Intelligence 38 (2010) 30–37

Contents lists available at ScienceDirect

Intelligence

# The dangers of unsystematic selection methods and the representativeness of 46 samples of African test-takers

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

*University of Amsterdam, The Netherlands*

ARTICLE INFO

*Article history:*
Received 13 October 2009
Accepted 6 November 2009
Available online 3 December 2009

*Keywords:*
Systematic literature review
National IQ
Group differences in IQ

ABSTRACT

In this rejoinder, we criticize Lynn and Meisenberg's (this issue) methods to estimate the average IQ (in terms of British norms after correction of the Flynn Effect) of the Black population of sub-Saharan Africa. We argue that their review of the literature is unsystematic, as it involves the inconsistent use of rules to determine the representativeness and hence selection of samples. Employing independent raters, we determined of each sample whether it was (1) considered representative by the original authors, (2) drawn randomly, (3) based on an explicated stratification scheme, (4) composed of healthy test-takers, and (5) considered by the original authors as normal in terms of Socio-Economic Status (SES). We show that the use of these alternative inclusion criteria would not have affected our results. We found that Lynn and Meisenberg's assessment of the samples' representativeness is not associated with any of the objective sampling characteristics, but rather with the average IQ in the sample. This suggests that Lynn and Meisenberg excluded samples of Africans who average IQs above 75 because they deemed these samples unrepresentative on the basis of the samples' relatively high IQs. We conclude that Lynn and Meisenberg's unsystematic methods are questionable and their results untrustworthy.



## 1. Introduction

In our previous paper (this issue), we estimated the average IQ, in terms of British norms after correction for the Flynn Effect, of the normal healthy Black population of sub-Saharan Africa (henceforth Africans) on the basis of systematic literature review. We arrived at an estimate close to 80, which replicated the estimate in another systematic review on the basis of the Raven's tests (Wicherts, Dolan, Carlson & van der Maas, 2009). Our estimate of the mean IQ of Africans stands in stark contrast with the estimate of 68 on the basis of the same corrected norms in the books of Lynn (2006) and Lynn and Vanhanen (2002, 2006). We pointed out that the reviews by Lynn and Lynn and Vanhanen were problematic because they did not employ systematic methodology developed for literature reviews (Cooper, 1998). These books contained neither an explication of inclusion and exclusion criteria, nor a list of excluded studies with reasons for exclusion. As Lynn and Meisenberg's response (this issue) makes abundantly clear, their estimate of the average IQ of Africans was not based on all the available data, but rather on a selection of samples that they deemed sufficiently representative. In this rejoinder, we argue that Lynn and Meisenberg's work suffers from much of the same problems as the previous work by Lynn (2006) and Lynn and Vanhanen (2002, 2006), in the sense that its methodology is unsystematic. Specifically, we show that Lynn and Meisenberg's assessment of the representativeness of samples is not a function of objective sampling characteristics, but rather of the mean IQ in the sample. This renders their methodology questionable and their estimate of the mean IQ of Africans untrustworthy.

* Corresponding author. Department of Psychology, Psychological Methods, University of Amsterdam, Roetersstraat 15, 1018 WB Amsterdam, The Netherlands. Tel.: +31 205257067; fax: +31 206390026.
*E-mail address:* J.M.Wicherts@uva.nl (J.M. Wicherts).

0160-2896/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.intell.2009.11.003

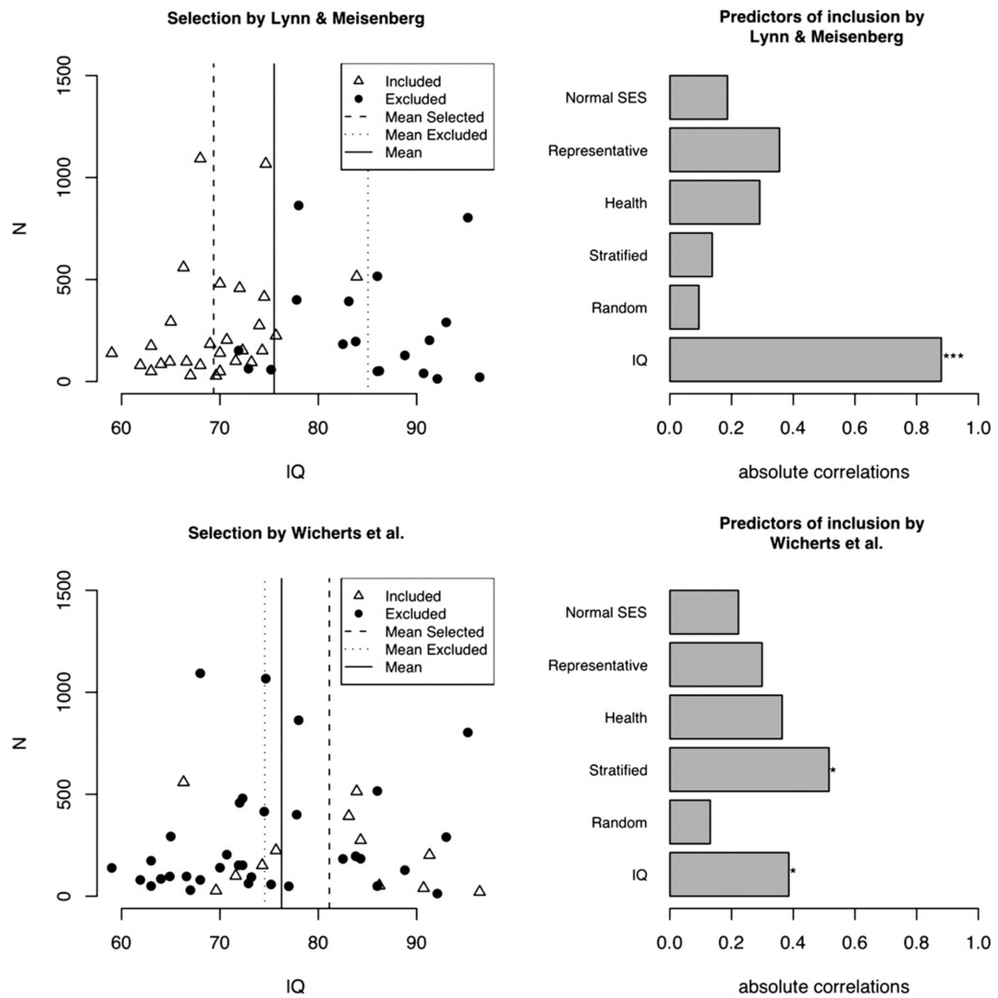


**Fig. 1.** Plot of mean IQs against samples sizes (left) for Lynn and Meisenberg's and Wicherts et al.'s selection of studies and prediction of inclusion of studies (right) by Lynn and Meisenberg and Wicherts et al. as predicted from mean IQs and the five criteria of representativeness.

## 2. The full database

Lynn and Meisenberg carefully checked our computations of the average IQs of the samples and noted some minor discrepancies with their own computations. The inter-rater reliability between our previous computations and Lynn's (2006) was poor at $r = .71$, but after correspondence with Dr. Lynn, we have resolved all but two of the discrepancies (Fahmy, 1964; Lloyd & Pidgeon, 1961). Apart from these two cases, the correlation between the results of our computations and the results of Lynn and Meisenberg now stands at .99.[1] Clearly, the difference between our and their results does not arise from any difference in the computed mean IQs, but rather from different selections of the available data. We used five criteria to determine the appropriateness of the available samples, while Lynn and Meisenberg stated to have used only one. On the basis of their assessment of the representativeness of samples, Lynn and Meisenberg selected 26 samples of our 44 samples, added two new samples, and concluded on the basis of these 28 unique[2] samples that the average IQ of Africans equals 69. As can be seen in the left-hand side of Fig. 1, our estimate was quite close to the average on the basis of all samples (i.e., 77), while Lynn and Meisenberg's estimate was considerably lower.

Fig. 2 depicts the results based on all the available data (including data from the Raven's tests) from Black sub-Saharan Africans who took cognitive ability tests that enable

[1] These two discrepancies are discussed below. Other minor discrepancies are due to our use of $N$ weighting in combining IQs from subsamples (Ferron, 1965) or the use of the mean standardized scores in another sample (Murdoch et al., 1994).

[2] Lynn and Meisenberg discussed two samples (Conant et al., 1999; Giordani, Boivin, Opel, Dia Nseyila & Lauer, 1996) that reported data from other studies by Boivin et al. We excluded data from Giordani et al., 1996 as it was composed of data from two other samples (Boivin et al., 1993; Boivin et al., 1995). Boivin et al. (M. Boivin, personal communication, June 9, 2006) have collected data from three unique samples in D.R. Congo (Boivin & Giordani, 1993; Boivin, Giordani & Bornefeld, 1995; Boivin et al., 1993). Note that we inadvertently referred to (Boivin & Giordani, 1993) as the source of another sample (Boivin et al., 1993). In the current analyses, we refer to the source of the third sample as: Boivin and Giordani (1993) and Conant et al. (1999). We simply used Lynn and Meisenberg's mean IQ estimate on the basis of the combined sample.

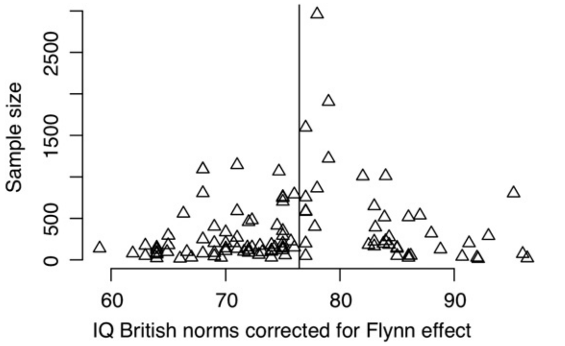


**Fig. 2.** Mean IQs of 109 African samples on the basis of British norms after correction for the Flynn Effect.

the computation of a mean IQ in terms of British norms after correction for the Flynn Effect. The full database[3] includes 109 samples totaling 37,811 test-takers. The *N*-weighted mean of these studies is 76.5 ($SD = 6.7$), and the median is 77.0. To arrive at a mean IQ close to Lynn and Meisenberg's estimate of 69, the majority of the data would have to be rejected as insufficiently representative or otherwise flawed. For instance, suppose that the 57 samples with average IQs above 74.5 were all unrepresentative, their exclusion from the database would result in an *N*-weighted or median average IQ of 69. However, this would amount to the exclusion of over 25,000 cases, or two-thirds of the available data.

## 3. Inconsistent rules to determine representativeness

Lynn and Meisenberg assessed the representativeness of our 44 samples, and excluded 18 as being insufficiently representative. In a systematic literature review, rules for inclusion and exclusion should be explicit and stated in advance, and these rules should be applied consistently across the board.[4] This enables replication of the selection of studies by independent raters (Cooper, 1998). Lynn and Meisenberg did not explicate the rules they employed in assessing representativeness. More importantly, as the following seven examples illustrate, they used their criteria to exclude samples inconsistently.

First, they rejected as unrepresentative data from the standardization of WAIS-III among English-speaking South Africans (Claassen, Krynauw, Paterson & Mathe, 2001), because it was "a well-educated sample". At the same time, Lynn and Meisenberg accepted as sufficiently representative, a sample of Africans (Klein, Pohl & Ndagijimana, 2007), who all "had finished high school [,] had studied for 1–2 years beyond high school [or] had studied 3 years or more beyond high school" (Klein et al., 2007, p. 458). We note that the excluded sample (average IQ well above 80) was carefully drawn on the basis of stratified procedures, while the Klein sample (average IQ of 70) was a convenience sample, consisting of immigrants to Belgium.

Second, we excluded a sample of Senegalese children (average IQ of 75; Boivin, 2002), because the IQ test in question had been altered and because the children had suffered from malaria, and so could be considered to be unhealthy. Although Lynn and Meisenberg agreed with us and excluded this sample, they included a sample (average IQ of 59) that took the same altered IQ test, and was composed of children suffering from malaria at the time of assessment (Holding et al., 2004).

Third, Lynn and Meisenberg rejected a particularly large sample of Nigerians ($N = 803$; average IQ of 95; Nenty & Dinero, 1981) because official IQs could not be determined in this study. However, this study includes a comparable sample from the US that enables the computation of the average IQ of the Nigerian sample. This computational method failed to meet our own criteria, but it was applied by Lynn (2006) on numerous occasions and by Lynn and Meisenberg in their computation (average IQ of 74) of the mean IQ of one sample (Lloyd & Pidgeon, 1961), despite the availability of more accurate official British norms (average IQ of 84).

Fourth, another sample (average IQ of 96) was considered unrepresentative by Lynn and Meisenberg because the children were from a "fee-paying school in Nigeria [which] is evidently an elite sample". Two other samples (average IQs of 59 and 72) were also drawn from fee-paying schools (Conant et al., 1999; Sternberg et al., 2002), but these samples were not excluded by Lynn and Meisenberg.

Fifth, Lynn and Meisenberg rejected a sample (average IQ of 91) in Ohuche and Ohuche (1973) as problematic in part because "the ages of the children are unknown". At the same time, Lynn and Meisenberg included samples (average IQs between 63 and 72) of which age information is lacking (Dent, 1937; Fahmy, 1964; Fahrmeier, 1975; Fick, 1929; Nissen, Machover & Kinder, 1935).

Sixth, Lynn and Meisenberg excluded data (Kashala, Elgen, Sommerfeldt, Tylleskar & Lundervold, 2005) from the Digit Span test (average IQ of 94), because it "correlates poorly with the Wechsler full scale IQ". Similar problems were established with respect to the Koh's blocks test in an African sample (Vernon, 1969), but Lynn and Meisenberg did not see this as a reason to exclude data from this particular test (Dent, 1937; average IQ 68). Likewise, they excluded data in one sample (average IQ of 91), because of a weak "correlation between IQs [with] tests of English, math and social science in grades 4–7, showing IQs have no validity for these ages." However, in numerous studies which they did include, test scores failed to correlate with other tests or criteria (Richter, Griesel & Wortley, 1989; Sternberg et al., 2002) average IQs 75 and 72, respectively). In fact, we documented a rather weak validity of the Raven's tests in African samples (Wicherts et al., 2009). However, despite its poor validity in many African settings, Lynn and Meisenberg state explicitly that the Raven's tests need to be considered to arrive at a good estimate of the average IQ of Africans.

Seventh, another reason for excluding data from one source (Ohuche & Ohuche, 1973) was that in this study, the

[3] In line with Lynn and Meisenberg, we excluded the samples that apparently included mainly non-Black Africans (Badri, 1965a; Khaleefa, Abdelwahid, Abdulradi & Lynn, 2008; Raveau, Elster & Lecoutre, 1976) and the samples of university students. We added two new samples that were located by Lynn and Meisenberg and another sample of 153 Black South-Africans (Wicherts & Dolan, submitted for publication). A full list of all studies, references, and computations is available upon request.

[4] Lynn and Meisenberg point out that we were inconsistent in our use of the Wisconsin Card Selection Test in estimating the IQ in two samples, but the WCST failed to meet our inclusion criteria and these data were only taken into account at the explicit request of one of our reviewers.

5–6 year olds showed an average IQ of 69.5, while the 7–12 year olds showed an average IQ of 94.2. According to Lynn and Meisenberg this discrepancy indicated “serious problems with the data”. However, a study involving Sudanese children (Fahmy, 1964) that was characterized by equally serious problems was included by Lynn and Meisenberg. Specifically, these Sudanese children lacked experience in using a pencil and averaged an IQ of 50 on the Draw a Man test. These same children averaged an IQ of 94 on a test that did not involve the use of a pencil. Lynn and Meisenberg did not consider the discrepancy between subtests' IQs problematic, and took the DAM scores into account in their mean estimate of this sample (average IQ of 69 according to Lynn and Meisenberg).

Thus, Lynn and Meisenberg applied numerous rules to determine the representativeness of samples, but they did so inconsistently. In the examples given above, the exclusion rule was applied only to samples that averaged relatively high IQs, but not to samples that averaged IQs near or below 70. This suggests that Lynn and Meisenberg's assessment of samples' representativeness is not independent of the mean IQ in the samples. We present the results of a statistical analysis that support this suggestion.

## 4. What is representative?

In order to determine the representativeness of samples, we judged the following five sampling characteristics as either

**Table 1**
Sampling characteristics of all samples from our previous review.

| Source | Random | Stratification | Healthy | Normal SES | Representative | Included by Lynn and Meisenberg | IQ British norms FE corrected |
|---|---|---|---|---|---|---|---|
| Akande (2000) | − | − | − | + | − | − | 72.9 |
| Ani and Grantham-McGregor (1998) | − | − | + | − | − | + | 73.2 |
| Ashem and Janes (1978) | − | + | + | + | + | − | 88.8 |
| Avenant (1988) | − | − | + | + | − | + | 70.0 [b] |
| Badri (1965b) | − | − | + | + | − | + | 61.9 |
| Bakare (1972) | − | + | + | + | − | − | 83.1 |
| Bardet, Moreigne and Sénécal (1960) | − | + | + | + | − | + | 66.3 |
| Boivin et al. (1993) | − | − | − | + | − | + | 66.6 |
| Boivin & Giordani, 1993; Conant et al., 1999 | + | − | − | − | − | + | 61/59 [a] |
| Boivin et al. (1995) | + | − | + | + | + | + | 64.9 |
| Boivin (2002) | − | + | − | + | − | − | 75.2 |
| Buj (1981) | − | + | + | + | + | +/− | 75.7 |
| Claassen et al. (2001) | − | + | + | + | + | − | 83.8 |
| Dent (1937) | − | + | + | + | − | + | 68.0 |
| Dunstan (1961, cited in Ferron, 1965) | − | − | + | + | − | − | 77.8 |
| Fahmy (1964) | − | + | + | + | + | + | 84.3/69 [a] |
| Fahrmeier (1975) | − | + | + | + | − | + | 72.3 |
| Ferron (1965) primary | − | − | + | + | − | + | 72.3/70 [a] |
| Ferron (1965) entrance | − | − | + | + | − | − | 78.0 |
| Ferron (1965) grammar | − | − | + | + | − | − | 92.5 |
| Fick (1929) | − | + | + | + | − | + | 65.0 |
| Haward and Roland (1954) | − | − | + | + | + | + | 67.0 |
| Holding et al. (2004) | − | − | − | + | − | + | 63.0 |
| Hunkin (1950) | − | − | + | + | + | + | 74.7 |
| Kashala et al. (2005) | + | − | − | − | + | − | 82.5 |
| Klein et al. (2007) | − | − | + | + | − | + | 69.6 |
| Lloyd and Pidgeon (1961) | − | + | + | + | + | + | 84.3/74 [a] |
| Lynn and Owen (1994) | − | − | + | + | + | + | 68.0 |
| Minde and Kantor (1976) | − | + | + | + | + | +/− | 83.9 |
| Murdoch et al. (1994) | − | − | + | + | − | + | 77/70 [a] |
| Nell (2000) | − | − | + | + | − | − | 71.9 |
| Nenty and Dinero (1981) | − | − | + | + | − | − | 95.2 |
| Nissen et al. (1935) | − | − | + | + | − | + | 63.0 |
| Nwanze and Okeowo (1980) | − | − | − | + | − | − | 92.1 |
| Ohuche and Ohuche (1973) | + | − | + | + | − | − | 91.3 |
| Richter et al. (1989) | − | − | + | + | − | + | 74.5 |
| Shuttleworth Edwards et al. (2004) | − | + | + | + | − | − | 90.7 |
| Skuy, Taylor, O'Carroll, Fridjhon and Rosenthal (2000) | − | − | − | − | − | − | 96.4 |
| Skuy, Schutte, Fridjhon and O'Carroll (2001) sample 1 | − | + | + | − | + | + | 71.6 |
| Skuy et al. (2001) sample 2 | + | − | + | + | + | + | 74.3 |
| Sternberg et al. (2001) | − | − | − | + | − | + | 64.0 |
| Sternberg et al. (2002) | − | − | + | − | − | + | 72.0 |
| Vernon (1969) | − | − | + | − | − | − | 86.0 |
| Wilson, Mundy-Castle and Sibanda (1991) | − | − | + | − | − | − | 86.2 |
| Yoloye (1971) | − | − | + | + | − | − | 86.0 |
| Zindi (1994) | + | + | + | + | − | + | 71.6 [b] |

[a] For samples in which the results of IQ computations by Lynn and Meisenberg were different from ours, we included both our (first entry) and their (second entry) IQ estimate.

[b] These computations were changed slightly after discussions with Dr. Lynn.

positive or negative. (1) Random selection: Was the sample drawn randomly on the basis of a targeted population? Random selection is widely accepted as the best method to achieve representativeness of a sample. (2) Stratification: Were one or more stratification variables considered in selecting the sample? Representative samples are often drawn by stratification with respect to particular demographic, socio-economic, educational, or geographic variables. (3) Health status: As our goal was to estimate the average IQs of normal, healthy test-takers, we determined whether the sample composed only of healthy test-takers. (4) Normal SES: We determined whether the sample was described as different than average in terms of Socio-Economic Status (SES) for a particular country or region. Often school fees may result in selection of above-average SES children, so whenever school fees were mentioned the SES variable was scored negatively. (5) Representativeness: We assessed whether the sample was described as representative of a particular population by the original authors.

The online supplementary material provides the original sample descriptions from all sources. Two PhD students specializing in psychological methods independently rated all samples on the basis of a codebook and these sample descriptions. The two raters worked independently, they were unaware of both our own assessment and Lynn and Meisenberg's, and they were blind to the average IQ of the sample. In addition, the first author (JMW) rated each sample on the same characteristics independently from the other raters, although he was neither blind to Lynn and Meisenberg's assessment nor to the mean IQ in the sample. Cohen's Kappa was computed to assess inter-rater reliabilities and the final assessment was based on the modus of the three ratings.

Cohen's Kappa indicated substantial agreement between the raters for stratification (average $k=.78$) and health status (average $k=.73$), almost perfect agreement for random selection (average $k=.83$) and representativeness (average $k=.96$), and moderate agreement for SES (average $k=.43$). The ratings of the 46 samples are given in Table 1, along with the mean IQ in the sample and a column indicating Lynn and Meisenberg's assessment of the representativeness of the samples (i.e., their inclusion criterion).

Table 2 contains the results of the mixed model regression analysis in which we predict the mean IQ in all 46 samples from the five criteria of representativeness. As can be seen, none of the criteria for representativeness was predictive of the mean IQ. This suggests that the use of these criteria as alternative inclusion criteria would have little effect on the overall estimate of the mean IQ of Africans, and hence that sampling bias could not have affected the results of our review. For instance, if we include only the 9 samples ($N=1975$) that meet at least four of five criteria for representativeness, the overall mean IQ of Africans was 79.7 ($SE=0.28$) in a fixed effects model and 78.8 ($SE=2.41$) in a random effects model.

Lynn and Meisenberg stated that they used the representativeness of samples as their main inclusion criterion. We computed tetrachoric correlations between their inclusion of studies and the five criteria for representativeness. However, as can be seen from the results in the right-hand side of Fig. 1, there is little to no correlation between whether or not Lynn and Meisenberg included a study and the five objective criteria of representativeness (all $ps>.15$). In a logistic regression of inclusion (yes/no) by Lynn and Meisenberg on the five criteria for representativeness, the explained variance was only 6.6% (Box and Snell's $R^2$). The omnibus test for the significance of the 5 criteria as predictors was not significant ($\chi^2(DF=5)=3.15$, $p=.68$). At the same time, the average IQ[5] was a statistically significant and strong negative predictor of inclusion; $\chi^2(DF=1)=37.6$, $p<.0001$. With a biserial correlation of −.88, the average IQ predicted Lynn and Meisenberg's inclusion or exclusion of samples quite well (explained variance 55.8%).

The top graph in Fig. 3 displays the logistic curve relating the probability of inclusion in our original work to the mean IQ in the samples (as the sole predictor). Also depicted are empirical percentages of inclusion in each 5-IQ-point interval. It is clear that the logistic function on the basis of means IQs hardly predicts our original inclusion of samples. Low IQ samples had a somewhat smaller probability of meeting our inclusion criteria due to our exclusion of unhealthy samples and samples in which the test administration was problematic. However, even for samples with relatively high IQ, the chances of the sample meeting all our original inclusion criteria remained around 50%. The bottom graph of Fig. 3 displays the results of the same logistic regression analysis for the inclusion of samples by Lynn and Meisenberg. The logistic curve is quite steep and showed near perfect fit to the empirical inclusion of samples within the intervals. Lynn and Meisenberg's assessment of the representativeness of samples was such that they included all samples with IQs below 70, while the chances of inclusion declined sharply with increasing average IQs. Lynn and Meisenberg rejected as unrepresentative all African samples with average IQs above 85.

**Table 2**
Meta-regression of average IQs on five criteria of representativeness ($k=46$).

| Predictor | Effect (*SE*) | Test results |
|---|---|---|
| Random sampling vs. no random sampling | −3.13 (4.40) | $Z=-0.71$, $p=.48$ |
| Stratification scheme vs. no stratification scheme | 1.49 (3.27) | $Z=0.46$, $p=.65$ |
| Healthy sample vs. unhealthy sample | 1.80 (3.77) | $Z=0.48$, $p=.63$ |
| Normal SES vs. described as abnormal SES | −3.78 (3.88) | $Z=-0.92$, $p=.33$ |
| Described as representative or not | 1.13 (3.40) | $Z=0.33$, $p=.74$ |

Notes: Effect is scored such that positive effects indicate higher mean IQs for first option. Our IQ computations in Table 1 were used for this analysis; Sampling weights computed as described in Wicherts et al. (in press); Maximum Likelihood estimation.

## 5. Conclusion

We systematically reviewed the literature of IQ testing of Africans and found that the average IQ of Africans in terms of British norms after correction for the Flynn Effect is 80. Lynn and Meisenberg raised the possibility that our estimate of the mean IQ was too high, because we included samples of Africans that they considered to be insufficiently

[5] The results presented in Figs. 1 and 3 were based either on our IQ computations or on the IQ computations by Lynn and Meisenberg (Table 1), in the analyses of our work or Lynn and Meisenberg's work, respectively.

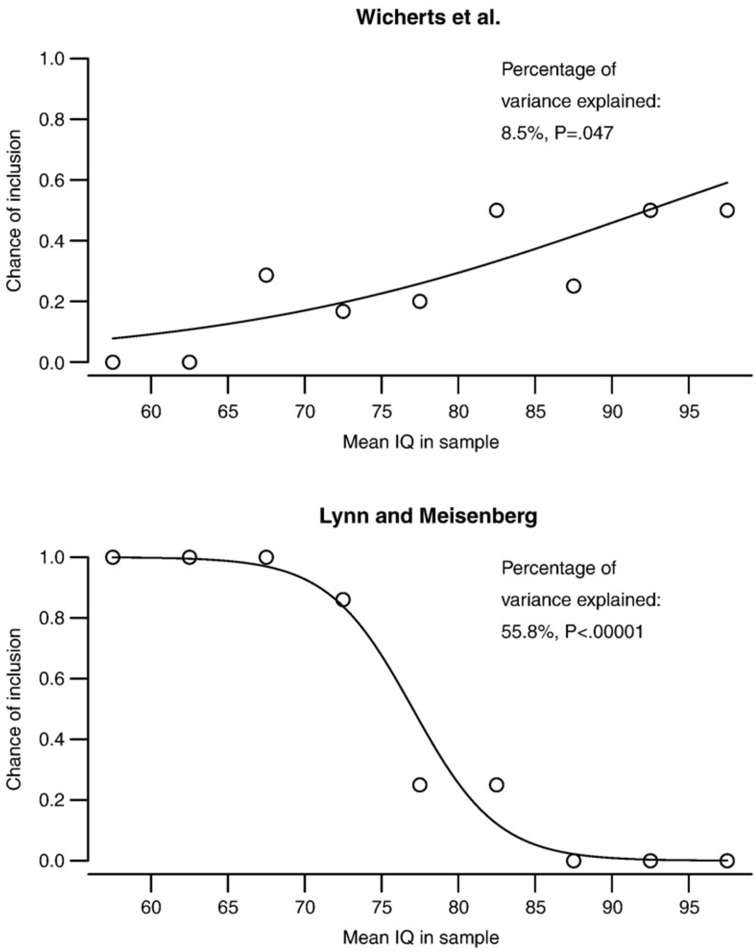


**Fig. 3.** Results of logistic regression of inclusion of sample on the average IQ in the sample for our original review (top) and Lynn and Meisenberg's review on the basis of the same samples.

representative. In light of this criticism, three independent raters assessed five objective criteria for representativeness of all samples and their ratings showed moderate to almost perfect agreement. We found that these criteria did not predict the mean IQs in the samples, which suggests that sampling bias could not have affected our results.

On the basis of their assessment of the representativeness of our 44 samples, Lynn and Meisenberg excluded 18 samples as insufficiently representative, and computed a median IQ of 69. Of the 18 samples with average IQs above 74.5, only two remained in Lynn and Meisenberg's analyses. They described these samples as "marginally acceptable" (Buj, 1981) or as "arguably [an] outlier" (Minde & Kantor, 1976), despite the fact that the original authors employed stratified sampling procedures and presented their samples as representative. Lynn and Meisenberg did not explicate their rules to assess the representativeness of samples and excluded samples because of a host of reasons, such as educational attainment, the presence of school fees, the lack of official IQ norms, poor validity of tests, and problems with age determination. However, these rules have been applied only to exclude samples with IQs above 74.5, but not to exclude samples that averaged IQs near or below 70.

More importantly, we found that the use of random sampling, stratification, normal SES samples, health status, and the original assessment of the representativeness were unrelated to Lynn and Meisenberg's assessment of samples' representativeness. Lynn and Meisenberg's sole inclusion criterion was so strongly associated with the average IQs ($r_b = -.88, p < .0001$) that it is hard to avoid the impression that Lynn and Meisenberg's assessment of representativeness was a function of the average IQ in the sample (cf. Figs. 1 and 3). In light of all the available IQ data of over 37,000 African test-takers, only the use of unsystematic methods to exclude the vast majority of data could result in a mean IQ close to 70. On the basis of sound methods, the average IQ remains close to 80. Although this mean IQ is clearly lower than 100, we view it as unsurprising in light of the potential of the Flynn Effect in Africa (Wicherts, Borsboom, & Dolan, 2010) and common psychometric problems associated with the use of western IQ tests among Africans.

## Acknowledgment

The preparation of this article was supported by VENI grant no. 451-07-016 from the Netherlands Organization for

Scientific Research (NWO). We thank Marjan Bakker and Kees-Jan Kan for their assistance.

## Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.intell.2009.11.003.

## 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.

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.

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 Hague: Mouton The Netherlands.

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.

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., et al. (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.

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.

Conant, L. L., Fastenau, P. S., Giordani, B., Boivin, M. J., Opel, B., & Nseyila, D. D. (1999). Modality specificity of memory span tasks among Zairian children: A developmental perspective. *Journal of Clinical and Experimental Neuropsychology*, *21*, 375–384.

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

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.

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.

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.

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.

Haward, L. R. C., & Roland, W. A. (1954). Some inter-cultural differences on the Draw-A-Man Test: Goodenough scores. *Man*, *54*, 86–88.

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.

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

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.

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.

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

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.

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

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.

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.

Murdoch, B. D., Fleming, J. H., Skuy, M., Painter, A., Schmidt, L., & Schutte, E. M. (1994). *Neuropsychological performance of a group of selected asymptomatic black South African children.* Pretoria, South Africa: Human Sciences Research Council.

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.

Raveau, F. H. M., Elster, E., & Lecoutre, J. P. (1976). Migration and differential acculturation. *Revue de Psychologie Appliquée*, *25*, 145–165.

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.

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.

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.

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

Wicherts, J. M., Borsboom, D., & Dolan, C. V. (2010). Why national IQs do not support evolutionary theories of intelligence. *Personality and Individual Differences*, *48*. 2010, 1–96. doi:10.1016/j.paid.2009.05.028.

Wicherts, J. M., & Dolan, C. V. (under review). The performance of English-speaking South Africans on the WAIS-III; Mean IQs of ethnic groups and measurement invariance.

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

Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. J. (in press). A systematic literature review of the average IQ of sub-Saharan Africans. Intelligence. doi:10.1016/j.intell.2009.05.002.

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.

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.