

Review of: "Facing the Facts About Test Score Gaps"

Meng Hu¹

¹ University of Hong Kong

Potential competing interests: No potential competing interests to declare.

The present article argues that the current evidence is supportive of the hereditarian hypothesis which states that ethnic group differences in intelligence (among other psychological traits) are partially (i.e., not entirely) due to genetic effects.

Indeed the article provides a good amount of evidence and is well organized. But I also believe the article missed a chunk of strong evidence in favor of the genetic hypothesis. Here are a couple remarks, by sections.

1.2. Censorship indeed is and has been a major threat for a long time. Arthur Jensen once mentioned one of his manuscript has been reviewed by 20+ reviewers and the procedure kept going on until the editor could find a reviewer willing to reject it. I believe it was "How much can we boost IQ and scholastic achievement?". More recently, I have noticed some editors' mails (that were sent/shared to me) stated explicitly they do not tolerate research on ethnic groups. In the U.S., the situation took a new turn since George Floyd. More researchers are being cancelled by activists (notably on Twitter) and research has become even more restricted. At this point, this should not be a debate whether this is true or not.

1.4. Here's a detail worth noting. The resemblance in predictive power across races indeed is widely accepted. But it also has been stable over time across races. One striking example is the SAT-grade relationship (Jensen, 1980, p. 486; Beard & Marini, 2015, 2018; Marini et al., 2019; Mattern et al., 2008).

2.1. Reardon et al. (2018) indeed reported an average of 0.7 SD gap in achievement test. They do show also that there is a large variability in their estimates, depending on areas/districts. They attribute this pattern to ethnic differences in parental SES and segregation, however all of these variables are correlated with parent IQ as well, which is probably the source of these differences. Murray (2007) reviewed the evidence regarding the black-white IQ convergence over time, supplied by his own analysis of the Woodcock-Johnson test. He concluded that the gap reduced in most studies by about $d=0.5$ before the 1970s. A popular explanation indeed seems to be educational improvement, however many other studies suggest that education does not improve IQ, or at least not the general factor of intelligence. Spitz (1986) and Besharov (2011) concluded that educational programs, most often targeted on the poor, with a larger sample of blacks, do not produce lasting gains. Studies on schooling often fail to produce far transfer (i.e, improvement in crystallized but not fluid ability), which is another evidence that intelligence was not improved (Carlsson et al., 2015; Finn et al., 2014; Dahmann, 2017; Karwowski & Milerski, 2021). Other studies which found a strong effect on IQ either did not separate crystallized and fluid composites but also did not separate the analyses by gender. My review suggests that, more often than not, when IQ gain was found, there was a strong gain among males but not for females. Because of this, the evidence is not

compelling. Therefore, while I don't believe schooling is the source of the convergence in IQ gap, I think the cause has not been well identified yet.

2.2. Wicherts et al. (2010) considered African studies to be of poor quality due to poor test administration, non-normal sample, and potential psychometric bias (i.e., non-invariance). If their estimates of 81 points for Africa is accurate, it is quite close to the Black American mean of 85. The remaining difference is, presumably, due to a mixture of environmental factors (culture bias? education? nutrition?) and genetic factors (US blacks on average have an ancestry proportion of 24% European).

2.3. The study of Spearman's hypothesis is indeed fascinating. I recognized that Warne's book (2020, p. 255) reviewed a few MGCFA studies but details are lacking. For this reason I would appreciate if studies which employ other methods are discussed as well. Compared to Multi-Group CFA, the method of correlated vectors (MCV) does not account for (sub)test bias nor does it evaluate alternative models. Te Nijenhuis et al. (2016) showed that the correlation between group differences and subtest g-loadings, using MCV, increases when biased items are removed (for a review of test bias, see: Hu, 2023). But while MCV generally is supportive of the Spearman's hypothesis, MGCFA studies do not always reach an agreement. Most studies found support for the hypothesis (Frisby & Beaujean, 2015; Lasker et al., 2021; Hu et al., 2019) when using bifactor models but not when using higher-order models (Dolan, 2000; Dolan & Hamaker, 2001) although there might be a few theoretical reasons for using bifactor over higher order models. Other, but less known, studies such as McDaniel & Kepes (2014) evaluate Spearman's by manipulating the g saturation of composite tests and found support for the hypothesis. Yet another method is by examining the group difference in Forward and Backward Digit Span, and it appears that the gap on BDS is $d=0.5$ larger than on FDS (Jensen, 1998, p. 370). Finally, what I believe to be the most underrated and yet most powerful and direct way of testing Spearman's is by examining ECT's complexity task. Jensen (1998, p. 391) reported high correlation between task complexity and the magnitude of the Black-White gap ($r=0.86$).

2.4. This section presents Jensen's theoretical argument as for why the Black-White difference cannot be entirely attributed to environment. This is compelling but I wished it discusses for instance, Jensen's (1998, p. 435) take on the two-realms hypothesis. He estimated a black-white difference in environment of about 2 SD if the heritability among blacks was 0 but 0.75 among whites. This 2 SD is obviously way larger than what is observed in real life. But by far the most convincing argument comes from deaf children data which offers a quasi-experimental study of the environmental effect. Braden's book (1994) summarizes the important points. Deaf children experience a quite severe environmental deprivation, leading to social isolation which extends into adulthood. Parental practice is anything but cognitive stimulating. Mother-child interaction is nonsupportive. Ostracism against deaf children occurs in schools. Yet deaf people perform equally well as normal-hearing people on performance IQ but 1 SD lower on verbal IQ. In a factor analysis, hearing loss, academic achievement and verbal IQ load on a common factor, but not nonverbal factor. In a correlated vector analysis, the environmental deprivation owing to deafness is inversely related with subtests' g-loadings (Braden, 1989). Overall these studies indicate that environmental deprivation, which is a common explanation for blacks' underachievement, does not lower intelligence and, therefore, has no impact on g.

2.5. I agree that admixture regression studies are at the moment the most robust analyses. The results so far are very

consistent. My only complaint is that often the survey data do not accommodate multi-racial categories well through weighting methods.

2.6. Brain size and intelligence indeed show a moderate correlation. However, I highly recommend that you mention one overlooked issue with these studies. Gignac & Bates (2017) showed that poor measurement quality negatively affects the correlation. Using only "excellent" measures, the obtained correlation is higher ($r=0.39$) than previously reported meta-analytic correlations.

This quote lacks a spacing: "Morton began his first and largest work, the *Crania Americana* of 1839"

2.7. I agree that polygenic risk scores would be more accurate if ancestry comparability is well established. It is still a point under debate. But I believe it may not necessarily underestimate the group differences, especially if it introduces random "errors".

Overall the article provides a comprehensive overview of the evidence in favor of the partial genetic hypothesis (i.e., partly genetic, partly environmental). However, as I suggested above, the article would improve by incorporating more studies which directly test the environmental hypothesis (e.g., studies on deaf children, schooling effects etc). This is because by far the most popular hypothesis as to why blacks are lagging behind in IQ test is due to their lack of education. So I'm surprised an important aspect of the environmental/cultural hypothesis was not discussed here. On the plus side, there was a lot of genetic studies mentioned here.

Despite this little issue, I believe the present article deserves publication as it highlights a problem too often ignored for political reasons and explains why it is problematic to ignore or refuse the debate or, worse, ban the research.

References cited:

Beard, J., & Marini, J. P. (2015). Validity of the SAT for predicting first-year grades: 2012 SAT validity sample (College Board Statistical Report 2015-2). New York: College Board.

Beard, J., & Marini, J. (2018). Validity of the SAT® for Predicting First-Year Grades: 2013 SAT Validity Sample. College Board.

Besharov, D. J., Germanis, P., Higney, C. A., & Call, D. M. (2011). The Abecedarian Project. Assessments of Twenty-Six Early Childhood Evaluations. College Park, MD: School of Public Policy, University of Maryland.

Braden, J. P. (1989). Fact or artifact? An empirical test of Spearman's hypothesis. *Intelligence*, 13(2), 149-155.

Braden, J. P. (1994). Deafness, deprivation, and IQ. Springer Science & Business Media.

Carlsson, M., Dahl, G. B., Öckert, B., & Rooth, D. O. (2015). The effect of schooling on cognitive skills. *Review of Economics and Statistics*, 97(3), 533-547.

Dahmann, S. C. (2017). How does education improve cognitive skills? Instructional time versus timing of instruction. *Labour Economics*, 47, 35-47.

Finn, A. S., Kraft, M. A., West, M. R., Leonard, J. A., Bish, C. E., Martin, R. E., ... & Gabrieli, J. D. (2014). Cognitive skills, student achievement tests, and schools. *Psychological science*, 25(3), 736-744.

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV

standardization data. *Intelligence*, 51, 79-97.

Gignac, G. E., & Bates, T. C. (2017). Brain volume and intelligence: The moderating role of intelligence measurement quality. *Intelligence*, 64, 18-29.

Hu, M. (2023). On The Validity of The GSS Vocabulary Test Across Groups. *OpenPsych*.

Jensen, A.R., 1980. Bias in mental testing.

Hu, M., Lasker, J., Kirkegaard, E. O., & Fuerst, J. G. (2019). Filling in the gaps: The association between intelligence and both color and parent-reported ancestry in the National Longitudinal Survey of Youth 1997. *Psych*, 1(1), 240-261.

Jensen, A.R., 1998. *The g Factor: The Science of Mental Ability*. Praeger, Westport, CT.

Karwowski, M., & Milerski, B. (2021). Intensive schooling and cognitive ability: A case of Polish educational reform.

Personality and Individual Differences, 183, 111121.

Lasker, J., Nyborg, H., & Kirkegaard, E. O. W. (2021). Spearman's Hypothesis in the Vietnam Experience Study and National Longitudinal Survey of Youth '79.

Marini, J. P., Westrick, P. A., Young, L., Ng, H., Shmueli, D., & Shaw, E. J. (2019). Differential Validity and Prediction of the SAT®: Examining First-Year Grades and Retention to the Second Year. College Board.

Mattern, K. D., Patterson, B. F., Shaw, E. J., Kobrin, J. L., & Barbuti, S. M. (2008). Differential Validity and Prediction of the SAT. (College Board Research Report 2008-4). New York: College Board.

McDaniel, M. A., & Kepes, S. (2014). An Evaluation of Spearman's Hypothesis by Manipulating g Saturation. *International Journal of Selection and Assessment*, 22(4), 333-342.

Spitz, H. H. (1986). The raising of intelligence: A selected history of attempts to raise retarded intelligence.

te Nijenhuis, J., Willigers, D., Dragt, J., & van der Flier, H. (2016). The effects of language bias and cultural bias estimated using the method of correlated vectors on a large database of IQ comparisons between native Dutch and ethnic minority immigrants from non-Western countries. *Intelligence*, 54, 117-135.

Wicherts, J. M., Dolan, C. V., & van der Maas, H. L. (2010). A systematic literature review of the average IQ of sub-Saharan Africans. *Intelligence*, 38(1), 1-20.