Intelligence xxx (2009) xxx–xxx

Contents lists available at ScienceDirect

Intelligence

A systematic literature review of the average IQ of sub-Saharan Africans ☆

Jelte M. Wicherts *, Conor V. Dolan, Han L.J. van der Maas

Department of Psychology, Psychological Methods, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands

ARTICLE INFO

Article history:
Received 8 October 2008
Received in revised form 6 May 2009
Accepted 12 May 2009
Available online xxxx

Keywords:
Group differences
Black–White differences
Flynn Effect
Race differences
Cross-cultural comparison
National IQ

ABSTRACT

On the basis of several reviews of the literature, Lynn [Lynn, R., (2006). Race differences in intelligence: An evolutionary analysis. Augusta, GA: Washington Summit Publishers.] and Lynn and Vanhanen [Lynn, R., & Vanhanen, T., (2006). IQ and global inequality. Augusta, GA: Washington Summit Publishers.] concluded that the average IQ of the Black population of sub-Saharan Africa lies below 70. In this paper, the authors systematically review published empirical data on the performance of Africans on the following IQ tests: Draw-A-Man (DAM) test, Kaufman-Assessment Battery for Children (K-ABC), the Wechsler scales (WAIS & WISC), and several other IQ tests (but not the Raven's tests). Inclusion and exclusion criteria are explicitly discussed. Results show that average IQ of Africans on these tests is approximately 82 when compared to UK norms. We provide estimates of the average IQ per country and estimates on the basis of alternative inclusion criteria. Our estimate of average IQ converges with the finding that national IQs of sub-Saharan African countries as predicted from several international studies of student achievement are around 82. It is suggested that this estimate should be considered in light of the Flynn Effect. It is concluded that more psychometric studies are needed to address the issue of measurement bias of western IQ tests for Africans.

Contents

☆ The preparation of this article was supported by VENI grant no. 451-07-016 from the Netherlands Organization for Scientific Research (NWO) awarded to JMW. We thank three anonymous reviewers for their insightful comments.
* Corresponding author. Tel.: +31 205257067; fax: +31 206390026.
E-mail address: J.M.Wicherts@uva.nl (J.M. Wicherts).

0160-2896/$ – see front matter © 2009 Published by Elsevier Inc.
doi:10.1016/j.intell.2009.05.002

Lynn and Vanhanen (2002, 2006) collated the results of a large number of published studies in which western IQ tests were administered to samples from countries all over the globe. They compared the mean IQs in these samples to UK norms in order to arrive at estimates of the average IQ of nations' inhabitants, which they denoted national IQs. Given several samples from a single country, they computed the median of the means to establish the national IQ. On the basis of their extensive literature reviews, Lynn (and Vanhanen)[1] concluded that the average IQ of the Black population of sub-Saharan Africa (henceforth Africans) lies below 70 (Lynn, 2003, 2006; Lynn & Vanhanen, 2002, 2006).

Lynn and Vanhanen's estimate of the average IQ of Africans is accorded a central role in the discussion on Black–White differences in IQ by Rushton and Jensen (2005). This estimate features prominently in several evolutionary theories of intelligence (Kanazawa, 2004; Lynn, 2006; Rushton, 2000). Moreover, Lynn and Vanhanen's (2002, 2006) estimates of national IQ have featured in over twenty scientific studies (Barber, 2005; Dickerson, 2006; Gelade, 2008a,b; Jones & Schneider, 2006; Kanazawa, 2006, 2008; Kirkcaldy, Furnham, & Siefen, 2004; Lynn, Harvey, & Nyborg, 2009; Meisenberg, 2004; Morse, 2006; Ram, 2007; Rindermann, 2006, 2007, 2008a,b; Rindermann & Meisenberg, in press; Rushton & Templer, in press; Shatz, 2008; Templer, 2008; Templer & Arikawa, 2006a,b; Voracek, 2004; Weede & Kampf, 2002; Whetzel & McDaniel, 2006; Woodley, 2009).

Although these estimates of national IQ are claimed to be “highly valid” (Rushton, 2003, p. 368) or “credible” (McDaniel, 2008, p. 732) by some authors, the work by Lynn (and Vanhanen) has also drawn criticism (Barnett & Williams, 2004; Ervik, 2003; Hunt & Carlson, 2007; Hunt & Sternberg, 2006; Lane, 1994). One point of critique is that Lynn (and Vanhanen)'s estimate of average IQ among Africans is primarily based on convenience samples, and not on samples carefully selected to be representative of a given, targeted, population (Barnett & Williams, 2004; Hunt & Sternberg, 2006). Unfortunately, in many developing countries, such representative samples are lacking (McDaniel, 2008).

A literature review is necessarily selective. Despite Lynn's objective of providing a “fully comprehensive review of the evidence” (Lynn, 2006, p. 2), a sizeable portion of the relevant literature was not considered in both his own review, and in reviews with Vanhanen. Nowhere in their reviews did Lynn (and Vanhanen) specify the details of their literature search. Our own searches in library databases resulted in additional relevant studies that may be used to estimate national IQ. For instance, Lynn and Vanhanen (2006) accorded a national IQ of 69 to Nigeria on the basis of three samples (Fahrmeier, 1975; Ferron, 1965; Wober, 1969), but they did not consider other relevant published studies that indicated that average IQ in Nigeria is considerably higher than 70 (Maqsud, 1980a,b; Nenty & Dinero, 1981; Okunrotifa, 1976). As Lynn rightly remarked during the 2006 conference of the International Society for Intelligence Research (ISIR), performing a literature review involves making a lot of choices. Nonetheless, an important drawback of Lynn (and Vanhanen)'s reviews of the literature is that they are *unsystematic*. Unsystematic literature reviews do not adhere to systematic methodology to control for potential biases in the many choices made by the reviewer (Cooper, 1998; Light & Pillemer, 1984). Lynn (and Vanhanen) failed to explicate the inclusion and exclusion criteria they employed in their choice of studies. Such criteria act as a filter, and may thus affect the estimate of national IQ. Lynn (and Vanhanen) excluded data from several sources without providing a rationale. For instance, they used IQ data from Ferron (1965), who provided averages in seven samples of children from Sierra Leone and Nigeria on a little-known IQ test called the Leone. For reasons not given, Lynn (2006) and Lynn and Vanhanen (2006) only used data from the two lowest scoring samples from Nigeria. Most of the remaining samples show higher scores, but those samples were not included in the estimation of the national IQ of Nigeria and Sierra Leone. Likewise, Lynn (and Vanhanen) did not consider several relatively high-scoring African samples from South Africa (Crawford Nutt, 1976; Pons, 1974). It is unfortunate that Lynn (and Vanhanen) did not discuss their exclusion criteria. In some cases (Crawford Nutt, 1976; Pons, 1974), the Raven's Progressive Matrices was administered with additional instruction. Although this instruction is quite similar to an instruction as described in the test manual (Raven, Court, & Raven, 1996), some have argued that this instruction artificially enhances test performance (cf. Rushton & Skuy, 2000). Given the likely differences in opinion on which samples to include or exclude in a review, inclusion and exclusion criteria should be explicated clearly and employed consistently. It is well known that unsystematic literature reviews may lead to biased results (Cooper, 1998; Light & Pillemer, 1984). Another problem is that the computation of statistics in literature reviews is quite error-prone. Indeed Lynn's work contains several errors (Loehlin, 2007).

The aim of our current paper is to estimate the current level of average IQ of Africans in terms of western norms by a thorough systematic literature review. We focus on the literature concerning IQ test performance of Africans on the Goodenough–Harris Draw-a-Man test (DAM; Goodenough, 1926; Harris, 1963), Kaufman-Assessment Battery for Children (Kaufman & Kaufman, 1983), the Wechsler scales

[1] Whenever we refer to “Lynn (and Vanhanen)” in this paper, we refer to Lynn and Vanhanen (2002, 2006) and to Lynn (2006), because the literature reviews in these three books overlap strongly. We refer to specific books whenever necessary.

(Wechsler, 1974, 1981), and several other IQ tests. Wicherts, Dolan, Carlson, and van der Maas (2009) considered all the available literature on the test performance of Africans on Raven's Coloured Progressive Matrices (CPM; Raven, 1956) and the Standard Progressive Matrices (SPM; Raven, 1960), and established that the average IQ on these tests is appreciably higher than Lynn and Vanhanen's estimates of national IQ of sub-Saharan African countries.

It is important to note that an observed IQ score does *not* necessarily equal a particular level of general intelligence or *g* (Bartholomew, 2004), as it is necessary to consider the issue of validity in interpreting an observed score as an indication of the position on a latent variable such as *g*. Several authors have questioned whether the IQ scores of Africans are valid and comparable to scores in western samples in terms of *g* (Barnett & Williams, 2004; Ervik, 2003; Hunt & Carlson, 2007; Hunt & Sternberg, 2006; Lane, 1994). Some (e.g., Berry, 1974) reject the very possibility of obtaining a valid measure of *g* in Africa with western IQ tests, while others (e.g., Herrnstein & Murray, 1994; Lynn, 2006; Rushton & Jensen, 2005) consider it relatively unproblematic. The psychometric issue of measurement invariance (Mellenbergh, 1989; Millsap & Everson, 1993) is crucial to the comparability of test scores across cultural groups in terms of latent variables, such as *g*. Alas, the number of studies addressing measurement invariance is small. In our review, we specifically focus on the issue of measurement invariance, and use the results of measurement invariance studies to determine whether a study should be included in the estimate of the average IQ of Africans. Before we provide the results of our review, we address the studies that are thought to support the accuracy of Lynn and Vanhanen's estimates of national IQ.

1. Scholastic achievement surveys

In several studies (Lynn, 2006; Lynn, Meisenberg, Mikk, & Williams, 2007; Lynn & Mikk, 2007; Lynn & Vanhanen, 2006; Rindermann, 2007), Lynn and Vanhanen's estimates of national IQ were correlated with data from several internationally comparable surveys of school achievement (e.g., TIMMS, PISA). In these surveys representative samples of primary and secondary students were given reading, mathematics, or science tests. These studies have quite clearly shown that national IQs and the means for countries from these scholastic surveys are highly correlated (cf. Hunt & Wittmann, 2008; Lynn & Mikk, 2009). According to Lynn and co-workers, these studies validate the national IQs (Lynn, 2006; Lynn et al., 2007; Lynn & Mikk, 2007; Lynn & Vanhanen, 2006).

In this section we consider specifically the mean performance of samples from sub-Saharan Africa as established in these achievement surveys. Because our review on the Raven's tests showed that the average IQ of Africans on these tests was markedly higher than 70, we expected that the Lynn (and Vanhanen) estimates of national IQ of sub-Saharan African countries were too low, and that the average performance of Africans in the achievement surveys would give rise to predicted national IQs in these countries that were higher than the estimates of Lynn (and Vanhanen). To test this hypothesis, we used a straightforward methodology in which we used linear regression to predict the national IQs in all countries *except* those in sub-Saharan Africa. Subsequently, we used the obtained regression equation to estimate the national IQ of those sub-Saharan African countries that participated in each particular achievement survey. We consider the four published studies separately.

Lynn (2006) and Lynn and Vanhanen (2006) report on a study in which they used a combination of international scholastic achievement surveys of Hanushek and Kimko (2000), in which the average IQs of Nigeria (average IQ of 69 according to Lynn & Vanhanen, 2006), Swaziland (average IQ of 68 according to Lynn & Vanhanen, 2006), and Mozambique (average IQ of 64 according to Lynn & Vanhanen, 2006) appear alongside that of 36 countries outside of sub-Saharan Africa. In Fig. 1, we display the results of his study. Lynn (2006) and Lynn and Vanhanen (2006) reported a correlation of 0.81 between these two variables, and claimed that this result validated their estimates of national IQs. However, a look at the scatter plot shows clearly the presence of three outliers, which are the three data points on the low-left side. In the absence of these three data points, the correlation is 0.86. These outliers correspond to the three countries from sub-Saharan Africa. Table 1 gives the estimated national IQs of these countries as well as the standardized residuals for these countries, which are all significantly different from zero ($p<.001$). The national IQs of the three African countries as predicted from the data of Hanushek and Kimko (2000) are well-above 80.

In Fig. 2, we display part of the results from the study by Rindermann (2007), who correlated Lynn and Vanhanen's estimates of national IQ with a student assessment mean score for 76 countries. This mean student assessment score was computed on the basis of the means of countries in the following surveys (cf. Rindermann, 2007): International Association for the Evaluation of Educational Achievement (IEA) Reading-Study of 1991, International Assessment of

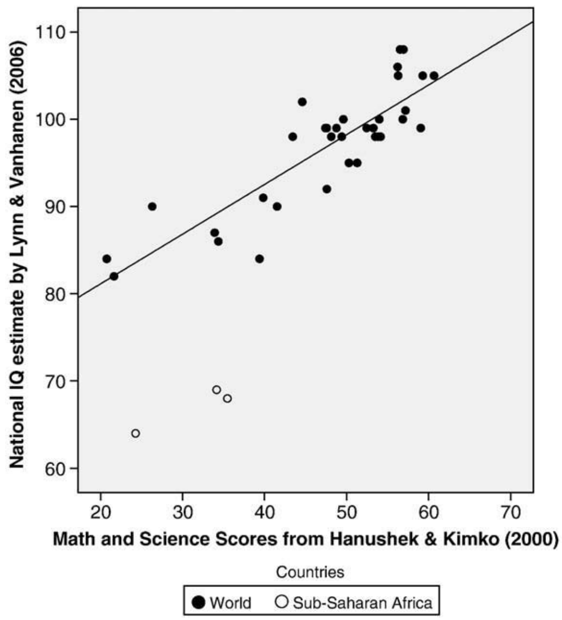

Fig. 1. Scatterplot of data from study by Lynn (2006) and Lynn and Vanhanen (2006).

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

Table 1
Comparison of national IQ estimated by Lynn and Vanhanen (2006) and those predicted by Math and Science studies.

Study	Dataset	Country	National IQ	Predicted national IQ	Standardized residual (*Z*)
Lynn and Vanhanen (2006)	Hanushek and Kimko (2000)	Mozambique	64	83.9	−5.64**
	Hanushek and Kimko (2000)	Nigeria	69	89.5	−5.80**
	Hanushek and Kimko (2000)	Swaziland	68[a]	90.2	−6.30**
Lynn et al. (2007)	TIMMS 1995	South Africa	72	80.5	−2.30*
	TIMMS 1999	South Africa	72	72.7	−.20
Lynn and Mikk (2007)	TIMMS 2003	Botswana	70[a]	80.5	−3.26*
	TIMMS 2003	Ghana	71	68.8	0.69
	TIMMS 2003	South Africa	72	67.4	1.42
Rindermann (2007)	Various	Botswana	70[a]	81.8	−3.35**
	Various	Ghana	71	71.8	−.22
	Various	Nigeria	69	87.1	−5.14**
	Various	South Africa	72	74.4	−.70
	Various	Zimbabwe	66	84.6	−5.27**

Note: *$p<.05$; **$p<.001$.
[a] National IQs estimated by Lynn and Vanhanen on the basis of neighboring countries.

Educational Progress (IAEP)-II 1991, TIMSS 1995, TIMSS 1999, TIMSS 2003, Programme for International Student Assessment (PISA) 2000, PISA 2003, Progress in International Reading Literacy Study (PIRLS) 2001. Note that some of these studies are also included in the studies we discuss below. Rindermann's analyses included five countries in sub-Saharan Africa, the national IQs of which are all below the regression line in Fig. 2. As can be seen in Table 1, Nigeria, Zimbabwe, and Botswana show significant negative residuals in the prediction of national IQ from these student assessment means, although for Ghana and South Africa the residuals are non-significant.

Lynn and Mikk (2007) report on the correlation between the means of 46 countries in the TIMMS 2003 and national IQ. The data for the Math scores in the 8th grade are depicted in Fig. 3. In these data, the residual of Botswana is again highly significant. The estimated national IQ of this country turns out to be 80.5. However, the data from South Africa and Ghana are no outliers in this analysis, which is in line with the results on the basis of Rindermann's data.

Lynn et al. (2007) report on the same TIMMS 2003 data as do Lynn and Mikk (2007), but added the results from the TIMMS 1995 and TIMMS 1999 surveys, in which South Africa again was the only sub-Saharan African country. The results for the TIMMS 1995 are depicted in Fig. 4, where it can be seen that South Africa is an outlier. In the prediction of national IQ from the TIMMS 1999 data, the South African mean is not a significant outlier, although the residual is again negative.

To conclude, we have used the results from four published studies on the validity of national IQ to estimate the national IQs of sub-Saharan African countries on the basis of their means in large-scale student assessment surveys. The average predicted national IQs for Botswana, Ghana, Mozambique, Nigeria, South Africa, Swaziland, and Zimbabwe are 81.4, 70.3, 83.9, 88.3, 73.8, 90.2, and 84.6, respectively. On average, the estimated national IQs of these sub-Saharan African countries

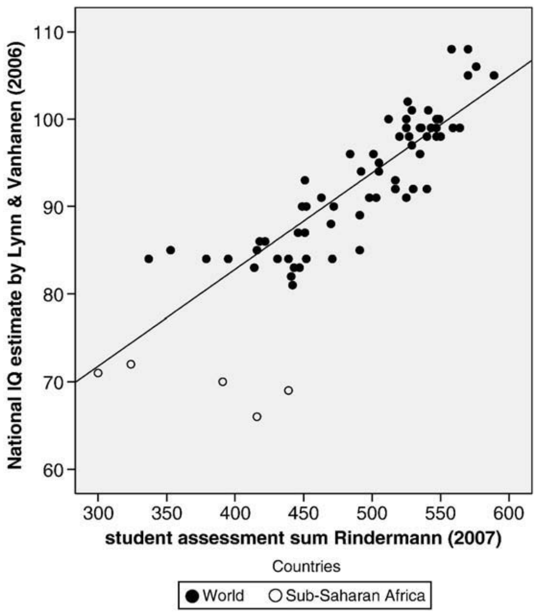

Fig. 2. Scatterplot of data from study by Rindermann (2007).

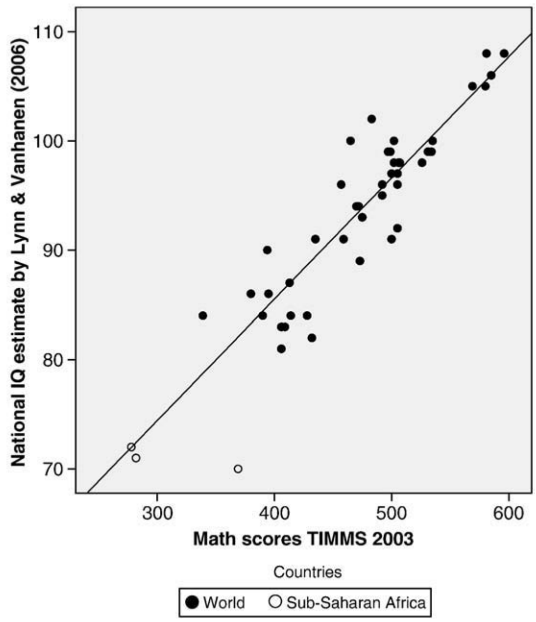

Fig. 3. Scatterplot of data from study by Lynn and Mikk (2007).

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

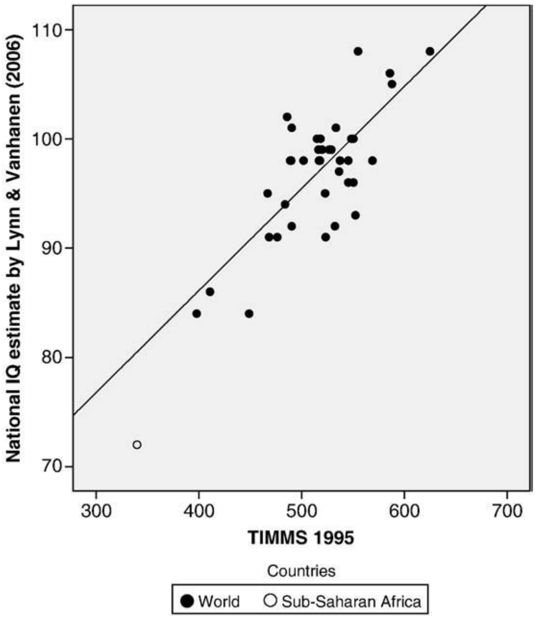

Fig. 4. Scatterplot of data from study by Lynn et al. (2007).

equals 81.8, which is higher than Lynn and Vanhanen's estimates of national IQ (i.e., $M = 68.6$). Thus, although these four studies appear to validate national IQs in other parts of the world, they do not appear to support the national IQs in sub-Saharan Africa.

2. A systematic review of the literature

In this section, we present the results of a systematic review of published studies in which western IQ tests (i.e., IQ tests developed and standardized in western countries) were administered to Black sub-Saharan Africans. Lacking of more specific data our emphasis is necessarily on overall (average) performance. Given the available data, we cannot do justice to the wide cultural and economic differences of the many and varied peoples on this continent. Nonetheless, we hope that our review may help to improve estimates of national IQ of African countries in future studies. In this review, we consider the studies that we found in a careful literature search. In view of the principle of "the totality of the available evidence" (Gottfredson, 2005), we discuss all the data sets used by Lynn (and Vanhanen) to support their claim that average IQ in Africa lies below 70. We also report on all samples that did not meet our inclusion criteria, thereby allowing the interested reader to compute alternative estimates of average IQ on the basis of his or her preferred inclusion criteria. We discussed the evidence on the Raven's tests elsewhere. Also, we do not consider samples of college and university students, as these are selected samples, and are discussed separately by Lynn (2006).

3. Method

3.1. Search of studies

In early 2009, we employed PsycINFO to search for additional studies of IQ test performance of Africans. We used the following search terms: "IQ", "intelligence", "cognitive ability", "abilities" combined with the countries' names and adjectives as well as the words "Africa", "African". This resulted in approximately 2800 unique references that were scanned for relevance. Because the Draw-a-Man test is administered often in Africa, we also searched for additional data of this test by using the search term "draw" in PsycINFO and by scanning in Web of Science the approximately 400 papers that referred to both test manuals (Goodenough, 1926; Harris, 1963). The titles and authors' names of all these papers were scanned for relevance. We used only books, papers, or reports that were available through the IBL system in the Netherlands, a system to which 400 Dutch libraries are connected. Although our search strategy resulted in a large number of studies, it is conceivable that we missed studies.

3.2. Our inclusion criteria

Our goal was to estimate the average IQ of samples of normal and healthy Africans on the basis of western norms on western IQ tests. Because of a scarcity of carefully selected representative samples on the IQ tests under review, we focus here on convenience samples. We considered the following inclusion criteria to arrive at an estimate of average IQ of Africans. Because readers may disagree with our inclusion criteria, we also provide alternative estimates on the basis of more lenient inclusion criteria.

3.2.1. Norms

For our preferred estimate of mean IQ, we used only test scores for which western standardized IQ norms were available. In several samples, the outdated concept of mental-age IQ (i.e., the mental-age times 100, divided by the chronological age) was used (e.g., Badri, 1965a; Fick, 1929). The reason is that these mental-age IQs often do not have the same SD as standardized IQs, which may result in inaccurate estimates of average IQ of samples that show a lower mean than the mean in the standardization sample (e.g., Jensen, 1980). Nonetheless, for some samples that took mental-age IQ tests it was possible to re-compute standardized IQs on the basis of standardization samples. Because of this inclusion criterion, we also did not take into account, in our estimate of average IQ, data from some samples in which data from western convenience samples were used to compute the average IQ of African samples. In several instances, Lynn and Vanhanen (2002, 2006) computed the mean IQ of an African sample by comparing the mean and SD of a particular IQ (sub)test in the African sample to the mean and SD of a western convenience sample. This method is problematic because it is based on strong and possibly arbitrary assumptions regarding the average IQ and variance in the western convenience sample. If these assumptions do not hold, this may bias the estimate of average IQ in the African sample. For instance, suppose that the difference between a sample of European university students and a sample of African students is characterized by an effect size of .80. The average IQ of the African sample could differ from 88 to 112, depending on (debatable) assumptions regarding the European sample's mean and SD (i.e., $M = 120$, $SD = 10$ vs. $M = 100$, $SD = 15$). Therefore, we only consider in our estimate of average IQ of Africans those IQ scores that are based on western standardization samples.

3.2.2. Standardized test administration of entire IQ test

The test should have been administered in accordance with the guidelines in the test manual. Deviations from the guidelines may act to enhance or lower the scores. For instance, the alteration of subtests or items may have an effect on the average IQ. In addition, we only used test scores of IQ tests that were administered in their entirety. The reason was that we were interested in IQ as an indicator of *g*, not in IQ scores based on single subtests or a collection of subtests that were aimed at measuring, say, spatial abilities or verbal abilities. We also did not consider scores on tests that are not meant to measure *g*, such as the Wisconsin Card Sorting Test (WCST), because this test is not an IQ test. However, because Lynn (and Vanhanen) did consider WCST scores in some African samples, we do discuss these samples below.

3.2.3. No reported problems during testing

Test administration should not have been described as problematic by the original authors. The idea behind this criterion was that the authors of the original study, from which the data were drawn, are in the best position to judge the IQ test's suitability for the African test takers, and to evaluate any problems that may have lowered test performance.

3.2.4. No measurement bias

We did not include in our estimate of average IQ of Africans data from tests that were found or expected (by the original authors) to be biased with respect to Africans. By measurement bias studies we mean studies that employed contemporary methods from item response theory to study differential item functioning, or measurement invariance studies that employed multi-group confirmatory factor analysis. The idea behind this criterion is that the use of national IQs and the comparison of IQ scores across groups is normally concerned with studying the cognitive abilities that the tests at hand are supposed to measure. If a test is found to be biased with respect to Africans, this suggests that their test scores cannot be interpreted in terms of the latent abilities underlying the test scores. Note that the number of studies of measurement invariance of Africans' test performance is small, so it is often impossible to ascertain whether IQ test scores of Africans can be interpreted in terms of these latent variables. Several of the original authors argued that the tests were invalid or biased with respect to Africans without reference to formal analyses (which are not always feasible given the samples sizes in many studies).

3.2.5. Normal samples

We excluded from our preferred estimate of average IQ of Africans' data from unhealthy or special populations. For instance, we excluded the WISC-R scores from a sample of deaf children from Nigerian (Alade, 1992) and we excluded samples of test takers that were specifically sampled because of their poor health.

It is quite possible that some readers may disagree with our criteria and the choices we made on the basis of these criteria. It will become clear that in several cases, our inclusion criteria differ from those which Lynn (and Vanhanen) apparently used. We strongly feel that there should be an open discussion of which studies to include in the estimates of national IQ and welcome the interested reader to compute his or her own average IQ of Africans on the basis of his or her own inclusion criteria. To that end, we report on all samples we found in our literature search and include our inclusion criteria as predictors of the mean IQs in a meta-analytic regression model.

3.3. Statistical analyses

The statistical analyses include all samples that we found in our literature search (i.e., an inclusionist strategy; Kraemer, Gardner, Brooks, & Yesavage, 1998). A meta-analytic regression model will be employed to evaluate the effects of our inclusion criteria. Besides the mean IQ, we also consider the SD of IQ in the African samples. We were able to compute the SD of IQ in 29 of the 44 samples in our review. The standard error (*SE*) of the standard deviation (Ferguson, 1966) for sample *i* equals: $SE_{SD_i} = \frac{SD_i}{\sqrt{2 * N_i}}$, where SD_i and N_i are the sample SD and sample size, respectively. We used the inverse of the square of this standard error (i.e., the inverse of the sampling variance) as weight in a random-effects model. So we assume that there is not one single population SD of IQ in African samples, and estimate the mean and variance of the population SDs. The standard error of the mean in each sample *i* was computed according to the usual formula $SE_{Mi} = \frac{SD_i}{\sqrt{N_i}}$. In the meta-analyses of the mean IQ, the means in the samples were weighted by the inverse of the square of these SE_{Mi}s: $W_i = \frac{N_i}{SD_i^2}$. Where the sample SD was not reported, the SE_{Mi} was based on the actual sample size and the SD as estimated from the meta-analytic estimate of the mean of population SDs.[2] The meta-analytical model we used was a random-effects model (i.e., we assume that there is heterogeneity in population mean IQ), and we used as moderators of the means indicators of our five inclusion criteria, as well as several other variables of interest. In addition, we studied the possibility of publication bias by using the trim and fill method (Duval & Tweedie, 2000) and two regression-based methods (Egger, Davey Smith, Schneider, & Minder, 1997; Sterne & Egger, 2005). Analyses were conducted using David B. Wilson's SPSS macros with maximum-likelihood estimation in SPSS 16 for MAC and the Comprehensive Meta-Analysis package.

4. Results

All samples are reported in Table 2, along with our assessments of whether or not these samples meet our inclusion criteria. We now discuss all samples per test, starting with those most commonly used in Africa.

4.1. Draw-a-Man test

The Goodenough–Harris Draw-a-Man test (DAM; Goodenough, 1926; Harris, 1963) is a non-verbal intelligence test for children aged two to thirteen in which children are required to draw a man. Although the scores on the Draw-a-

[2] We checked the robustness of our main results against the use of alternative estimates of the SD for those samples with missing SDs. The use of an SD of 15 had very little effect on our results. For instance, it raised the overall mean IQ by only .02 points, while it changed the mean IQ on the basis of the samples that meet our inclusion criteria by less than .01 points.

Table 2
Sub-Saharan African samples and IQ scores as discussed in text.

Source	Country	N	Sample	Stand. IQ norm	Stand. admin.	No probl.	No bias	Normal sample	M age	Test	MIQ	SD	IQUKFE	IQ Lynn
(Akande, 2000)	South Africa	63	Normally-achieving, low-achieving, and learning disabled school children	+/−	+	+	+	−	8.0	WPPSI WCST	81.0	–	72.9	–
Ani and Grantham-McGregor (1998)	Nigeria	94	Aggressive and pro-social school boys from Lagos	+	−	+	+	+/−	11.7	Similar. WISC-R	81.8	14.3	73.2	–
Ashem and Janes (1978)	Nigeria	128	Urban and rural normally and undernourished children	+	−	+	+	+/−	4.3	McCart.	92.6	19.5	88.8	–
Avenant (1988)	South Africa	140	Prison wardens	+	−	+	−	+	25.0	WAIS-R	76.0	8.5	72.5	71*
Badri (1965a)	Sudan	293	Boys from rural and urban areas	−	+	+	−	+	9.0	DAM	85.9	14.3	73.1	74
Badri (1965b)	Sudan	80	Culturally deprived preschool boys	−	+	−	−	+	6.0	DAM	75.0	–	61.9	64
Bakare (1972)	Nigeria	393	Upper-class and lower-class children	+	+	+	+	+	8.0	DAM	87.2	12.0	83.1	–
Bardet et al. (1960)	Senegal	559	School children in Dakar and Khombole	+	+	+	+	+	11.0	DAM	77.9	–	66.3	–
Boivin and Giordani (1993)	D.R. Congo	97	Children with intestinal parasites	+	−	+	−	−	8.6	K-ABC	71.0	9.1	66.6	62
Boivin et al. (1995)	D.R. Congo	97	Children with intestinal parasites	+	−	+	−	−	7.6	K-ABC	69.9	10.4	64.9	68
Boivin (2002)	Senegal	58	Children with an history of malaria	+	−	+	−	−	9.8	K-ABC	82.3	15.6	75.2	–
Buj (1981)	Ghana	225	Representative sample of urban adults	+	+	+	+	+	≈30	CFT	82.2	–	75.7	80
Claassen et al. (2001)	South Africa	196	Representative sample of English-speaking adults	+	−	+	−	+	30.0	WAIS-III	86.4	13.3	83.8	–
Dent (1937)	South Africa	80	Schooled and unschooled children	−	+	−	+	+	12.0	Koh Bl.	67.0	11.0	68.0	68*
Dunstan (1961, cited in Ferron, 1965)	Sierra Leone	400	Form I students from Freetown	−	+	−	−	+	13.5	M.H.T.	77.8	–	77.8	–
Fahmy (1964)	Sudan	184	Children from primitive tribe	−	+	+	+	+	11.0	Alex. Pass	94.4	≈20	84.3	69
Fahrmeier (1975)	Nigeria	152	Schooled and unschooled children	+	−	+	+	+	10.6	PMA spat.	77.6	–	72.3	–
Ferron (1965)	Nigeria	1633	Primary school children	−	+	+	−	+	≈11	Leone	77.0	–	77.0	–
Fick (1929)	South Africa	293	Sample of children	−	+	+	−	+	11	Army-B	64.8	–	65.0	65
Holding et al. (2004)	Kenya	174	Rural children suffering from malaria	−	−	−	−	−	6.0	K-ABC	63.0		63.0	63
Hunkin (1950)	South Africa	1076	Fairly representative sample of urban school children	+	+	+	−	+	6.8	DAM	83.0	16.2	74.7	70*
Kashala et al. (2005)	D.R. Congo	183	Schoolchildren with ADHD and normal controls	+	−	+	+	+/−	8.4	DSP	93.5	12.5	82.5	–
Khaleefa et al. (2008)	Sudan	1345	Representative sample of school children in 1964	−	+	−	−	+	7	DAM	83.5	–	70.1	–
Khaleefa et al. (2008)	Sudan	1345	Representative sample of school children in 2006	−	+	+	+	+	7	DAM	95.6	10.1	69.6	–
Klein et al. (2007)	Various	28	Adults immigrated to Belgium	+	+	+	+	+	32	CFT	83.1	9.1	69.6	
Lloyd and Pidgeon (1961)	South Africa	275	Children in urban and rural school	+	+	+	+	+	11.5	NVT	88.7	7.8	84.3	74*
Lynn and Owen (1994)	South Africa	1093	Secondary school children	−	+	+	−	+	16.5	JAT	68.0	13.4	68.0	68*
Minde and Kantor (1976)	Uganda	514	Children in three primary schools	+	+	+	+	+	11.0	DAM	88.6	17.1	83.9	–
Nell (2000)	South Africa	152	Employed men with fewer than 12 years of education	+	−	+	−	+	43.4	3 subtests WAIS-R	79.0	9.0	71.9	69*
Nenty and Dinero (1981)	Nigeria	803	Secondary school students	−	−	+	−	+	13.5	CFT3	97.2	13.5	95.2	–

(continued on next page)

Table 2 (*continued*)

Source	Country	*N*	Sample	Stand. IQ norm	Stand. admin.	No probl.	No bias	Normal sample	*M* age	Test	MIQ	SD	IQUKFE	IQ Lynn
Nissen et al. (1935)	Guinea	50	Unschooled children	−	+	−	−	+	10.0	Army-B	63.0	11.0	63.0	63
Nwanze and Okeowo (1980)	Nigeria	13	Children with reading problems	+	+	+	+	−	7.3	DAM	98.6	5.3	92.1	–
Ohuche and Ohuche (1973)	Sierra Leone	202	Children in experimental school	+	+	+	+	+	8.5	DAM	94.8	–	91.3	–
Richter et al. (1989)	South Africa	415	Urban school children	+	+	+	−	+	9.0	DAM	84.0	–	74.5	77*
Shuttleworth Edwards et al. (2004)	South Africa	40	Educated adults	+	+	+	+	+	24.0	WAIS-III	94.2	16.2	90.7	–
Skuy et al. (2000)	South Africa	21	Children with learning difficulties	+	+	+	+	+/−	9.5	K-ABC	102.9	10.1	96.4	–
Skuy et al. (2001)	South Africa	100	Soweto secondary school children	+	+/−	+	+	+	17.0	DAM& WISC-R	82.9	16.0	71.6	–
Skuy et al. (2001)	South Africa	152	Secondary school children	+	+	+	+	+	14.0	WISC-R &WCST	76.0	12.5	74.3	67*
Sternberg et al. (2001)	Kenya	85	Children from rural area	+	−	+	+	+	13.0	Mill–Hill	64.0	–	64.0	64
Sternberg et al. (2002)	Tanzania	458	Rural school children	−	−	+	−	+	12.0	WCST	72.7		72.0	72
Vernon (1969)	Uganda	50	Boys of above average SES	−	+	+	−	+	12.0	Various	88.0	11.7	86.0	80*
Wilson et al. (1991)	Zimbabwe	52	Primary school girls	+	+	+	+	+	9.0	PMA	96.0	–	86.2	–
Yoloye (1971)	Nigeria	516	Secondary school boys in Idaban	+	−	+	+	+	15.8	Lorge–Th.	88.0	12.8	86.0	–
Zindi (1994a)	Zimbabwe	204	High school children	+	+	+	−	+	13.0	WISC-R	79.0	11.9	71.6	61

Note: IQ: IQ estimate on the basis of US or UK norms; *Considered by Lynn (2006), but not considered by Lynn and Vanhanen (2006) in their estimates of national IQ; IQFEUK: IQ estimate in terms of UK norms with correction for Flynn Effect; IQ Lynn: IQ estimate by Lynn (2006); Army-B: Army Beta test; BD: Block Design; CFT: Culture Fair Intelligence Test; DAM: Draw-a-Man test; DSP: Digit Span subtest from WISC-R; JAT: Junior Aptitude Test; K-ABC: Kaufman-Ability Battery for Children; KB: Koh's Blocks; Lorge–Th: Lorge–Thorndike intelligence test; M.H.T.: Moray House Test no. 44; McCart: McCarthy Scale of Children's Abilities; Mill–Hill: Mill–Hill Vocabulary; PMA: Primary Mental Abilities; WAIS: Wechsler Adult Intelligence Scale; WAIS-III: Wechsler Adult Intelligence Scale-3rd edition; WCST: Wisconsin Card Sorting Test; WISC-R: Wechsler Intelligence Scale for Children-Revised; WPPSI: Wechsler Preschool and Primary Scale of Intelligence.

Man test have been shown to correlate reasonably well with scores on cognitive ability tests such as the Stanford–Binet (e.g., Williams, 1935) and the SPM (Carlson, 1970), the DAM test is not generally considered as good an indicator of general intelligence as regular IQ tests or tests like the SPM or CPM (e.g., Jensen, 1980). Nonetheless, we accept the DAM as an IQ test, which is often administered in Africa. Lynn (2006) used DAM test scores to estimate the average IQ of Africans in five African samples (Badri, 1965a,b; Fahmy, 1964; Hunkin, 1950; Richter, Griesel, & Wortley, 1989). Lynn and Vanhanen (2006) used three of these samples (Badri, 1965a,b; Fahmy, 1964) to estimate national IQ in Sudan.

Nonetheless, in several of these African samples the administration of the DAM was fraught with difficulties. The first of these is Fahmy's (1964) study among schooled and unschooled children in a primitive tribe in Sudan. These children had an average IQ around 50 on the DAM, but scored considerably higher on three other tests (which are discussed below). Fahmy indicated that "[the] children who had no schooling, never used a pencil, and have no experience in how to conceptualize their visual image" (p. 172). Moreover, most of the unschooled children, that Fahmy "recruited from [...] the bush" (p. 164), were naked. It is noteworthy that within Goodenough's scoring scheme of the DAM test, five out of a total of 51 points are awarded for including (drawn) clothes. Not surprisingly, Fahmy considered the DAM test to be unsuitable for these Sudanese children, regardless of school attendance. Likewise, on the basis of a study with the DAM in Sudan, Badri (1965a) suggested that new scoring rules should be developed for Sudanese children, because they often draw traditional Sudanese clothing, which offers little scoring opportunities on the DAM. In another study in Sudan, Badri (1965b) noted that: "Many [children from remote villages] hold pencils in unusual ways and state they have never before made a drawing on paper" (p.333). Like Fahmy, Badri concluded that the DAM was unsuited for Sudanese children. Generally, the DAM appears to be unsuitable for African children without schooling (cf. Serpell, 1979). The most obvious reasons for this are inexperience with pencils and pencil drawing, and the unfamiliarity with two-dimensional pictures, which is often encountered among these children.

Hunkin (1950) administered the DAM to a "fairly representative" sample of South African urban schoolchildren, and indicated that "Classroom conditions were not ideal from the point of view of scientific test administration" (Hunkin, 1950, p. 54). This may have lowered test performance in this sample. Hunkin found that the average item scores in her sample of African children correlated quite well with the average items scores in the US standardization sample. On the other hand, she also found that the properties of some items of the DAM were different in these two samples, which suggests the presence of differential item functioning.

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

In two South African samples the administration of the DAM appeared at least to be successful (Hunkin, 1950; Richter et al., 1989). Because mental-age IQs were used in the Hunkin study, we recomputed standardized IQs on the basis of means and SD of the raw scores in the South African sample and the US standardization sample (Hunkin, 1950), resulting in a mean IQ of 83 ($N = 1076$). Richter et al. (1989) indicated that the DAM scoring scheme underestimated the abilities of Black South African children above the age of 8, because some details "do not appear to be part of [their] culturally informed knowledge" (p. 5), which again suggests differential item functioning in the DAM.

We found additional studies with the DAM in Africa (Bakare, 1972; Bardet, Moreigne, & Sénécal, 1960; Minde & Kantor, 1976; Nwanze & Okeowo, 1980; Ohuche & Ohuche, 1973; Skuy, Schutte, Fridjhon, & O'Carroll, 2001). For instance, Ohuche and Ohuche (1973) administered the DAM to 202 children aged 5–11 in an experimental school in Sierra Leone, and found an average IQ of 95 in terms of US norms (Harris, 1963). Moreover, Bakare (1972) administered the DAM to 393 upper-class and lower-class Nigerian school children, and found an average IQ of 87. In another study (Bardet et al., 1960), Senegalese school children averaged an IQ of 78 in terms of US norms. Also, in a recent publication, Khaleefa and colleagues report on two standardization samples of the DAM in Sudan (Khaleefa, Abdelwahid, Abdulradi, & Lynn, 2008). These samples were described as representative samples of school children from 1964 and 2006, respectively.[3]

It should be noted that the 1963 norms of the DAM have been criticized for being inaccurate (Howard Scott, 1981). Howard Scott suggested that 10 IQ points should be added to the IQs. We chose not to employ this correction, because we want to compare our results to Lynn's (who did not employ the correction). It is noteworthy that several authors (Badri, 1965a; Minde & Kantor, 1976; Munroe & Munroe, 1983; Serpell, 1979), including the test developers themselves (Goodenough & Harris, 1950), have argued that the comparison of DAM scores across cultures is problematic because of cultural differences in experience with pencil drawing on paper, and because several aspects of the scoring scheme are clearly culturally loaded. These problems signal a strong need for more insight into differential item functioning of the DAM test. Unfortunately, we were unable to find any studies of measurement invariance of the DAM between western and African samples.

4.2. Kaufman-Assessment Battery for Children

In a series of studies, Boivin, Giordani and co-workers administered the K-ABC to children in the Democratic Republic of Congo (Boivin & Giordani, 1993; Boivin, Giordani, & Bornefeld, 1995; Giordani, Boivin, Opel, Dia Nseyila, & Lauer, 1996).[4] Lynn (and Vanhanen) used these data sets to estimate the average IQ of Africans. However, according to our criteria, the African data from the K-ABC do not provide a sound basis for calculating the average IQ of the African population. The first problem with these data is that the studies were mainly concerned with the effect of intestinal parasites (Boivin & Giordani, 1993) and malaria (Boivin et al., 1993) on cognitive development. For that reason, the children in these samples were all from underdeveloped rural areas. In some studies, children were specifically selected for their poor health (Boivin & Giordani, 1993). Of course, malaria and intestinal parasite infections are common in tropical Africa, but such selective samples are far from ideal to estimate the average IQ in healthy African populations.

More importantly, in these samples, the K-ABC tests were adapted for use in rural children in Africa (Giordani et al., 1996). To this end, the instructions and items were changed. The extent to which this altered the measurement properties of the K-ABC is unknown. For all of these children, individual cognitive assessment was an entirely new experience. More importantly, for most of the children, it was their first encounter with color-printed material. Giordani et al. (1996) studied the psychometric properties of the K-ABC in their rural African samples. They were reticent with respect to the comparability of these African scores to US norms. For instance, in some K-ABC subtests, items included objects that were rather unfamiliar to these test takers (e.g., telephones). As the original authors indicate, it is therefore likely that at least some items in the K-ABC show differential item functioning (Giordani et al., 1996), and that several subtests are not comparable across Western and African samples. Lynn (2006) and Lynn and Vanhanen (2006) also considered the scores in a sample of 184 Kenyan rural children, who all suffered from malaria (Holding et al., 2004). In this study, *all* subtests of the K-ABC were altered, so it is entirely unclear to what degree the alterations in the test allow for the comparison to US norms. These samples did not meet our inclusion criteria because of the changes in this test and the special nature of the samples.

We came across additional data from the K-ABC in Africa. First, in one study (Boivin, 2002), the average IQ on the basis of the (adapted) K-ABC in a sample of Senegalese children with a history of malaria ($N = 58$) equaled 82. In yet another study (Skuy, Taylor, O'Carroll, Fridjhon, & Rosenthal, 2000), the (unadapted) K-ABC was administered to 21 Black South African children, and the average IQ was found to be 103.

4.3. Wechsler Scales

Lynn (2006) included in his review of average IQ of Africans two studies using the Wechsler Adult Intelligence Scale-Revised (WAIS-R; Wechsler, 1981) (Avenant, 1988, cited in Nell, 2000), and three studies using the Wechsler Intelligence Scale for Children-Revised (WISC-R; Wechsler, 1974) (Fernández-Ballesteros, Juan Espinosa, Colom, & Calero, 1997; Skuy et al., 2001; Zindi, 1994a) and Lynn and Vanhanen (2002, 2006) used the data from two of these studies to estimate national IQs of Equatorial Guinea (Fernández-Ballesteros et al., 1997) and Zimbabwe (Zindi, 1994a).

Nell (2000) argued strongly that the WAIS-R and WISC-R underestimate the cognitive abilities of South African Blacks, and he provided the results of the Avenant study and a study

[3] The 1964 data included the data from Badri (Badri, 1965a,b), so we do not consider Badri's samples in our estimate of average IQ of Africans. However, we did include Badri (1965a) in our estimate of the SD of IQ because Khaleefa et al. did not report on a SD of IQ in the 1964 sample.

[4] The analyses in Giordani et al. (1996) were based on the samples described in Boivin & Giordani (1993) and Boivin et al. (1995) (M. Boivin, personal communication, June 9th, 2006), so we do not consider it separately here.

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

of his own to illustrate his point. Lynn (2006) presented the results obtained in these samples in support of a low IQ among sub-Saharan Africans. Nell concluded on the basis of these studies that a "language artefact" would affect the scores and that "the Wechsler tests lack validity for these subjects" (p. 27). Of course, Lynn has every right to disagree with Nell's assessment of the unsuitability of the Wechsler scales for African test takers. However, the data from these samples did not meet our inclusion criteria because the tests were either adapted or not administered in their entirety. In the first study, Avenant (1988, cited in Nell, 2000) reports on the performance of a sample of prison wardens ($N=140$) who took the WAIS-R. The test was "adapted […], with the wording of some items changed to prevent purely local difficulties in understanding" (Avenant, 1988, p. 3–4, cited in Nell, 2000). Nell computed an average IQ of 73 for the sample, but his computations are incorrect (i.e., Digit Span test scores were not taken into account): the average IQ should be 76. SDs of subtest scores were also reported, which enabled us to estimate the SD of IQ at around 8.5. In the second study reported by Nell, a sample of 157 employed Black South African men with fewer than 12 years of education took the Digit Symbol, Block Design, and Digit Span subtests of the WAIS-R. Average IQ on the basis of these three subtests is around 79 with a SD of approximately 9.

One study, often referred to in the literature (Lynn, 2006; Rushton & Jensen, 2005), is Zindi (1994a). This particular study was concerned with the suitability of the US version of the WISC-R for Zimbabwean high school children. Zindi clearly indicated that the WISC-R needed adaptation to remove language difficulties, and he stressed that some instructions and items in the WISC-R may not be appropriate for Zimbabwean children. In a subsequent study, Zindi (1994b) found that small alterations in the WISC-R greatly enhanced average IQ in Zimbabwean children. It is also noteworthy that the average IQ in the Zimbabwean sample appears to be higher than the figure of 67 reported by Zindi himself. We recomputed the average IQ using the average scaled scores reported by Zindi and the relevant IQ table in the WISC-R manual (Wechsler, 1974). This resulted in an average IQ of 79 for this Zimbabwean sample.

Skuy et al. reported on two samples of Soweto high school students. The first sample ($N=100$) took the DAM, and averaged an IQ of 83.2 ($SD=17.9$), and they also took six subtests from the WISC-R, on which their average score was equivalent to an IQ of 82.6 ($SD=14.1$). These students also took the WCST, on which their mean standardized score was somewhat lower. We used the averages of WISC-R and DAM in Table 2. Lynn (2006) used only WCST data ("IQ of 64"), but did not include additional data from the DAM or data from six WISC-R subtests to estimate IQ. The second sample from Skuy et al. ($N=152$) took the full WISC-R, the WCST, and several other neuropsychological tests. In this sample average IQ is lowered due to the low performance on the vocabulary subtest and other verbal subtests (Rushton & Jensen, 2005). Skuy and colleagues indicate that "language has a considerable effect on test performance" (Skuy et al., 2001, p. 1422). For most Black Africans, English is not the native language, and it is well known that the Wechsler scales have a strong English language component. Many items in the WISC-R are tailored to the US, and it is quite possible such items (e.g., those related to geography, history, and politics of the US) are biased. For instance, it is hard to imagine that an Information subtest item concerning the distance between a west-coast city and an east-coast city in the US will function equivalently in African and American samples. In addition, several of the non-verbal (performance) subtests in the Wechsler scales have items displaying typically western objects and situations that may be less familiar to African test takers. Thus, the possibility of cultural bias of the Wechsler scales cannot simply be ignored. The Soweto students in second sample of Skuy et al. scored higher on the WCST, with an average around 89.5, and we used, somewhat arbitrarily, the average of the PIQ and the WCST in Table 2.

Unfortunately, reliability, inter-subtest correlations, or validity were reported in none of the studies using the WAIS-R or WISC-R discussed above. Besides, we are not familiar with any confirmatory factor analyses of (western) Wechsler scales among Africans. The WISC-R data from Zindi (1994a) and Skuy et al. (2001) were submitted to analyses with the method of correlated vectors (Rushton, 2001; Rushton & Jensen, 2003). However, these analyses did not test whether the factorial structure the WISC-R of both these African samples was comparable to that of western samples. Moreover, the method of correlated vectors is not a suitable method to study measurement invariance (Dolan, 2000; Dolan, Roorda, & Wicherts, 2004; Lubke, Dolan, & Kelderman, 2001), and had been found wanting on other grounds (Ashton & Lee, 2005). Finally, we were unable to locate any studies of measurement invariance of the Wechsler tests in western and African samples.

In the last study with the WISC-R in Africa that Lynn (and Vanhanen) reported to substantiate their claim of low IQ among Africans, the IQ of a small sample 10–14 year-old children was found to be around 60 (Lynn & Vanhanen, 2002). However, the use of this sample is an error. The average IQ of the people of Equatorial Guinea is based on a lengthy book chapter (Fernández-Ballesteros et al., 1997). Although this chapter reports research conducted among members of an illiterate tribe in Equatorial Guinea, the WISC-R was not administered to these Africans. The forty-eight children, to whom the WISC (not the WISC-R) was administered, were from Spain, not from Equatorial Guinea. Clearly, Lynn (and Vanhanen) made a mistake in using this sample to estimate average IQ of Africans.[5]

There exist additional data on Wechsler IQ of Black South Africans. Akande (2000) reported on a sample of 63 normally-achieving, low-achieving, and learning disabled Black South African children, who took the Wechsler Preschool and Primary Scale of Intelligence (WPPSI) and the WCST. The sample cannot be considered normal, because more than half of the children were selected for their low IQ scores. Nonetheless, the average IQ on the WPPSI was 77.5,

[5] The chapter clearly indicates that this study with 48 subjects was conducted in Spain. The mean IQ is mentioned two times, the first time as follows: "A similar design was used in our second experiment with forty-eight subjects, 10 to 14-year-olds, attending a school for handicapped children (63.025 IQ mean)" (Fernández-Ballesteros et al., 1997, p. 253). In a later part of the chapter, the text clearly states that half of the subjects were diagnosed as having brain organic disorders. The lead author (R. Fernández-Ballesteros, personal communication, May 29th, 2007) indeed indicated to us that this sample was Spanish.

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

while the average performance on the WCST is equivalent to an average IQ of 84.5 (both in terms of US norms). Akande raised some doubts as to whether the WPPSI accurately reflected intelligence of these children, because English was often not the home language of the children. Nonetheless, Akande considered the WCST suitable for use with the Black South African children. We computed the average of the WCST and the WPPSI scores for our analyses. In another study, 40 educated adults scored an average IQ of 94 on the US WAIS-III (Shuttleworth Edwards et al., 2004).[6] In yet another study, the average WISC-R IQ of 21 Black children was found to be 84 (Skuy et al., 2000). These are the same children that were administered the K-ABC (average IQ: 103). Skuy and colleagues again noted language problems among the Black children on the WISC-R; they consider the IQ from the K-ABC to be more valid.

A slightly adapted version of the WAIS-III was standardized among English-speaking South Africans in the late 1990s (Claassen, Krynauw, Paterson, & Mathe, 2001), but Lynn (and Vanhanen) did not use these data. In terms of US norms, the average IQ of the representative sample of English-speaking Black South Africans ($N = 196$) was 86.4 (SD = 13.3) (Shuttleworth Edwards et al., 2004). Strictly speaking, US norms of the WAIS-III do not apply here because of the changes to the South African WAIS-III. However, the mean difference between Black and White South Africans in terms of South African WAIS-III norms is 15.8 IQ points, or about one SD. These data suggest that the Black–White difference in average IQ in South Africa is comparable to the Black–White difference in average IQ in the US. In addition, Claassen et al. conducted rigorous differential item functioning analyses of most WAIS-III subtests. They found several items that showed bias with respect to Black South Africans, particularly in the Verbal subscales.

Two additional studies used subtests of the WISC-R in Africa. In a study in Lagos, Nigeria, the Similarities subtest of the WISC-R was administered to a sample of aggressive and pro-social boys (Ani & Grantham-McGregor, 1998). The pro-social boys achieved an average score equivalent to an IQ of 87, while the aggressive boys' average was around 77. Average Similarities score of the entire sample was equivalent to a mean IQ of 81.8 ($SD = 14.3$). In another study (Kashala, Elgen, Sommerfeldt, Tylleskar, & Lundervold, 2005), the Digit Span subtest from the WISC-R was administered to 183 Congolese children, with and without ADHD. The averages of these children are equivalent to an IQ of 93.5 ($SD = 12.5$), but these data do not meet our second inclusion criterion. These children also took the CPM, which resulted in an average IQ of 78.5.

4.4. Culture Fair Intelligence Test

Buj (1981) provided the results of the administration of the Culture Fair Intelligence Test (Cattell & Cattell, 1973) to 225 adults in the Ghanaian capital of Accra. In this sample, which was stratified for gender, age (6 groups), and Socio-Economic Status (3 levels), average IQ was 82.2. Moreover, Nenty and Dinero (1981) administered the Culture Fair Intelligence Test to 803 students in seven secondary schools in both urban and rural areas in Nigeria. They found that these Nigerian adolescents scored on a par with a sample of 600 high school students from four schools in Portage County, Ohio. It is not possible to compute an average IQ of these samples on the basis of US norms, because test takers were given extra time to complete the subtests of the Culture Fair Intelligence Test. Nonetheless, it's hard to imagine that average IQ in both the American and the Nigerian sample is below 90 (we assumed that the mean IQ in the US sample was 100 to compute the mean in Table 2). In contrast to most studies considered thus far, this study actually considered the possibility of measurement bias, which was studied using contemporary psychometric modeling. Some evidence for differential item functioning was found, although the effects were not large and mixed in direction.

In an experimental study into the effects of stereotype threat (Steele, 1997) on intelligence test performance, Klein, Pohl, and Ndagijimana (2007) administered the Culture Fair Intelligence Test to African immigrants in Belgium under different conditions. These conditions differed in the degree to which negative stereotypes concerning the cognitive ability of Africans were made salient for the test takers. In the conditions in which the stereotypes were stressed, the Africans ($N = 30$) averaged an IQ of 74.9 ($SD = 13.6$), while Africans ($N = 28$) in the control condition (i.e., no stereotype threat induced) averaged an IQ of 89.5 ($SD = 12.0$). We included the latter sample in Table 2 and in our analyses. Because of the small sample size, the addition of the former sample does not change any of our estimates.

4.5. Other IQ tests

Lynn (and Vanhanen) also considered studies in Africa using several less-well known IQ tests. One of these studies was concerned with the effects of coaching on test performance (Lloyd & Pidgeon, 1961). In this study, a sample of South African Zulu children were given the Non-Verbal Test, a test normed on English children, and published in 1951 (Buros, 1959). The Zulu children ($N = 275$) had an average IQ of 88.7 on the pretest (i.e., without coaching). Lynn (2006) did not discuss how he arrived at his estimate of an IQ of 74 for this sample.

In a cross-cultural study, Vernon (1969) administered a battery of cognitive tests to fifty Ugandan boys and hundred English boys. He computed average IQs of each test on the basis of score distributions in the English sample. On 16 of the 21 tests, mean IQ of the Ugandan boys was above 80. The mean on the 21 subtests equals 86 (which is the figure we used in our analyses). This equals 88 if we leave aside the IQ of an English vocabulary subtest on which these boys scored very lowly ($M = 57$). Lynn (2006) assigned an estimate of 80 to this sample, but provided no rationale for his downward correction. Vernon's data does not meet our first inclusion criterion, because the IQs were based on an English convenience sample, and not on published norms. Besides, Vernon himself computed inter-subtest correlations in this sample, and found no indication of a *g* factor comparable to that in other samples. A later factor analysis of part of

[6] Two minor changes were made in the administration of the WAIS-III in this South African sample, viz. a change of Dollars in Rand and of Miles in Kilometers in the items of the Arithmetic subtest (A. Shuttleworth-Edwards, personal communication, June 21st, 2007).

Vernon's data by Hakstian and Vandenberg (1979), corroborated that "the cognitive structure among Ugandan subjects may be slightly different from that of other cultures" (p. 87). This is an interesting result, if only because several tests used by Vernon were also used in older studies in Africa.

In one of those old studies, fifty 5–13 year-old children from the Sousou tribe in rural Guinea were administered the Army Beta Test (Nissen, Machover, & Kinder, 1935). These unschooled test takers suffered from "inexperience in manipulating a pencil" (p. 325), which is a handicap in taking the Army Beta. Moreover, on some subtests it was clear that most test takers did not understand what was expected from them. For instance, "[t]he subjects appeared utterly bewildered" (Nissen et al., 1935, p. 331) when confronted with the Manikin and Feature Profile subtest of the Army Beta. These difficulties notwithstanding, Lynn (and Vanhanen) assigned this particular sample an IQ of 63. The Army Beta was also administered to 293 Black South African children by Fick (1929). With respect to representativeness of samples, Fick clearly stated that "sweeping generalizations regarding whole groups should be avoided" (p. 904). He also acknowledged that the test scores may have been lowered due to the fact that "the native does not grow up with pictures and diagrammatic representations of things" and that "some of the items […] are not quite fair to the native child" (p. 909). In light of these difficulties, and because of the absence of *any* indication of the reliability, validity, or correlational structure of the Army Beta tests in this sample, these data fail to meet our inclusion criteria. Dent (1937) is another old study on the suitability of western intelligence tests among Black South Africans. Dent considered his sample of 80 test takers too small to make any generalizations. With regard to the use of the Koh's Block test (the predecessor of the Block Design test in the Wechsler scales), Dent remarks that "all subjects experienced difficulty with this test" (p. 462). We take this to mean that the subjects did not understand instructions, so these data do not meet our third inclusion criterion.

The studies reported in the paragraph above were severely criticized as early as the 1940s (Biesheuvel, 1943), and cannot be taken seriously today. To begin with, the Army Beta test originates from the first years of intelligence testing, and is now completely obsolete (Jensen, 1982, called this test "primitive"). More importantly, administering a paper and pencil test with pictures and diagrammatic representations to persons inexperienced with pencils and unfamiliar with such representations is unlikely to provide a valid indication of intelligence. The situation is exacerbated by the fact that the pictures used in the Army Beta are likely to be culturally biased, because the pictures display typically American objects and situations. For instance, one item displays a tennis match; the test takers are required to draw the missing tennis net (Lane, 1994). By modern psychometric standards, these old studies cannot be taken seriously, certainly not to estimate average IQ of the African population.

Ferron (1965) reports test results with an unknown IQ test in seven samples of children in Nigeria and Sierra Leone (combined $N=1633$). Ferron considers this test unsuitable for African children. Despite this, Lynn (2006) and Lynn and Vanhanen (2006) included in their reviews the average IQs from the two lowest scoring samples. Lynn (2006) briefly discussed a third sample with an average IQ of 81, which was not included by Lynn (2006) and Lynn and Vanhanen (2006). Unfortunately, Lynn (and Vanhanen) did not explain why they excluded the scores of the four higher scoring samples in Ferron's overview of the results with the Leone test in Africa. The average IQ in all African samples discussed by Ferron ($N=1633$) is around 77.

Besides the DAM test, Fahmy administered three additional tests to his sample of Sudanese children, namely Alexander's Passalong, Porteus Maze test, and the Goddard Formboard. The average IQ on these tests was 94, 76.5, and 73.5, respectively. Fahmy claimed that the latter two tests are "unsuitable" (p. 169) and "of little use in measuring intelligence" (p. 166) among these Sudanese children. Also, it appears that the average IQ of this sample was considerably lowered because of the unfamiliarity of these children with pencil drawing, resulting in their low performance on the DAM test. Fahmy considered Alexander's Passalong test as the only test suited for these Sudanese children, so we used this IQ mean in the analyses.

Lynn (and Vanhanen) used WCST data from a study in Tanzania by Sternberg et al. (2002) to argue for a low IQ level among Africans. Sternberg and colleagues aimed to show that without coaching, the WCST is inappropriate for African children, so these data do not meet our inclusion criteria. In another study (Sternberg et al., 2001), rural children from Kenya were given the Mill–Hill vocabulary scale, but English was not the home language of the children. The WCST and the Mill–Hill vocabulary scale are not meant to be measures of *g*, so these data do not meet our second inclusion criterion.

Lynn (2006) mentioned data from the Junior Aptitude Test or JAT in South African Blacks (Lynn & Owen, 1994). This test has been shown to be severely biased, both at the subtest level (Dolan et al., 2004) and at the item level (Owen, 1989). We do not consider these data because of the established bias, and because Owen (1989) concluded that the test was unsuitable for Black students. This sample also took the Raven's test (Lynn, 2006 included this sample twice), and we did include these scores in our systematic review on the performance of Africans on the Raven's. We also included in Table 2 the scores of Nigerian children (Fahrmeier, 1975) who took the Spatial Relations subtest of the Primary Mental Abilities (PMA; Thurstone, 1963).[7]

Finally, we found additional studies showing higher average IQs for Africans. First, in a study in Zimbabwe, Wilson, Mundy-Castle, and Sibanda (1991) administered the full PMA to fifty-two Black school girls aged nine. In terms of US norms, their average IQ was approximately 96. Second, Ashem and Janes (1978) administered a *translated* version of the McCarthy Scale of Children's Abilities to a sample of ($N=128$) "well-to-do" and poor children from rural and urban areas in Nigeria. The sample included normally-nourished, reasonably nourished, and malnourished children. The average IQ of the entire sample was 92.6 ($SD=19.5$). Third, Yoloye (1971) administered the Lorge–Thorndike intelligence tests to a sizeable sample of secondary school boys in Idaban, Nigeria. Test administration was not conducted according to official guidelines, because these boys were given

[7] These children also took the CPM, on which they averaged much lower scores. Previously, Lynn (1991) used both PMA and CPM to estimate this sample's average IQ, but in later work (Lynn, 2006; Lynn & Vanhanen, 2002, 2006) the mean IQ was based on CPM data only.

extra time to complete the test. Average IQ equaled 88 on the basis of then-recent US norms. Fourth, Ferron (1965) refers to a study by Dunstan (1961, cited in Ferron, 1965) on a sample of "over 400" form I students in Freetown, Sierra Leone. These students took one of the Moray House Tests (no. 44) and Ferron indicated that their mental-age IQ was around 77.8 (i.e., they had a chronological age of 13.5 and a mental-age of 10.5). Ferron indicated that Dunstan reported problems during test administration and a "language handicap". These data do not meet our inclusion criteria.

4.6. Meta-analytic analyses

Table 2 includes a total of 44 samples of African test takers for which average IQs could be determined. For obvious reasons, we did not include the deaf children (Alade, 1992) and the Spanish children (Fernández-Ballesteros et al., 1997) in the analyses. The table gives the raw mean IQ of the sample along with the SD and average age. These IQ estimates are all based on US norms, except for the samples from Lloyd and Pidgeon (1961), Dunstan (1961, cited in Ferron, 1965), and Vernon (1969), which were based on UK norms. Lynn (and Vanhanen)'s estimates of national IQ are expressed in terms of UK norms corrected for outdated norm tables due to the Flynn Effect. Average IQ in the United States is approximately 2 points lower than the average in Great Britain (Lynn & Vanhanen, 2002; Raven et al., 1996), so the IQs from US norms need to be lowered by 2 points to arrive at an average IQ in terms of British norms. In line with Lynn (and Vanhanen)'s approach, we employed a correction for the Flynn Effect of 3 points per decade. So Table 2 includes the raw average IQs as well as the average IQs in terms of UK norms with Flynn Effect correction. The latter can be compared to the estimates of average IQ by Lynn (2006) and Lynn and Vanhanen (2006).

Before we turn to the mean IQs, we consider the SDs in the 29 samples in which the SDs was either reported or could be computed (cf. Table 2). The meta-analytic estimate of the *mean* of the population SDs was 12.6, with a 95% confidence interval of [11.5–13.7]. The Dersimonian–Laird estimate of the *variance* of the population SDs, or Tau^2 was equal to 7.9. The I^2 (Higgins, Thompson, Deeks, & Altman, 2003) was 96.6., suggesting sizeable heterogeneity in SDs across studies.[8] The heterogeneity is such that 68% of the population SDs lies between 9.8 and 15.4. This suggests that the SD of IQ in most African samples is somewhat lower than it is in western population samples (i.e., $SD = 15$). It is noteworthy that the SD of IQ from the WAIS-III in the representative South African sample described by Claassen et al. (2001) was 13.3, and that the SD of IQ in the recent Sudanese representative sample of the DAM was equal to 10.1. The third sample described as representative by the authors (Hunkin, 1950) showed an SD of 16.2. A meta-analytic estimate of the SD of IQ based on these three representative samples was equal to 13.2 (95% confidence interval: [8.8, 17.6]). Thus, the relatively low SD of IQ of Africans is also apparent in representative samples.

It is well known that volunteer subjects are often of above average IQ (Rosenthal & Rosnow, 1975), which may have introduced *sampling bias* of the samples of Africans. It is to be expected that the primary school samples in the current study do not suffer from this problem, because the majority of children in (contemporary) Africa attend primary school. If sampling bias were present in the samples under review, the SDs of IQ should be smaller in samples from secondary schools and in samples of (volunteer) adults. In addition, under this type of sampling bias, we would expect the convenience samples to have a smaller SDs and higher mean IQs than more representative samples. We checked this possibility by means of a meta-regression model with two predictors of sample SD (and M): an indicator of representativeness of the sample (three samples are considered representative, two of which reported SDs; cf. Table 2) and an indicator of whether sample was under the age of 13 (i.e., primary school ages) or older (i.e., secondary school and adults). Both predictors failed to significantly predict the SD of IQ in the samples: $Zs < .40$, $p > .70$. We also used these indicators to predict the *mean* IQ in the samples, and found similar results: $Zs < .70$, $p > .40$. Hence, the SD of IQ in African samples appears to be around 13 and sampling bias does not appear to be an issue in the current review. In addition, we performed an analysis to verify whether the secondary school samples have higher mean IQs and smaller SDs than the other samples (i.e., are more selective), but this was not the case: $Zs < .6$, $p > .50$. The mean IQ on the basis of the seven secondary school samples ($N = 3268$) was 77.8 (95% confidence interval: [68.9, 86.7]).

Table 3
Results of meta-regression of samples mean IQ ($N = 42$) on the basis of our inclusion criteria.

Predictor		Mean (SE)	95% conf. inter.
Criterion 1	Standardized western IQ norms	78.7 (3.0)	[72.8, 84.6]
	No standardized western IQ norms	73.6 (2.5)	[68.8, 78.5]
	$Z = 1.68$, $p = .092$, variance = 79.6 ($SE = 17.7$)		
Criterion 2	Test administered according to guidelines	78.0 (3.0)	[72.1, 83.8]
	Test not administered according to guidelines	75.5 (2.4)	[70.8, 80.1]
	$Z = 0.83$, $p = .401$, variance = 83.7 ($SE = 18.6$)		
Criterion 3	No reported administration problems	78.2 (4.2)	[70.1, 86.4]
	Reported administration problems	68.4 (3.9)	[60.8, 76.1]
	$Z = 2.36$, $p = .018$, variance = 74.9 ($SE = 16.7$)		
Criterion 4	No psychometric problems/no measurement bias	80.0 (2.7)	[74.6, 85.4]
	Psychometric problems/ measurement bias.	73.4 (2.0)	[69.4, 77.4]
	$Z = 2.41$, $p = .016$, variance = 74.7 ($SE = 16.6$)		
Criterion 5	Normal sample	77.7 (4.2)	[69.5, 85.9]
	Non-normal sample	73.7 (3.8)	[66.3, 81.1]
	$Z = 0.96$, $p = .339$, variance = 83.2 ($SE = 18.5$)		

[8] We present different measures of heterogeneity in this paper. First, the Q test is a statistical test of heterogeneity of true effect sizes. A significant Q test suggests heterogeneity because it indicates that observed differences in effect sizes show more variance than is to be expected from sampling error variance. Second, the Tau^2 is the Dersimonian–Laird estimate of the *variance* of the population effect sizes based on the *Q* test. Third, the I^2 is a proportion that indicates the degree to which observed effect sizes in a meta-analysis are due to true differences in the underlying effect sizes (Higgins et al., 2003). In this analysis, it estimates the proportion of the observed variance between sample's SDs that reflects differences in true population SDs.

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

Table 4
Results by subsets of samples.

Subset	No. of samples	*N*	Mean IQ (SE)	95% conf. inter.	Q [b]	Tau^2 (SE)	I^2
All samples in Table 2	42	14129	77.1 (1.3)	[74.6, 79.5]	5845	66.7 (23.1)	99.3
Samples that meet our five inclusion criteria	12	2544	82.6 (2.6)	[77.6, 87.6]	1132	76.8 (45.0)	99.0
Samples considered by Lynn and Vanhanen (2006)	11	2056	67.4 (1.4)	[64.6, 70.2]	268	21.4 (11.1)	96.3
Samples not considered by Lynn and Vanhanen (2006) [a]	27	7759	80.4 (1.8)	[76.9, 83.8]	3757	83.9 (31.4)	99.3
Army Beta	3	423	65.4 (1.3)	[63.0, 67.9]	7	3.3 (4.8)	72.0
Culture Fair Intelligence Test	3	1056	86.7 (7.5)	[72.0, 101.3]	408	165.6 (210.0)	99.5
Draw-A-Man test	10	5953	77.7 (2.2)	[73.3, 82.0]	1481	47.8 (30.6)	99.4
Kaufman-Assessment Battery for Children	5	447	73.0 (4.1)	[65.0, 80.9]	215	80.0 (66.7)	98.1
WISC-R	4	633	75.4 (2.7)	[70.2, 80.7]	83	27.6 (24.5)	96.4
WAIS-R/WAIS-III	4	528	79.4 (3.4)	[72.7, 86.0]	153	44.2 (43.2)	98.0

[a] Only those studies published before 2006.
[b] Distributed as Chi-square with degrees of freedom equal to number of samples minus one.

As can be seen in Table 2, most of the samples fail to meet our inclusion criteria. The reasons to exclude data were (1) unavailability of standardized western norms (14 samples), (2) test not an IQ test or test not administered according to official guidelines (15 samples), (3) test administration described by original authors as problematic (6 samples), (4) test found or deemed to be biased (19 samples), and (5) inappropriate sample (9 samples). We used our inclusion criteria to predict the mean IQ in the samples. The results are given in Table 3. As can be seen, our inclusion criteria have a positive effect on the predicted average IQ. The effects of criteria 1, 2 and 5 are quite small, because the analysis includes both low-scoring and high-scoring samples that did not meet these criteria. Combined, the criteria explained 27% of the variance in mean IQs. The effects of reported problems during test administration (Criterion 3), and supposed or established measurement bias (Criterion 4) on the predicted average IQ are substantial: around 10 and 6.5 IQ points, respectively. The *predicted* average IQ of the samples that meet or did not meet our inclusion criteria are given in Table 3. In Table 4 we report various estimates of the mean IQ of Africans on the basis of alternative selection of samples. The estimated mean IQ on the basis of our five inclusion criteria is 82.6, but there was a clear indication of heterogeneity in mean IQs ($Q=1132$, $DF=11$, $p<.001$). The estimate of the variance of population means or Tau^2 was 76.8 (i.e., $SD=8.8$), suggesting that 68% of the population mean IQs are expected to lie between 73.8 and 91.4. The use of all samples in Table 2 resulted in a mean estimate that was lower at 77.1. Table 4 also includes the estimate of mean IQ on the basis of Lynn and Vanhanen's (2006) selection of samples (of the IQ tests under review here) and estimates based on specific IQ tests. It can be seen that there are clear differences between IQ tests in the estimated mean IQ, with the Army Beta tests showing the lowest estimate and the Culture Fair Intelligence Test the highest estimate. Because of confounding of tests used and

Table 5
Estimates of mean IQs per country on the basis of studies in Table 2 and studies from the Raven's study.

Country	Nat. IQ	Based on all samples in Table 2				Based on samples that meet incl. crit. and Raven's test data [a]		
	L and V (06)	No. of samples	Quality rating	Mean IQ (SE)	95% conf. inter.	No. of samples [b]	Mean IQ (SE)	95% conf. inter.
Centr. Afr. Rep.	64	0	(only 1 study with the Raven's)			0+1	71.0 (0.4)	[70.1, 71.9]
R. Congo	64	0	(only studies with the Raven's)			0+3	75.9 (0.9)	[74.2, 77.7]
D.R. Congo	65	3	2.5 out of 5	71.3 (5.7)	[60.2, 82.5]	0+4	77.8 (2.1)	[73.7, 81.9]
Eritrea	68 [c]	0	(only 1 study with the Raven's)			0+1	85.0 (1.2)	[82.6, 87.4]
Ethiopia	64	0	(only studies with the Raven's)			0+4	69.4 (0.8)	[67.9, 71.0]
Ghana	71	1	5 out of 5	75.7 (0.8)	[74.1, 77.3]	1+1	73.3 (2.3)	[68.7, 77.9]
Guinea	67	1	2 out of 5	63.0 (1.6)	[59.9, 66.1]	None of the samples meet inclusion criteria		
Kenya	72	2	2 out of 5	63.3 (0.8)	[61.8, 64.9]	0+3	80.4 (3.2)	[74.1, 86.7]
Madagascar	82	0	(only 1 study with the Raven's)			0+1	82.0 (1.2)	[79.5, 84.5]
Mali	69 [c]	0	(only studies with the Raven's)			0+3	74.1 (1.2)	[71.7, 76.5]
Namibia	70 [c]	0	(only 1 study with the Raven's)			0+1	74.0 (1.4)	[71.3, 76.7]
Nigeria	69	8	3.6 out of 5	83.5 (3.1)	[77.4, 89.5]	0+9	83.8 (2.0)	[79.9, 87.6]
Senegal	66 [c]	2	3.5 out of 5	70.5 (4.4)	[61.8, 79.2]	1+0	66.3 (0.5)	[65.3, 67.3]
Sierra Leone	64	2	3.5 out of 5	84.5 (6.8)	[71.3, 97.8]	1+0	91.3 (0.9)	[89.6, 93.0]
South Africa	65 [d]	14	3.8 out of 5	76.1 (1.9)	[72.4, 79.8]	7+9	77.1 (1.7)	[73.8, 80.4]
Sudan	71	3	3.3 out of 5	74.0 (1.8)	[70.6, 77.4]	None of the samples meet inclusion criteria		
Tanzania	72	1	2 out of 5	72.0 (0.6)	[70.8, 73.2]	None of the samples meet inclusion criteria		
Uganda	73	2	4 out of 5	84.4 (0.9)	[82.6, 86.2]	1+0	83.9 (0.8)	[82.3, 85.5]
Zambia	71	0	(only studies with the Raven's)			0+6	78.5 (1.9)	[74.7, 82.3]
Zimbabwe	66	2	4.5 out of 5	78.8 (7.3)	[64.5, 93.1]	1+1	81.5 (4.6)	[72.5, 90.5]

Notes:
[a] See Wicherts et al. (2009) for details on the studies on the Raven's tests.
[b] First number indicates number of samples from Table 2 that meet our inclusion criteria and the second number indicates the number of samples that meet the inclusion criteria of the Raven's study.
[c] National IQ estimated by Lynn and Vanhanen (2006) on the basis of neighboring countries.
[d] Estimate based on Black South Africans only.

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

the types of samples that took these tests, these differences between IQ tests need to be interpreted with caution. At the request of the reviewers, we also provide estimates of mean IQ per country in Table 5. We discuss these estimates below.

4.7. Publication bias

Several mechanisms may give rise to *publication bias* in the current literature review. First, our sample of studies can be biased such that higher-scoring samples are more often reported in the literature and/or are more often used in studies. Second, it is possible that authors of papers may shy away from reporting mean IQs in cases where the mean IQ of Africans is (much) lower than in western samples. Third, it is possible that we simply failed to take into account higher-scoring samples in our review, e.g., because studies with higher-scoring samples are more often published in poorly disseminated African journals. Note that the first two mechanisms are expected to result in lower-IQ samples missing from the review, i.e., a positive publication bias. The third mechanism may give rise to higher-scoring samples being missed in the review, i.e., a negative publication bias. In the case of positive publication bias in a review (e.g., only studies with high average IQ are published and/or report the mean IQ), one would expect a *positive* relation between the mean in the sample and the standard error of the sample's mean. In the case of a negative publication bias, i.e., if higher-scoring samples are missing, there will be a negative relation between the standard error and the mean in the sample. In case of no publication bias, the funnel plot will be symmetric (Light & Pillemer, 1984), and the standard errors will not be predictive of the means in the samples.

Fig. 5 displays the mean IQs of all African samples that meet our inclusion criteria. The heterogeneity in mean IQs is quite clear. There is no clear asymmetry in this funnel plot, although the relatively low-scoring sample of Sengalese children (Bardet et al., 1960) does introduce some asymmetry (this sample took the DAM test with relatively old norms). We formally tested for publication bias by using various methods. The Egger test (Egger et al., 1997) did not show an indication of publication bias: $t(10) = 0.78$, $p = .22$, nor did an alternative version of this test in a mixed effects model (Sterne & Egger, 2005): $Z = 1.75$, $p = .08$. However, the trim and fill method[9] (Duval & Tweedie, 2000) did fill in one potentially missing study on the left-hand side of the funnel plot. This lowered the estimated mean slightly to 81.4 (95% confidence interval: [76.5, 86.3]. In addition, we conducted a cumulative meta-analysis (see Appendix A) that showed hardly any trend in the estimate of the mean when less precise samples are added (although it is still evident that the Sengalese data from Bardet et al. represents somewhat of an outlier). Thus, publication bias does not appear to strongly affect our results. At the same time, however, negative publication bias appears to be present in Lynn and Vanhanen's (2006) review (with respect to the IQ tests under review here); they used 11 of the samples in Table 2 to estimate national IQ in various countries. If we consider the studies that were published prior to 2006, it is clear that the samples (see Fig. 6), which Lynn and Vanhanen (2006) considered, show significantly lower mean IQs than the samples that they did not consider: $Z = 4.50$, $p < .001$. The same applies to the 21 samples that were considered by Lynn (2006): $Z = 5.25$, $p < .001$. Thus, the studies considered by Lynn (and Vanhanen) are not representative for published studies with IQ tests in Africa.

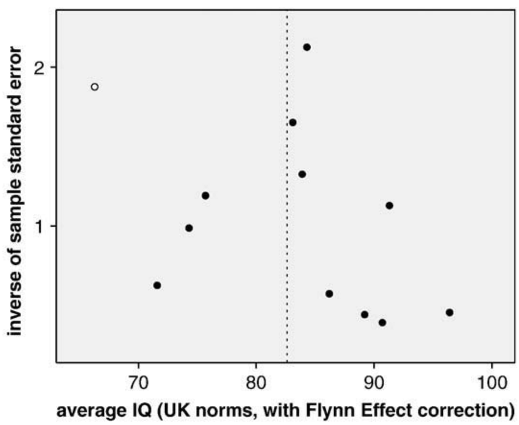

Fig. 5. Mean of samples that meet our inclusion criteria against the inverse of the standard error.

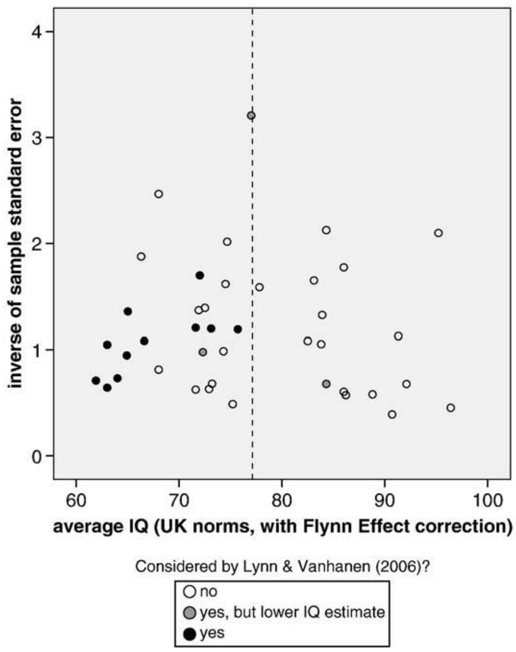

Fig. 6. Mean of samples from studies published prior to 2006 against the inverse of the standard error.

5. Conclusion

The assertion that the average IQ of Africans is below 70 is not tenable, even under the most lenient of inclusion criteria.

[9] The trim and fill method (Duval & Tweedie, 2000) enables a correction for publication bias by filling in fictional studies that are missing due to publication bias. Duval (2005) has indicated that the method may not be robust against heterogeneity in effect sizes, so our results should be interpreted with caution. However, the use of a fixed effects model in performing this analysis would be problematic given the clear indication of heterogeneity in population means.

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

The inclusion of all studies in Table 2 results in an estimate of an average IQ of 77. This estimate represents an underestimate of the true average IQ, because it is based on (1) inaccurate (often ad hoc) IQ norms or norms based on mental-age IQs, (2) IQ tests that were not administered according to official guidelines (e.g., adapted), (3) studies of IQ in which the test administration was problematic, (4) IQ tests found to be biased with respect to Africans, and (5) IQ test scores of non-normal (e.g., unhealthy) samples. According to our inclusion criteria, the average IQ (in terms of UK norms) of the African samples on the basis of the tests featured in our review is 81 or 82. This is close to mean of the national IQs as estimated from the scholastic achievement surveys, which was 81.8. Our systematic review of the average performance of Africans on the Raven's tests suggested that the average IQ of Africans in terms of western norms on the Raven's is approximately 78 (UK norms) or 80 (US norms). Thus, it can be concluded that the average IQ test performance of Africans in terms of UK norms is around 81. We found neither an indication of sampling bias, nor an indication of publication bias. We also found that the SD of IQ in African samples appears to be somewhat lower than the value of 15 found in most western samples. Although this may in part be caused by the common use of convenience samples in Africa, the representative samples also evidenced an SD of IQ for Africans that was lower than 15. In addition, we found a clear indication of heterogeneity of mean IQs across the African samples. This is likely due to the use of different IQ tests, different types of samples, differences between countries, and differences in sampling methods. However, the convenience samples did not show higher means or smaller SDs than the representative samples. In addition, the means and SDs in the samples of primary school children did not differ from the means and SDs in the samples of older-aged test takers. Of our inclusion criteria, only documented problems during test administration and psychometric problems were predictive of the mean IQs. Thus, it is important to consider these issues in future research and to take them into account in estimating the mean IQ of African samples.

Needless to say, the use of alternative inclusion criteria results in different estimates of the average IQ of Africans. For instance, one of the reviewers of this paper remarked that some samples of secondary school students may be considered selective samples of above average IQ. However, average IQ on the basis of the samples secondary school students was not appreciably higher than the average IQ on the basis of the other samples. Regardless of one's choice of criteria, it is important that all studies that meet one's criteria be included in one's review. We excluded several samples from our review because the western standardized norms were not applicable or unavailable. If one wishes to drop this particular criterion, then the sample reported by Nenty and Dinero (1981) ($N = 803$, average IQ well-above 90), the Sudanese DAM data from Khaleefa et al. (2008, $N = 2690$, average IQ around 71), the data from all the samples described by Ferron (1965; $N = 2033$, average IQ around 77), and the standardization sample ($N = 196$, average IQ of 86) of the WAIS-III for black South Africans (Claassen et al., 2001) should also be included. In systematic reviews, inclusion criteria should be explicated in advance and used consistently. In systematic reviews, samples should never be excluded without a clear, and hopefully sound, rationale. As a past editor of Psychological Bulletin puts it: "[Those] who accumulate and integrate other people's data ought to be held to similar standards of methodological rigor as the researchers whose evidence forms the bases of their review" (Cooper, 2003, p. 3). On the other hand, in *unsystematic* literature reviews, the grounds of important decisions, which may affect the outcome of the review, remain unknown. This is at odds with the scientific principles of verification, openness, and replicability.

Our estimate clearly differs from that of Lynn (and Vanhanen). First, Lynn (and Vanhanen) apparently used different inclusion criteria. Unfortunately, their inclusion criteria are neither mentioned nor discussed (bar some rare cases). Second, we came across several downward errors in the computation of average IQ (e.g., Fahmy, 1964; Lloyd & Pidgeon, 1961). Third, our extensive search for relevant studies resulted in additional studies of IQ in Africa that Lynn (and Vanhanen) missed. This was partly caused by the fact that we had access to African journals that did not show up in Lynn (and Vanhanen)'s work. Because Lynn (and Vanhanen) missed a sizeable portion of the relevant literature, their estimate of average IQ of Africans is clearly too low. Combined, the current systematic review and the results of our review of Raven's tests suggest that the average IQs of African test takers is close to 80. We believe that the accuracy of estimates of national IQ of sub-Saharan African countries can be improved considerably.

In Table 5, we report estimates of the mean IQ per country for all countries that are featured in Table 2. We also report Lynn and Vanhanen's (2006) estimates of national IQ as well as the average number of inclusion criteria that is met by the samples in each country. These estimates should be interpreted with caution, because they also include studies with inappropriate samples that do not meet our inclusion criteria. Moreover, the estimates on the basis of the studies in Table 2 do not take into account the results from the many studies in Africa with the Raven's tests. Therefore, on the right-hand side of Table 5, we provide estimates of the mean IQ of African countries on the basis of those samples from the current review that meet our five inclusion criteria and samples for each country from our review of the Raven's tests. Also included are mean estimates for countries for which there were only data from the Raven's. Our inclusion criteria in our review of the Raven's test performance of Africans were quite similar to the criteria we employed here. For instance, we only included samples for which there exist valid UK norms and only included scores from the Raven's that were administered according to the test guidelines. We refer to Wicherts et al. (2009; available upon request) for a detailed discussion of these studies. The estimates of the average IQs of these countries are quite close to the predicted national IQs on the basis of the scholastic achievement surveys. It is important to note that these estimates of the mean IQ performance of African populations, just like the national IQs from Lynn and Vanhanen (2002, 2006), are based for the most part on convenience samples rather than representative samples. This may have affected the results. These estimates are unlikely to be the final word on this topic, for some readers may disagree with our inclusion criteria.

Numerous studies in the literature have used Lynn and Vanhanen's (2006) estimates of national IQ. Although one study suggested that the precise estimates of national IQ of low-scoring countries had little effect on the correlation between national IQ and Gross-Domestic Product (Whetzel & McDaniel, 2006), the effect on other correlations remains to be studied. Generally, the correlations between national IQs and means of scholastic achievement surveys improve when the national IQs of the African countries are set at the values in Table 5 or at 81 (for countries not featured in Table 5) instead of the lower values from Lynn and Vanhanen. For instance, the correlations reported by Lynn (2006) and Lynn and Vanhanen (2006) between national IQs and the collation of results from Hanushek and Kimko (2000) increases from .796 to .863. This appears to suggest that Lynn and Vanhanen's (2006) estimates of national IQs in other parts of the world are more accurate than those in sub-Saharan Africa. At the same time, other ecological correlations from the literature are lowered by increasing the national IQs of African countries to the values in Table 5 or to 81 (for the other countries). For instance, several of the correlations reported by Templer (2008) and Kanazawa (2008) are affected negatively when national IQs in sub-Saharan Africa are corrected upwards. For instance, the correlation between distance from Ethiopia and national IQs in Kanazawa's (2008) study ($N = 113$) drops from $r = .198$ ($p < .05$) to $r = .113$ ($p = .23$) and the correlation between national IQs and HIV/AIDS rate ($N = 70$) reported by Templer (2008) changes from $r = -.635$ to $r = -.481$. Likewise, Rushton and Templer (in press) correlated national IQs with several crime-related variables, but after correcting the national IQs in Africa the correlations changed from $r = -.253$ to $r = -.261$ for homicide, from $r = -.290$ ($p = .002$) to $r = -.229$ ($p = .015$) for rape, and from $r = -.215$ ($p = .02$) to $r = -.162$ ($p = .09$) for assault. The robustness of these and other findings against alternative estimates of national IQs in Africa should be addressed in future studies.

There can be little doubt that Africans average lower IQs than do westerners. Several factors may cause this. Lynn (2006), Rushton (2000), and Kanazawa (2004) have proposed evolutionary theories to explain the relatively low scores of Africans on IQ tests. However, the fact is that African countries are developing countries, and we view this as highly relevant in the explanation of the low IQ test performance of Africans. Specifically, Flynn (1987, 2007) has shown that IQ levels have increased considerably in the developed world over the course of the twentieth century. African countries below the Sahara have not experienced the improvements in the variables that have been proposed to have caused the Flynn Effect in the developed world. These include improvements in nutrition and health (care), increases educational attainment, improvements in educational practices, urbanization, large-scale dissemination of visual–spatial toys, etc. Although it cannot be precluded that genetic effects play a role in the low IQ performance of Africans, we view environmental circumstances as potentially more relevant to the present-day difference in mean. The average IQ level of 81 for Africans in terms of western norms may appear to be low, but from a historical perspective it is not. For instance, due to the Flynn Effect, the average IQ of the Dutch population in the 1950s, compared to contemporary norms, would also be around 80 (Flynn, 1987, 2007). Note that in terms of societal development, contemporary African countries are more similar to developed countries in the first half of the twentieth century than to present-day developed countries.

There are very few studies that have addressed the psychometric meaning of the IQ performance of Africans on the Culture Fair Intelligence Test, DAM, K-ABC, and the Wechsler tests. In none of the samples in our review was measurement invariance tested with contemporary methods of item response theory or multi-group confirmatory factor analysis and found to be tenable. For IQ tests that were subject to a rigorous test of measurement invariance, measurement bias was found to be quite severe (Claassen et al., 2001; Dolan et al., 2004; Nenty & Dinero, 1981; Owen, 1989). In the absence of studies of differential item functioning in the Wechsler scales, the DAM test, and most of the other tests in Table 2, it is unclear whether these scores are measurement invariant. Accordingly, the degree to which measurement bias has affected IQ levels in African samples is unclear. We consider the conclusion that the average IQ of people in sub-Saharan Africa is lower than average IQ of people in western countries to be tenable, but the degree to which these low scores reflect lower levels of *g* is unknown. These low scores might not reflect an accurate or valid assessment of *g*, but future studies are needed to address this issue. For instance, studies like those by Fagan and Holland (2007) can be conducted in order to shed light on the meaning of IQ test scores of Africans. In addition, modern psychometric techniques should be applied to (existing) data sets in order to study the true nature of differences in IQ test performance between African and western test takers.

Appendix A

Table A1
Cumulative meta-analysis of samples that meet our five inclusion criteria.

Study	Cumul. *N*	Mean	SE	Lower	Upper
Lloyd and Pidgeon (1961)	275	84.30	0.47	83.38	85.22
Bardet et al. (1960)	817	75.30	9.00	57.66	92.94
Bakare (1972)	1210	77.90	5.88	66.38	89.42
Minde and Kantor (1976)	1724	79.40	3.85	70.32	88.47
Buj (1981)	1949	78.66	3.67	71.11	86.21
Ohuche and Ohuche (1973)	2151	80.76	3.30	74.57	87.95
Skuy et al. (2001)	2303	79.84	3.07	73.38	86.30
Skuy et al. (2001)	2403	78.83	2.88	72.82	84.85
Wilson et al. (1991)	2455	79.63	2.79	73.98	85.28
Skuy et al. (2000)	2476	81.23	2.67	75.77	86.71
Klein et al. (2007)	2504	81.93	2.56	76.70	87.16
Shuttleworth Edwards et al. (2004)	2544	82.62	2.56	77.60	87.65

Note: Samples ordered from smallest to largest standard error of sample's mean.

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002

References

Akande, A. (2000). Order effects on neuropsychological test performance of normal, learning disabled and low functioning children: A cross-cultural study. *Early Child Development and Care*, *165*, 145−161.

Alade, E. B. (1992). Determining intelligence coefficients of Nigeria deaf children with the Wechsler Intelligence Scale for Children—Revised. *Early Child Development and Care*, *80*, 103−107.

Ani, C. C., & Grantham-McGregor, S. (1998). Family and personal characteristics of aggressive Nigerian boys. *Journal of Adolescent Health*, *23*, 311−317.

Ashem, B., & Janes, M. D. (1978). Deleterious effects of chronic undernutrition on cognitive abilities. *Journal of Child Psychology and Psychiatry*, *19*, 23−31.

Ashton, M. C., & Lee, K. (2005). Problems with the method of correlated vectors. *Intelligence*, *33*, 431−444.

Avenant, T. J. (1988). The establishment of an individual intelligence scale for adult South Africans. *Report on an exploratory study conducted with the WAIS-R on a sample of Blacks. (No. P-91). Pretoria, South Africa: Human Sciences Research Council. Document Number)*.

Badri, M. B. (1965). Influence of modernization on Goodenough quotients of Sudanese children. *Perceptual and Motor Skills*, *20*, 931−932.

Badri, M. B. (1965). The use of finger drawing in measuring the Goodenough quotient of culturally deprived Sudanese children. *Journal of Psychology*, *59*, 333−334.

Bakare, C. G. M. (1972). Social-class differences in the performance of Nigerian children on the Draw-a-Man test. In L. J. Cronbach, & P. J. Drenth (Eds.), *Mental tests and cultural adaptation* (pp. 355−363). The Netherlands: Mouton: The Hague.

Barber, N. (2005). Educational and ecological correlates of IQ: A cross-national investigation. *Intelligence*, *33*, 273−284.

Bardet, C., Moreigne, F., & Sénécal, J. (1960). Application de test de Goodenough à des écoliers africains de 7 à 14 ans [Application of the Goodenough test to African school children ages 7 to 14]. *Enfance*, 199−208.

Barnett, S. M., & Williams, W. M. (2004). National intelligence and the emperor's new clothes. *Contemporary Psychology*, *49*, 389−396.

Bartholomew, D. J. (2004). *Measuring intelligence. Facts and fallacies.* Cambridge, UK: Cambridge University Press.

Berry, J. W. (1974). Radical cultural relativism and the concept of intelligence. In J. W. Berry, & P. R. Dasen (Eds.), *Culture and cognition: Readings in cross-cultural psychology* (pp. 225−229). London: Methuen & Co Ltd.

Biesheuvel, S. (1943). *African intelligence.* Johannesburg, South Africa: South African Institute of Race Relations.

Boivin, M. J. (2002). Effects of early cerebral malaria on cognitive ability in Senegalese children. *Journal of Developmental and Behavioral Pediatrics*, *23*, 353−364.

Boivin, M. J., & Giordani, B. (1993). Improvements in cognitive performance for schoolchildren in Zaire, Africa, following an iron supplement and treatment for intestinal parasites. *Journal of Pediatric Psychology*, *18*, 249−264.

Boivin, M. J., Giordani, B., & Bornefeld, B. (1995). Use of the tactual performance test for cognitive ability testing with African children. *Neuropsychology*, *9*, 409−417.

Boivin, M. J., Giordani, B., Ndanga, K., Maky, M. M., Manzeki, K. M., Ngugu, N., & Muamba, K. (1993). Effects of treatment for intestinal parasites and malaria on the cognitive abilities of schoolchildren in Zaire, Africa. *Health Psychology*, *12*, 220−226.

Buj, V. (1981). Average IQ values in various European countries. *Personality and Individual Differences*, *2*, 168−169.

Buros, O. K. (1959). *The fifth mental measurements yearbook.* Highland Park, NJ: Gryphon.

Carlson, J. S. (1970). A note on the relationship between the Draw-a-Man test, the progressive matrices, and conservation. *Journal of Psychology*, *74*, 231−235.

Cattell, R. B., & Cattell, A. K. S. (1973). *Culture Fair Intelligence Test Scale 2.* Chicago, IL: Institute for Personality and Ability Testing.

Claassen, N. C. W., Krynauw, A. H., Paterson, H., & Mathe, M. W. G. (2001). *A standardisation of the WAIS-III for English-speaking South Africans.* Pretoria, South Africa: Human Sciences Research Council.

Cooper, H. (1998). *Synthesizing research. A guide for literature review*, (3rd ed.) London, UK: Sage Publications.

Cooper, H. (2003). Editorial. *Psychological Bulletin*, *129*, 3−9.

Crawford Nutt, D. H. (1976). Are Black scores on Raven's Standard Progressive Matrices an artifact of method of test presentation? *Psychologia Africana*, *16*, 201−206.

Dent, G. R. (1937). The educability of the Bantu. In E. G. Malherbe (Ed.), *Educational adaptations in a changing society* Capetown, South Africa: Juta & Co., Ltd.

Dickerson, R. E. (2006). Exponential correlation of IQ and the wealth of nations. *Intelligence*, *34*, 291−295.

Dolan, C. V. (2000). Investigating Spearman's hypothesis by means of multi-group confirmatory factor analysis. *Multivariate Behavioral Research*, *35*, 21−50.

Dolan, C. V., Roorda, W., & Wicherts, J. M. (2004). Two failures of Spearman's hypothesis: The GAT-B in Holland and the JAT in South Africa. *Intelligence*, *32*, 155−173.

Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 127−144). New York: Wiley.

Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455−463.

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal*, *315*, 629−634.

Ervik, A. O. (2003). Book review: IQ and the wealth of nations. *The Economic Journal*, *113*, 406−408.

Fagan, J. F., & Holland, C. R. (2007). Racial equality in intelligence: Predictions from a theory of intelligence as processing. *Intelligence*, *35*, 319−334.

Fahmy, M. (1964). Initial exploring of the intelligence of Shilluk children: Studies in the Southern Sudan. *Vita Humana*, *7*, 164−177.

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

Ferguson, G. A. (1966). *Statistical analysis in psychology and education*, (2nd ed.) New York: McGraw-Hill.

Fernández-Ballesteros, R., Juan Espinosa, M., Colom, R., & Calero, M. D. (1997). Contextual and personal sources of individual differences in intelligence: Empirical results. In J. S. Carlson, J. Kingma, & W. Tomic (Eds.), *Advances in cognition and educational practice: Reflections on the concept of intelligence*, *Vol. 4*. (pp. 221−274)London, England: JAI Press Inc.

Ferron, O. (1965). The test performance of "coloured" children. *Educational Research*, *8*, 42−57.

Fick, M. L. (1929). Intelligence test results of poor white, native (Zulu), coloured, and Indian school children and the educational and social implications. *South African Journal of Science*, *16*, 904−920.

Flynn, J. R. (1987). Massive IQ gains in 14 nations: What IQ tests really measure. *Psychological Bulletin*, *101*, 171−191.

Flynn, J. R. (2007). *What is intelligence? Beyond the Flynn Effect.* Cambridge, UK: Cambridge University Press.

Gelade, G. A. (2008). IQ, cultural values, and the technological achievement of nations. *Intelligence*, *36*, 711−718.

Gelade, G. A. (2008). The geography of IQ. *Intelligence*, *36*, 495−501.

Giordani, B., Boivin, M. J., Opel, B., Dia Nseyila, D. N., & Lauer, R. E. (1996). Use of the K-ABC with children in Zaire, Africa: An evaluation of the sequential-simultaneous processing distinction within an intercultural context. *International Journal of Disability, Development and Education*, *43*, 5−24.

Goodenough, F. L. (1926). *Measurement of intelligence by drawings.* Chicago, IL: World Book Company.

Goodenough, F. L., & Harris, D. B. (1950). Studies in the psychology of children's drawings II. 1928–1949. *Psychological Bulletin*, *47*, 360−433.

Gottfredson, L. S. (2005). What if the hereditarian hypothesis were true? *Psychology, Public Policy, and Law*, *11*, 311−319.

Hakstian, A. R., & Vandenberg, S. G. (1979). The cross-cultural generalizability of a higher-order cognitive structure model. *Intelligence*, *3*, 73−103.

Hanushek, E. A., & Kimko, D. D. (2000). Schooling, labor-force quality, and the growth of nations. *American Economic Review*, *90*, 1184−1208.

Harris, D. B. (1963). *Children's drawings as measures of intellectual maturity.* New York, NY: Harcourt, Brace & World, Inc.

Herrnstein, R. J., & Murray, C. A. (1994). *The Bell curve: Intelligence and class structure in American life.* New York, NY: Free Press.

Higgins, J. P. T., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *British Medical Journal*, *327*, 557−560.

Holding, P. A., Taylor, H. G., Kazungu, S. D., Mkala, T., Gona, J., Mwamuye, B., et al. (2004). Assessing cognitive outcomes in a rural African population: Development of a neuropsychological battery in Kilifi District, Kenya. *Journal of the International Neuropsychological Society*, *10*, 246−260.

Howard Scott, L. (1981). Measuring intelligence with the Goodenough–Harris Drawing test. *Psychological Bulletin*, *89*, 483−505.

Hunkin, V. (1950). Validation of the Goodenough Draw-a-man Test for African children. *Journal for Social Research*, *1*, 52−63.

Hunt, E. B., & Carlson, J. S. (2007). Considerations relating to the study of group differences in intelligence. *Perspectives on Psychological Science*, *2*, 194−213.

Hunt, E. B., & Sternberg, R. J. (2006). Sorry, wrong numbers: An analysis of a study of a correlation between skin color and IQ. *Intelligence*, *34*, 131−137.

Hunt, E. B., & Wittmann, W. (2008). National intelligence and national prosperity. *Intelligence*, *36*, 1−9.

Jensen, A. R. (1980). *Bias in mental testing.* London: Methuen & Co., Ltd.

Jensen, A. R. (1982). The debunking of scientific fossils and straw persons. *Contemporary Education Review*, *1*, 121−135.

Jones, G., & Schneider, W. J. (2006). Intelligence, human capital, and economic growth: A Bayesian Averaging of Classical Estimates (BACE) approach. *Journal of Economic Growth*, *11*, 71−93.

Kanazawa, S. (2004). General intelligence as a domain-specific adaptation. *Psychological Review*, *111*, 512−523.

Kanazawa, S. (2006). Mind the gap…in intelligence: Re-examining the relationship between inequality and health. *British Journal of Health Psychology*, *11*, 623−642.

Kanazawa, S. (2008). Temperature and evolutionary novelty as forces behind the evolution of general intelligence. *Intelligence*, *36*, 99−108.

Kashala, E., Elgen, I., Sommerfeldt, K., Tylleskar, T., & Lundervold, A. (2005). Cognition in African children with attention-deficit hyperactivity disorder. *Pediatric Neurology*, *33*, 357−364.

Kaufman, A. S., & Kaufman, N. L. (1983). *Kaufman Assessment Battery for Children: Administration and scoring manual.* Circle Pines, MN: American Guidance Service.

Khaleefa, O., Abdelwahid, S. B., Abdulradi, F., & Lynn, R. (2008). The increase of intelligence in Sudan 1964–2006. *Personality and Individual Differences*, *45*, 412−413.

Kirkcaldy, B., Furnham, A., & Siefen, G. (2004). The relationship between health efficacy, educational attainment and well-being among 30 nations. *European Psychologist*, *9*, 107−119.

Klein, O., Pohl, S., & Ndagijimana, C. (2007). The Influence of intergroup comparisons on Africans' intelligence test performance in a job selection context. *The Journal of Psychology*, *141*, 453−467.

Kraemer, H. C., Gardner, C., Brooks, J. O., & Yesavage, J. A. (1998). Advantages of excluding underpowered studies in meta-analysis: Inclusionist versus exclusionist viewpoints. *Psychological Methods*, *3*, 23−31.

Lane, C. (1994, December 1). *Tainted sources. The New York Review of Books.*

Light, R. J., & Pillemer, D. B. (1984). *Summing up. The science of reviewing research.* Cambridge, MA: Harvard University Press.

Lloyd, F., & Pidgeon, D. A. (1961). An investigation into the effects of coaching on non-verbal test material with European, Indian and African children. *British Journal of Psychology*, *31*, 145−151.

Loehlin, J. C. (2007). Book review: R. Lynn Race differences in intelligence. *Intelligence*, *35*, 93−94.

Lubke, G. H., Dolan, C. V., & Kelderman, H. (2001). Investigating group differences on cognitive tests using Spearman's hypothesis: An evaluation of Jensen's method. *Multivariate Behavioral Research*, *36*, 299−324.

Lynn, R. (1991). Race differences in intelligence: A global perspective. *Mankind Quarterly*, *31*, 255−296.

Lynn, R. (2003). The geography of intelligence. In H. Nyborg (Ed.), *The scientific study of general intelligence: Tribute to Arthur R. Jensen* (pp. 127−146). Oxford, UK: Elsevier Science Ltd.

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

Lynn, R., Harvey, J., & Nyborg, H. (2009). Average intelligence predicts atheism rates across 137 countries. *Intelligence*, *37*, 11−15.

Lynn, R., Meisenberg, G., Mikk, J., & Williams, A. (2007). National IQs predict differences in scholastic achievement in 67 countries. *Journal of Biosocial Science*, *39*, 861−874.

Lynn, R., & Mikk, J. (2007). National differences in intelligence and educational attainment. *Intelligence*, *35*, 115−121.

Lynn, R., & Mikk, J. (2009). National IQs predict educational attainment in math, reading and science across 56 nations. *Intelligence*, *37*, 305−310.

Lynn, R., & Owen, K. (1994). Spearman's hypothesis and test score differences between Whites, Indians, and Blacks in South Africa. *Journal of General Psychology*, *121*, 27−36.

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

Lynn, R., & Vanhanen, T. (2006). *IQ and global inequality.* Augusta, GA: Washington Summit Publishers.

Maqsud, M. (1980). Personality and academic attainment of primary school children. *Psychological Reports*, *46*, 1271−1275.

Maqsud, M. (1980). Relationship between sociometric status and moral judgment in secondary school girls. *West African Journal of Educational and Vocational Measurement*, *5*, 13−17.

McDaniel, M. A. (2008). Book review. R. Lynn & T. Vanhanen (2006). IQ and global inequality. *Intelligence*, *36*, 731−732.

Meisenberg, G. (2004). Talent, character, and the dimensions of national culture. *Mankind Quarterly*, *45*, 123−168.

Mellenbergh, G. J. (1989). Item bias and item response theory. *International Journal of Educational Research*, *13*, 127−143.

Millsap, R. E., & Everson, H. T. (1993). Methodology review: Statistical approaches for assessing measurement bias. *Applied Psychological Measurement*, *17*, 297−334.

Minde, K., & Kantor, S. (1976). Instructing Ugandan primary schoolchildren in the execution of an "intelligence" test: A controlled evaluation. *Journal of Cross Cultural Psychology*, *7*, 209−222.

Morse, S. (2006). Making development simple. The genetic deterministic hypothesis for economic development. *Ecological Economics*, *56*, 79−88.

Munroe, R. L., & Munroe, R. H. (1983). Drawings and values in three East African societies. *Journal of Social Psychology*, *119*, 135−136.

Nell, V. (2000). *Cross-cultural neuropsychological assessment: Theory and practice.* Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Nenty, H. J., & Dinero, T. E. (1981). A cross-cultural analysis of the fairness of the Cattell Culture Fair Intelligence Test using the Rasch model. *Applied Psychological Measurement*, *5*, 355−368.

Nissen, H. W., Machover, S., & Kinder, E. F. (1935). A study of performance tests given to a group of native African Negro children. *British Journal of Psychology*, *25*, 308−355.

Nwanze, H. O., & Okeowo, P. A. (1980). The usefulness of a developmental profile in predicting reading retardation in Nigerian children. *West African Journal of Educational and Vocational Measurement*, *5*, 50−59.

Ohuche, N. M., & Ohuche, R. O. (1973). The Draw-A-Man Test as a predictor of academic achievement. *West African Journal of Educational and Vocational Measurement*, *1*, 20−27.

Okunrotifa, P. O. (1976). A comparison of the entry behaviours of Nigerian rural and urban children in geography. *West-African Journal of Educational and Vocational Measurement*, *3*, 1−6.

Owen, K. (1989). *Test and item bias: The suitability of the Junior Aptitude tests as a common battery for White, Indian, and Black students in Standard 7. Pretoria.* South Africa: Human Sciences Research Council Document Number.

Pons, A .L. (1974). Administration of tests outside cultures of their origin. Paper presented at the 26th Annual congress of the South African Psychological Association, Johannesburg, South Africa.

Ram, R. (2007). IQ and economic growth: Further augmentation of Mankiw–Romer–Weil model. *Economics Letters*, *94*, 7−11.

Raven, J. C. (1956). *Guide to using the Coloured Progressive Matrices.* London: H.K. Lewis & Co. Ltd.

Raven, J. C. (1960). *Guide to the Standard Progressive Matrices.* London: H.K. Lewis & Co. Ltd.

Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's Progressive Matrices and Vocabulary Scales.* Oxford, UK: Oxford Psychologists Press.

Richter, L. M., Griesel, R. D., & Wortley, M. E. (1989). The Draw-a-Man test: A 50-year perspective on drawings done by black South African children. *South African Journal of Psychology*, *19*, 1−5.

Rindermann, H. (2006). Was messen internationale Schulleistungsstudien? Schulleistungen, Schülerfähigkeiten, kognitive Fähigkeiten, Wissen oder allgemeine Intelligenz? [What do international student assessment studies measure? School performance, student abilities, cognitive abilities, knowledge or general intelligence?]. *Psychologische Runschau*, *57*, 69−86.

Rindermann, H. (2007). The g factor of international cognitive ability comparisons: The homogeneity of results in PISA, PIRLS and IQ tests across nations. *European Journal of Personality*, *21*, 667−706.

Rindermann, H. (2008). Relevance of education and intelligence at the national level for the economic welfare of people. *Intelligence*, *36*, 127−142.

Rindermann, H. (2008). Relevance of education and intelligence for the political development of nations: Democracy, rule of law and political liberty. *Intelligence*, *36*, 306−322.

Rindermann, H., & Meisenberg, G. (in press). Relevance of education and intelligence at the national level for health: The case of HIV and AIDS. *Intelligence*. doi:10.1016/j.intell.2009.03.005.

Rosenthal, R., & Rosnow, R. L. (1975). *The volunteer subject.* New York: Wiley.

Rushton, J. P. (2000). *Race, evolution, and behavior. A life history perspective.* 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. (2003). The bigger Bell curve: Intelligence, national achievement, and the global economy. *Personality and Individual Differences*, *34*, 367−372.

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

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

Rushton, J. P., & Templer, D. I. (in press). National differences in intelligence, crime, income, and skin color. *Intelligence*. doi:10.1016/j.intell.2009.04.003.

Serpell, R. (1979). How specific are perceptual skills? A cross-cultural study of pattern reproduction. *British Journal of Psychology*, *70*, 365−380.

Shatz, S. M. (2008). IQ and fertility: A cross-national study. *Intelligence*, *36*, 109−111.

Shuttleworth Edwards, A. B., Kemp, R. D., Rust, A. L., Muirhead, J. G. L., Hartman, N. P., & Radloff, S. E. (2004). Cross-cultural effects on IQ test performance: A review and preliminary normative indications on WAIS-III test performance. *Journal of Clinical and Experimental Neuropsychology*, *26*, 903−920.

Skuy, M., Schutte, E., Fridjhon, P., & O'Carroll, S. (2001). Suitability of published neuropsychological test norms for urban African secondary school students in South Africa. *Personality and Individual Differences*, *30*, 1413−1425.

Skuy, M., Taylor, M., O'Carroll, S., Fridjhon, P., & Rosenthal, L. (2000). Performance of Black and White South African children on the Wechsler Intelligence Scale for Children-Revised and the Kaufman Assessment Battery. *Psychological Reports*, *86*, 727—737.

Steele, C. M. (1997). A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist*, *52*, 613—629.

Sternberg, R. J., Grigorenko, E. L., Ngorosho, D., Tantufuye, E., Mbise, A., Nokes, C., et al. (2002). Assessing intellectual potential in rural Tanzanian school children. *Intelligence*, *30*, 141—162.

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

Sterne, J. A. C., & Egger, M. (2005). Regression methods to detect publication and other bias in meta-analysis. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 99—110). New York: Wiley.

Templer, D. I. (2008). Correlational and factor analytic support for Rushton's differential K life history theory. *Personality and Individual Differences*, *45*, 440—444.

Templer, D. I., & Arikawa, H. (2006). Association of race and color with mean IQ across nations. *Psychological Reports*, *99*, 191—196.

Templer, D. I., & Arikawa, H. (2006). Temperature, skin color, per capita income, and IQ: An international perspective. *Intelligence*, *34*, 121—139.

Thurstone, T. G. (1963). *Primary mental abilities for grades 4–6.* Chicago, IL: Science Research Associates.

Vernon, P. E. (1969). *Intelligence and cultural environment.* London, UK: Methuen & Co. Ltd.

Voracek, M. (2004). National intelligence and suicide rate: an ecological study of 85 countries. *Personality and Individual Differences*, *37*, 543—553.

Wechsler, D. (1974). *WISC-R Manual.* New York, NY: The Psychological Corporation.

Wechsler, D. (1981). *WAIS-R Manual.* New York, NY: The Psychological Corporation.

Weede, E., & Kampf, S. (2002). The impact of intelligence and institutional improvements on economic growth. *Kyklos*, *55*, 361—380.

Whetzel, D. L., & McDaniel, M. A. (2006). Prediction of national wealth. *Intelligence*, *34*, 449—458.

Wicherts, J. M., Dolan, C. V., Carlson, J. S., & van der Maas, H. L. J. (2009). Raven's test performance of Africans: Average performance, psychometric properties, and the Flynn Effect. Unpublished manuscript. University of Amsterdam.

Williams, J. H. (1935). Validity and reliability of the Goodenough intelligence test. *School and Society*, *41*, 653—656.

Wilson, D., Mundy-Castle, A., & Sibanda, P. (1991). Cognitive effects of LOGO and computer-aided instruction among Black and White Zimbabwean primary school girls. *Journal of Social Psychology*, *131*, 107—116.

Wober, M. (1969). Meaning and stability of Raven's Matrices test among Africans. *International Journal of Psychology*, *4*, 229—235.

Woodley, M. A. (2009). Inbreeding depression and IQ in a study of 72 countries. *Intelligence*, *37*, 268—276.

Yoloye, E. A. (1971). The effect of schooling on the performance of bilingual students in tests of intelligence. *Research in Education*, *5*, 25—34.

Zindi, F. (1994). Differences in psychometric performance. *The Psychologist*, *7*, 549—552.

Zindi, F. (1994). Towards the standardization of the WISC-R for early childhood assessment in Zimbabwe. *IFE Psychologia: An International Journal*, *2*, 19—32.

Please cite this article as: Wicherts, J. M., et al., A systematic literature review of the average IQ of sub-Saharan Africans, *Intelligence* (2009), doi:10.1016/j.intell.2009.05.002