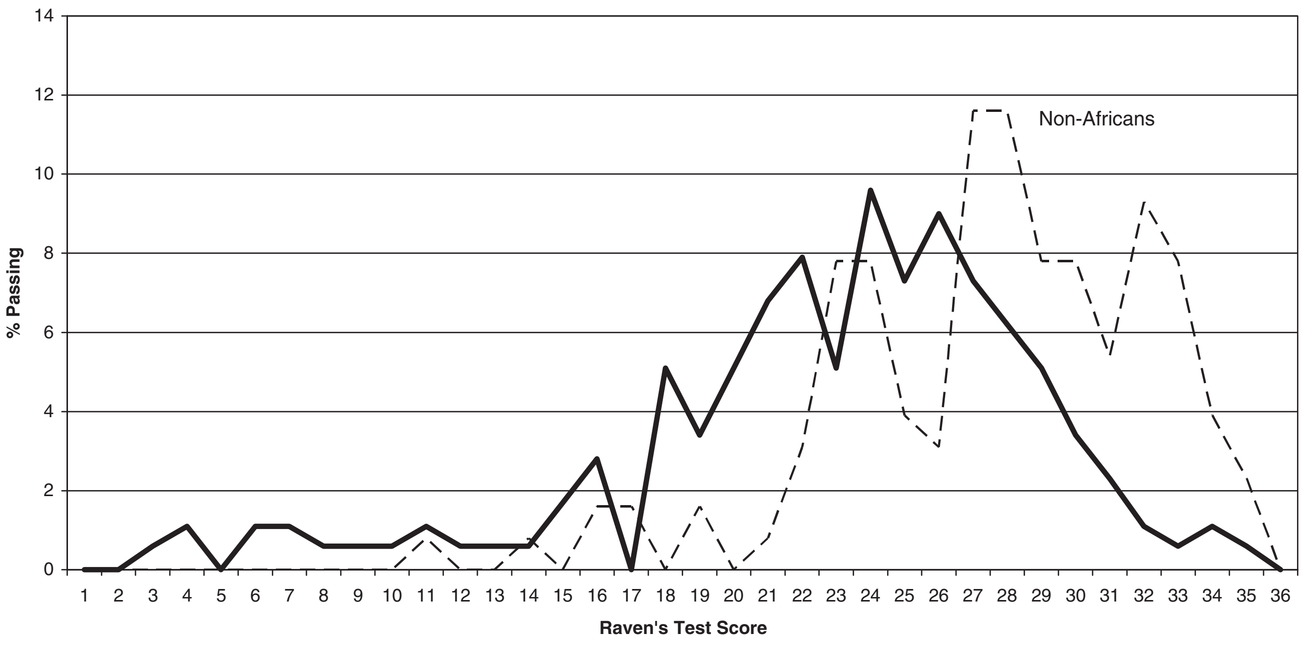

Figure 1. Percentage of 177 African and 129 non-African 17- to 23-year old first-year engineering students attaining various scores on the Raven's Advanced Progressive Matrices Test.

Table 1. Proportion of sample selecting the correct answer on items of the advanced progressive matrices by race

Item	African	Indian	White	Item	African	Indian	White	Item	African	Indian	White
1	.99	1.00	1.00	13	.67	.67	.82	25	.46	.60	.78
2	.94	.96	.99	14	.85	.93	.96	26	.52	.60	.78
3	.93	.98	1.00	15	.77	.88	.89	27	.32	.53	.58
4	.90	.91	.99	16	.80	.86	.94	28	.36	.39	.49
5	.90	.89	.97	17	.74	.82	.89	29	.29	.39	.43
6	.93	1.00	1.00	18	.47	.72	.76	30	.40	.61	.76
7	.91	.96	.97	19	.67	.75	.81	31	.36	.47	.65
8	.86	.88	.96	20	.72	.79	.72	32	.20	.35	.49
9	.90	.96	1.00	21	.67	.84	.93	33	.39	.40	.60
10	.77	.93	.99	22	.58	.70	.81	34	.28	.37	.50
11	.88	.95	.96	23	.62	.67	.86	35	.39	.60	.64
12	.90	.93	.99	24	.42	.61	.72	36	.10	.07	.24

Note: African, $N = 177$; Indian, $N = 57$; White, $N = 72$.

magnitude of the African/non-African differences was .34 ($p < .05$; Pearson's r) and .42 ($p < .01$, Spearman's ρ) using the item total correlations for the non-African group, but only .22 and .21 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.)

To test further whether the African/non-African differences can be attributed in part to the *g* factor, we followed the procedure used in the studies by Carretta and Ree (1995), Ree and Carretta (1995), and performed a multi-group confirmatory factor analysis of the single-factor *g* solution (Figure 2). In doing so we limited the analysis to those 14 items in Table 1, which had difficulty levels 0.20 and 0.80 for all groups. A χ^2 test was conducted to determine if the *g* loadings for the three groups were the same. To do this, the model parameters were first estimated with the loadings for all groups constrained to the same value, and then with the constraint removed, thereby letting each group's factor loading have a different value. If the χ^2 of the difference between the constrained and unconstrained models is not significant, the loadings on *g* are the same for all three ethnic groups; if the χ^2 were significant, separate parameter estimates for each group would provide a better fit. The results indicated that the same model – single-factor *g* – fit the data for all three ethnic groups (e.g., $\Delta\chi^2 = 34.28$, df = 26; CFI = .938; TLI = .934; IFI = .941; GFI = .897; also RMSEA = .015, $\chi^2 = 275.09$, df = 257, $p < .001$). The reliable variance in the Raven's because of *g* is: for Africans, 19%; for Whites, 22%; and for East Indians, 21%. The standardized factor solution for the hypothesized model is shown in Table 3. Thus, the factor structure of cognitive ability in this study is equivalent for Africans and non-Africans.

Table 2. Item–total correlations for the advanced progressive matrices by race

Item	African	Indian	White	Item	African	Indian	White	Item	African	Indian	White
1	.25	-	-	13	.38	.24	.15	25	.39	.32	.39
2	.51	.34	.20	14	.53	.26	.08	26	.35	.48	.52
3	.58	.42	-	15	.53	.03	.27	27	.28	.17	.53
4	.39	.47	.14	16	.44	.61	.49	28	.26	.33	.36
5	.53	.34	.25	17	.47	.16	.12	29	.25	.28	.36
6	.49	-	-	18	.46	.63	.54	30	.38	.34	.27
7	.52	.20	.02	19	.38	.31	.26	31	.40	.51	.49
8	.30	.47	.27	20	.49	.44	.50	32	.28	.38	.47
9	.56	.54	-	21	.54	.54	.29	33	.37	.27	.14
10	.56	.26	.35	22	.39	.46	.44	34	.43	.39	.59
11	.62	.56	.32	23	.50	.52	.36	35	.39	.43	.46
12	.49	.14	.38	24	.42	.53	.36	36	.24	–.03	.44

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

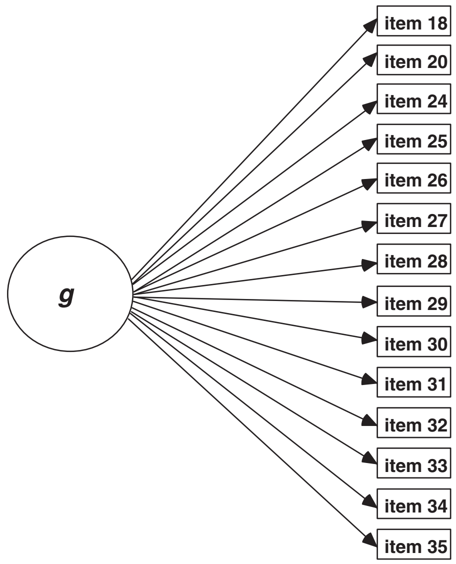

Figure 2. Factor structure model of item *g* loadings for goodness-of-fit test.

Table 3. Standardized factor solution for hypothesized *g* factor model

	African	White	East Indian
Item 18	.427	.546	.497
Item 20	.338	.380	.377
Item 24	.351	.418	.390
Item 25	.267	.347	.278
Item 26	.323	.414	.353
Item 27	.363	.403	.343
Item 28	.294	.314	.318
Item 29	.255	.263	.246
Item 30	.290	.347	.294
Item 31	.400	.443	.414
Item 32	.343	.325	.319
Item 33	.305	.299	.305
Item 34	.582	.627	.543
Item 35	.338	.370	.347

Concurrent Validities

To test for the concurrent validity of the Matrices, the scores for the Africans and non-Africans were correlated with their scores on the English Comprehension Test, the Similarities subscale of the new South African Wechsler Adult Intelligence Scale (total scores as scaled), university course grades (based on an aggregate of at least four courses using standard scores), and high-school grade-point average (recovered from Dean's office). Ethnic differences in mean scores on all these measures paralleled those found on the Raven's Advanced Matrices, with the means for the Africans, East Indians, and Whites, respectively, being: English Comprehension test, 47, 70, and 73 (SDs = 17, 13, 10; $F(2, 300) = 47.20$, $p < .001$); Similarities test, 10, 11.5, and 12 (SDs = 1.7, 1.8, 1.9; $F(2, 300) = 12.3$, $p < .001$); university grades, 56, 58, and 63 (SDs = 12, 11, 9; $F(2, 300) = 3.43$, $p < .05$); and high-school grades, 22, 28, and 28 (SDs = 6.0, 6.6, 7.3; $F(2, 300) = 7.30$, $p < .01$).

The pattern of correlations between the various test scores and the two sets of grades were the same in both the African and the non-African students, thereby indicating the test's predictive validity and lack of bias (Table 4). 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 (Figure 3; t [1,302] = .26, NS). Similarly, a standardized composite of the three cognitive tasks (Raven's, Similarities, and English Comprehension) predicted a composite of the high-school and university grades for both Africans ($r = .32$, $p < .01$) and non-Africans ($r = .35$, $p < .01$), with the slopes of the two regression lines again not significantly different from each other (Figure 4; t [1,302] = .83, NS). Both intercept differences, however, as expected, were significant (Figure 3, t [1,303] = 4.55; $p < .001$; Figure 4, t [1,303] = 3.64;

Table 4. Correlations between Advanced Progressive Matrices (APM), English Test, Similarities Test, High-school Grade Point Average, and University Grade Point Average fro 177 African (above diagonal) and 129 non-Africans (below diagonal)

	Raven's APM	English Test	Similarities Test	High-school Grade Point Average	University Grades
Raven's APM	–	.285**	.144*	.216**	.118
English Test	.253**	–	.256**	.270**	.097
Similarities Test	.259*	.210*	–	.243**	.177**
High-school Grade Point Average	.159*	.390**	.341**	–	.452**
University Grades	.109	.204*	.146*	.639**	–

Note: $^*p < .05$; $^{**}p < .01$.

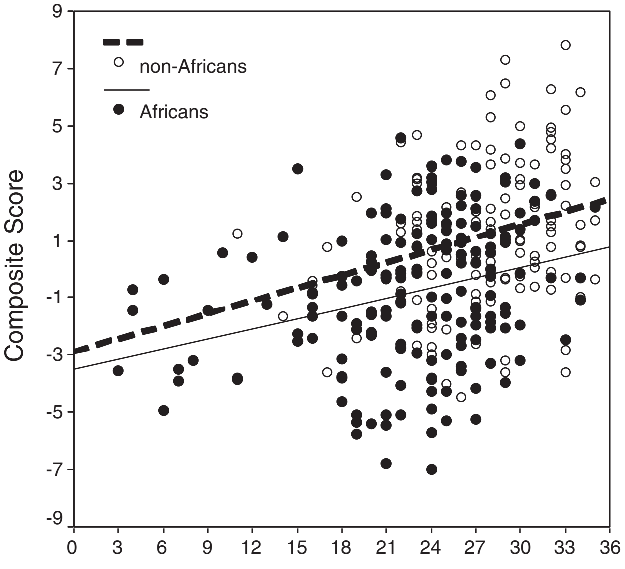

Figure 3. Regression of Raven's Advanced Progressive Matrices scores on a composite criterion of English Comprehension test, Similarity subtest from the South African Wechsler Adult Intelligence Scale, and high-school and university course grades for 177 Africans and 129 non-Africans.

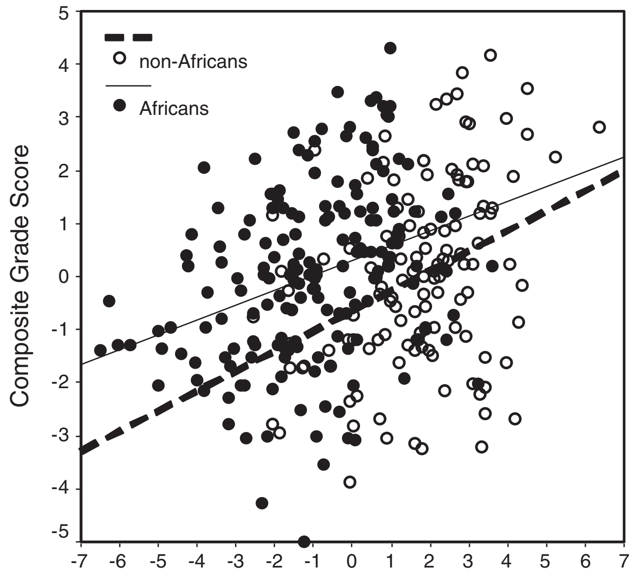

Figure 4. Regression of Raven's Advanced Progressive Matrices scores plus the English Comprehension Test plus the Similarities Scale from the South African Wechsler Adult Intelligence Scale, and high-school and university course grades for 177 Africans and 129 non-Africans.

$p < .001$). Effect sizes (d) obtained on the difference between the regression lines were .53 and .42 for Figures 3 and 4, respectively. Effect sizes of this magnitude are usually considered from moderate to substantial. They show that the average performance of the non-Africans (i.e., the 50th percentile) exceeds the performance of 69% and 66% of the Africans, respectively.

Discussion

The unique contribution of this paper is to show that the logical network of relationships, both within and between tests, is the same for Africans as it is for non-Africans. This conclusion is based on a wider array of criteria and a fuller set of analyses than in any of our previous research (Rushton *et al*., 2002, 2003). We established this equivalence using correlations (A) among the Raven's Advanced Matrices, the English Comprehension Test, the Similarities subscale from the South African Wechsler Intelligence Scale, the end-of-year university grades, and high-school grade point average; and (B) among the items in the test, including from a confirmatory factor analysis. We conclude that these results, as they stand, indicate that both the internal and external validity of the Raven's (and by implication other *g* loaded tests) are sufficient to enable their use in selection among highly educated Africans, despite the large mean group differences between non-Africans and Africans.

The African students solved an average of 23 out of the 36 problems on Set II of the Advanced Progressive Matrices, placing them at the 57th percentile with an IQ equivalent of 103 on the 1993 US norms. The East Indian and White students solved an average of 26 and 29 out of 36 problems, respectively, placing them at the 64th and 86th percentiles, with IQ equivalents of 106 and 117. These African engineering students are the highest scoring African sample tested to date, similar to other African engineering students at the University of the Witwatersrand with an IQ of 103 (Rushton *et al*., 2003), and to the math and science students at the University of the North with an IQ of 100 (Zaaiman *et al*., 2001).

There was no evidence of test bias revealed in this study. Inter-item matrices were the same for African students and for non-African students, both having similar α coefficients, item difficulty levels, and *g* factor loadings. In other words, 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. External validities were the same for both groups, with the Raven's scores correlating with performance on the composite English Comprehension Test, the Similarities Test, the university course grades, and the high-school grade point average ($r = .28$). The validity coefficients are likely lower in these university students than those in less selected samples or in the general population because of the restriction of range effects. These results support those from earlier studies showing the predictive utility of mental ability test scores in Africans (Kendall *et al*., 1988; Rushton *et al*., 2003; Sternberg *et al*., 2001).

The present study supports the view that African/non-African differences reflect the *g* factor of intelligence, rather than any culturally specific ways of thinking. This conclusion must be qualified, however. First, the *g* estimates based on the African group's item–total correlations are lower than those based on the non-African group's, so only a partial demonstration of the effect was achieved using this technique. Second, the confirmatory factor analysis was based on only 14 items, with the data being scored 0 or 1, and with the East Indian and White sample sizes being small ($n = 57$ and 72, respectively). Nonetheless, other studies, using other procedures (e.g., factor analysis of subtests; Rushton & Jensen, 2003) have found the African/non-African differences to be mainly on *g*.

The results presented here corroborate the low mean IQ score of 70 reported for the African general population (Lynn & Vanhanen, 2002) if it is assumed that African engineering students are two standard deviations (30 IQ points) above the population mean. In the US and many developed nations, engineers typically do average up to two standard deviations above the general population. In US, for example, the mean Verbal+Quantitative+Analytic scores of engineering students on the GRE is about 1800, while psychology and education students average about 1500, a difference of about one standard deviation (Educational Testing Service, 1998). Psychology students in turn average about one standard deviation higher than the general population average. Consistently, African psychology students typically average an IQ of about 84, one standard deviation above the general African population (Rushton & Skuy, 2000; Skuy *et al*., 2002). Altogether, the results from seven other studies conducted at universities in South Africa, including by other investigators and at a medical school (described in the Introduction), yield a median IQ of 84 (range = 77–103).

Several caveats need to be raised about our study. One reviewer suggested that it seemed unlikely that Africans with so low an IQ could complete engineering school and then practice the profession, and so these lower scores cannot mean the same as they would for students in the US. Two responses are possible. The first is that the low African mean IQ score *does* accurately represent the probable level of cognitive performance for the population and that, indeed, commensurate work performance is predicted (see Lynn & Vanhanen, 2002). The second is that although individual differences in IQ score *within populations* are predictive of individual differences in various criteria within that population (as in Figures 3 and 4), differences *between populations* are attributable to such factors as poverty and cognitive deprivation so that high motivation is able to outweigh their predicted role in determining performance. According to this view, Euro-American test norms are not valid for Africans. Future research, especially longitudinal studies using some of the real-life criteria identified by Gottfredson (2003) could be undertaken to better resolve this enigma. Other research might address the question of whether the measured group differences are related to educational, socio-economic, and language background in the same way for Africans and non-Africans (Van de Vijver, 1997). Thus more research may yet establish that these lower IQ scores among Africans do not really mean the same as they would for Americans.

There are other potential threats to the validity of our study. We imposed a 30-min time limit for completing the test and this might have lowered the scores of those who work more slowly and carefully. If Africans were over-represented in this group, this would have lowered the scores of Africans. There is also some evidence that the standard deviations are smaller for the non-Africans than for the Africans, perhaps suggesting more selectivity in admissions. Variability reduction in the non-African group would restrict range and systematically depress their correlations, thus making the concurrent validities appear more equivalent than they are.

Overall, the results of this study join those previously reported providing evidence of construct validity for IQ tests among sub-Saharan Africans. The results show that the Raven's Advanced Progressive Matrices measure the same psychological construct in both Africans and in non-Africans, and its scores predicted university grades and other criteria, even for this highly selected group of engineering students. Perhaps future research should be undertaken to ascertain the value of the Raven's test to set a cutoff score for admission to the university program, and thereby minimize the risk of drop out or failure, and increase average grades and future on-the-job performance of South African engineering students, thus enhancing their contributions to the society.

References

Avenant, T.J. (1988) *The Establishment of an Individual Intelligence Scale for Adult South Africans. Report on An Exploratory Study Conducted with WAIS-R on a Aample of Blacks (Report No. P-91)*. Pretoria: South Africa Human Sciences Research Council.

Carretta, T.R. and Ree, M.J. (1995) Near identity of cognitive structure in sex and ethnic groups. *Personality and Individual Differences*, **19**, 149–155.

Educational Testing Service. (1998) *Graduate Record Examinations: Guide to the use of scores*. Princeton, NJ: Educational Testing Service.

Fahrmeier, E.D. (1975) The effect of school attendance on intellectual development in Northern Nigeria. *Child Development*, **46**, 281–285.

Glewwe, P. and Jacoby, H. (1992) *Estimating the determinants of cognitive achievement in low-income countries: The case of Ghana*. Washington, DC: The World Bank.

Gottfredson, L.S. (2002) Where and why *g* matters: Not a mystery. *Human Performance*, **15**, 25–46.

Gottfredson, L.S. (2003) *g*, jobs, and life. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 293–342). New York: Pergamon.

Grieve, K.W. and 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. and Murray, C. (1994) *The bell curve*. New York: Free Press.

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.

Kamin, L. (1995) Lies, damned lies, and statistics. In R. Jacoby and N. Glauberman (Eds.), *The bell curve: History, documents, opinions* (pp. 81–105). New York, NY: Random House.

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, UK: Cambridge University Press.

Luria, A.R. (1979) *The making of mind*. Cambridge, MA: Harvard University Press.

Lynn, R. and 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. and Vanhanen, T. (2002) *IQ and the wealth of nations*. Westport, CT: Praeger.

Neisser, U., Boodoo, G., Bouchard, T.J. Jr., Boykin, A.W., Brody, N., Ceci, S.J., Halpern, D., Loehlin, J.C., Perloff, R., Sternberg, R.J. and Urbina, S. (1996) Intelligence: Knowns and unknowns. *American Psychologist*, **15**, 77–101.

Nell, V. (2000) *Cross-cultural neuropsychological assessment: Theory and practice*. London, UK: Erlbaum.

Ones, D.S. and Viswesvaran, C. (2002) Introduction to the special issue: Role of general mental ability in industrial, work, and organizational psychology. *Human Performance*, **15**, 1–2.

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. (2000) The Raven's Progressive Matrices: Change and stability over culture and time. *Cognitive Psychology*, **41**, 1–48.

Raven, J. *et al*. (1990) *Manual for Raven's Progressive Matrices and Vocabulary Scales. Research Supplement No. 3: American and International Norms* (2nd Edn.). Oxford, England: Oxford Psychologists Press.

Raven, J., Raven, J.C. and Court, J.H. (1998) *Manual for Raven's Advanced Progressive Matrices* (1998 Edn.). Oxford, England: Oxford Psychologists Press.

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

Roth, P.L., Bevier, C.A., Bobko, P., Switzer, F.S. III and Tyler, P. (2001) Ethnic group differences in cognitive ability in employment and educational settings: A meta-analysis. *Personnel Psychology*, **54**, 297–330.

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: A life history perspective* (3rd Edn.). 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. (2002) Jensen Effects and African/Colored/Indian/White differences on Raven's Standard Progressive Matrices in South Africa. *Personality and Individual Differences*, **33**, 65–70.

Rushton, J.P. and Jensen, A.R. (2003) African–White IQ differences from Zimbabwe on the Wechsler Intelligence Scale for Children-Revised are mainly on the *g* factor. *Personality and Individual Differences*, **34**, 177–183.

Rushton, J.P. and Skuy, M. (2000) Performance on Raven's Matrices by African and White university students in South Africa. *Intelligence*, **28**, 251–265.

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.

Schmidt, F.L. and Hunter, J.E. (1998) The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychological Bulletin*, **124**, 262–274.

Skuy, M., Gewer, A., Osrin, Y., Khunou, D., Fridjhon, P. and 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., Schutte, E., Fridjhon, P. and 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.

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.

Sternberg, R.J., Grigorenko, E.L., Ngrosho, D., Tantufuye, E., Mbise, A., Nokes, C., Jukes, M. and Bundy, D.A. (2002) Assessing intellectual potential in rural Tanzanian school children. *Intelligence*, **30**, 141–162.

te Nijenhuis, J. and 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., Evers, A. and Mur, J.P. (2000) Validity of the Differential Aptitude Test for the assessment of immigrant children. *Educational Psychology*, **20**, 99–115.

te Nijenhuis, J., Tolboom, E., Resing, W. and Bleichrodt, N. (in press) Does cultural background influence the intellectual performance of children from immigrant groups? Validity of the RAKIT intelligence test for immigrant children. *European Journal of Psychological Assessment*, **20**, 10–26.

Van de Vijver, F. (1997) Meta-analysis of cross-cultural comparisons of cognitive test performance. *Journal of Cross-Cultural Psychology*, **28**, 678–709.

Vernon, P.A. (1983) Speed of information-processing and general intelligence. *Intelligence*, **7**, 53–70.

Viswesvaran, C. and Ones, D.S. (2002) Agreements and disagreements on the role of general mental ability (GMA) in industrial, work, and organizational psychology. *Human Performance*, **15**, 211–231.

Wigdor, A.K. and Garner, W.R. (Eds.) (1982) *Ability testing: Uses, consequences, and controversies, Part 1. Report of the Committee, Part 2: Documentation section*. Washington, DC: National Academy Press.

Zaaiman, H., van der Flier, H. and 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.