

Research Article

Spearman's Hypothesis Tested Between and Within Genetic Ancestry Groups in the HRS-HCAP Sample

Vladimir Shibaev¹, John Fuerst²

1. Admiral Nevelskoy Maritime State University, Vladivostok, Russia; 2. University of Maryland, United States

We tested Spearman's hypothesis, that cognitive differences between groups are concentrated on the general factor of intelligence (g), using multi-group confirmatory factor analysis (MG-CFA) and the method of correlated vectors (MCV) in the Health and Retirement Study Harmonized Cognitive Assessment Protocol (HRS-HCAP). We examined differences both between European and African ancestry groups ($N_{\text{European ancestry}} = 1805$, $N_{\text{African ancestry}} = 341$), defined by both self-identification and genetic ancestry principal components, as well as between gradients of African ancestry within the African ancestry group. A bifactor model with g and three group factors (Memory, Visuospatial Reasoning, and Speed/Attention) achieved partial strict measurement invariance across ancestry groups and also achieved strict invariance across genetic ancestry gradients within the African ancestry group. The weak form of Spearman's hypothesis fit the ancestry group comparison best. Regarding the intra-African ancestry comparison, where higher African ancestry was associated with lower cognitive scores, while the $\Delta\chi^2$ test indicated that weak Spearman's hypothesis fit better than the contra model, the evidence overall was less decisive. Nonetheless, results from the strict invariance model implied that g explained 63% of the total absolute latent mean differences across ancestry gradients, which was comparable to the proportion explained between the European and African ancestry groups (52.5%). Correlated vectors analyses using both CFA and EFA-derived g -loadings showed strong convergence with observed between-group differences ($r = .76$) and within-group African ancestry associations ($r = .50$ to $.65$). Overall, the results suggest that Spearman's hypothesis extends to genetic ancestry gradients within the admixed African American population.

Correspondence: papers@team.qeios.com — Qeios will forward to the authors

1. Introduction

Spearman's Hypothesis (SH) posits that the magnitudes of subtest cognitive differences between self-identified racial/ethnic (SIRE) groups are proportional to each subtest's loading on the general factor of intelligence, g . Originally formulated to explain differences between self-identified Black and White Americans^[1], the hypothesis has since been extended to numerous other group comparisons. The hypothesis is important because, if true, it implies that understanding the nature and source of group differences fundamentally requires understanding the nature of g itself and the sources of variation in it^{[2][3][4]}.

Tests of SH have yielded overwhelming support for it in its original context. Research on Black-White differences in the United States has drawn on two main analytical approaches: the method of correlated vectors (MCV) and multi-group confirmatory factor analysis (MGCFAs).

Meta-analyses of studies using MCV report strong correlations between g -loadings and subtest score differences^{[5][4]}, while studies using MGCFAs have largely confirmed that these differences are predominantly associated with the g factor^[6].

In light of recent findings that genetic ancestry and ancestry correlates predict cognitive outcomes within admixed American SIRE groups, several researchers have proposed that SH may extend to ancestry-associated differences within admixed populations. Table 1 summarizes results to date. In African American samples, subtest associations with self-reported European ancestry, interviewer-rated skin color, and measured genetic ancestry all correlate positively with g -loadings, constituting what is frequently called a "Jensen Effect"^[7]. European genetic ancestry also conformed to SH tested across the ancestry continuum using local structural equation modeling (LSEM) in combined European and African American samples from the PING and TCP studies^[6]. However, the evidence base remains narrow. Only one study has reported MCV results for genetic ancestry within African American or Hispanic American groups specifically, and no study has yet to apply MGCFAs or LSEM to within-group ancestry effects. The present work aims to address this gap in the literature.

Author	Sample	SIRE	Predictor	MI Tested	SH Tested with SEM	Subtest #	Jensen Effect (r)
Scarr et al. [8]*		African Americans	European blood groups	No	No	5	$r_{\text{African}} = .84$
Hu et al. ^[9]	NLSY97	African Americans	Self-reported European ancestry	SIRE	No	12	$r_{\text{African}} = .68$; $r_{\text{Combined}} = .72$
Hu et al. ^[9]	NLSY97	African Americans	Lighter skin color (β)	SIRE	No	12	$r_{\text{African}} = .85$; $r_{\text{Combined}} = .88$
Hu et al. ^[9]	NLSY97	European Americans	Self-reported African ancestry	SIRE	No	12	$r_{\text{European}} = .59$; $r_{\text{Combined}} = .57$
Lasker ^[6]	TCP	European & African American	European genetic ancestry	SIRE & Ancestry	Yes – confirmed	10	Not reported
Lasker ^[6]	PING	European & African American	European genetic ancestry	SIRE & Ancestry	Yes – confirmed	7	Not reported
Hu ^[10]	NLSY79	African Americans	Self-reported European ancestry	SIRE	No	10	$r_{\text{African}} = .40$
Fuerst et al. ^[11]	TCP	African Americans	European genetic ancestry	SIRE & Ancestry	No	10	$r_{\text{Combined}} = .85$
Fuerst et al. ^[11]	TCP	Hispanic Americans	European genetic ancestry	SIRE & Ancestry	No	10	$r_{\text{Combined}} = .92$
Fuerst et al. ^[11]	TCP	European, African, & Hispanic sample	European genetic ancestry	SIRE & Ancestry	No	10	$r_{\text{Combined}} = .89$

Author	Sample	SIRE	Predictor	MI Tested	SH Tested with SEM	Subtest #	Jensen Effect (r)
Fuerst et al. ^[12]	ABCD	European, African, & Hispanic sample	European genetic ancestry	SIRE & Ancestry	No	11	$r_{Combined} = .86$
Shibaev & Fuerst. ^[13]	Add Health	African Americans	African genetic ancestry	SIRE & Ancestry	No	4	$r_{Combined} = .94$ $r_{African} = .87$
Shibaev & Fuerst. ^[13]	Add Health	Hispanic Americans	Amerindian genetic ancestry	SIRE & Ancestry; non-invariant across ancestry at metric level	No	4	Assumptions violated
Shibaev & Fuerst. ^[13]	Add Health	Hispanic Americans	African genetic ancestry	SIRE & Ancestry	No	4	$r_{Combined} = .88$ $r_{Hispanic} = .84$
This study	HCAP	African American	African genetic ancestry	SIRE & Ancestry	Yes – partially confirmed	16	$r_{Combined} = .65$ $r_{African} = .50$ to $.51$

Table 1. Studies Testing SH with Ancestry-Related Predictors in Admixed Populations

Note: MI = measurement invariance; SH = Spearman's hypothesis; SEM = structural equation modeling; SIRE = socially-identified race/ethnicity; Jensen Effect (r) = Pearson's correlation between subtest g-loadings and the magnitude of ancestry effects across subtests. Subscripts on r indicate the sample used to derive the reference g-loading: African, Hispanic, European, or Combined. Scarr et al.'s data were re-analyzed and reported in Lasker et al.^[14]. Jensen Effects are coded such that a positive correlation indicates that subtests with higher g-loadings show stronger associations with the ancestry predictor, regardless of the direction of the ancestry effect itself (e.g., more positive with European ancestry, more negative with African ancestry).

Beyond its theoretical importance (e.g., [4]), this issue has practical implications. If genetic variation within groups is correlated with (e.g., vertically-transmitted cultural) factors which affect cognitive ability primarily through g , then the choice of cognitive measure matters in admixture regression research, as a poor measure of g will tend to produce attenuated associations. Unfortunately, few available samples combine genetic ancestry data with a sufficiently broad cognitive battery.

The HRS Harmonized Cognitive Assessment Protocol (HCAP) is a notable exception (Health and Retirement Study [Data set]). It provides a representative sample of older Americans (aged 65+) assessed on 16 cognitive measures spanning memory, reasoning, processing speed, attention, language, and orientation[15]. Crucially, a subset of participants also has accompanying genetic data, enabling parallel analyses using both polygenic scores and admixture-based ancestry estimates.

In this study, we first assess whether cognitive differences across European and African ancestry groups are measurement invariant. We then formally test for SH among these group differences using both MGCEA and MCV. Next, we split the African ancestry group into two subgroups (lower and higher African ancestry, based on the primary European–African principal component) and repeat the analyses. We hypothesize:

1. *Strict measurement invariance will hold between European and African ancestry groups.*
2. *Both MGCEA and MCV will support SH between European and African ancestry groups.*
3. *Strict measurement invariance will hold across ancestry levels within the African ancestry group.*
4. *Both MGCEA and MCV will support SH across ancestry levels within the African ancestry group.*

2. Method

2.1. Sample

Participants were drawn from the HCAP sample, which is part of the Health and Retirement Study, and included individuals aged 65 and older. Ancestry groups were identified based on a combination of self-reported race/ethnicity and genetic ancestry principal components, computed from genome-wide SNP data for the genotyped HRS sample. The European ancestry group included self-reported non-Hispanic Whites who fell within ± 1 SD of the group mean on the first and second principal components. The African ancestry group was defined analogously but ± 2 SD was used as the cutoff on the first (European–African) principal component and ± 1 SD on the second, relative to the mean of all genotyped self-

identified non-Hispanic African Americans^[16]. Together, these criteria yield non-overlapping genetic ancestry clusters.

We excluded participants who were not sampled for the HCAP survey, were older than 90 years, had a 50% or greater estimated probability of dementia in 2016, or were not US-born. The resulting analytic sample consisted of 2,146 participants (1,805 European ancestry, 341 African ancestry). Mean ages were 76.05 ($SD = 6.68$) for the European ancestry group and 73.98 ($SD = 6.47$) for the African ancestry group.

2.2. Variables

The primary variables used in this analysis are listed below.

2.2.1. Cognitive Measures

Sixteen cognitive measures from the HCAP battery were analyzed, spanning multiple cognitive domains, as follows:

1. *Consortium to Establish a Registry for Alzheimer's Disease (CERAD) Delayed Recall*. A measure of verbal episodic memory. Participants learn a 10-word list over three trials and freely recall the words after a delay. Scored as the number of words correctly recalled.
2. *Weschler Memory Scale (WMS): Logical Memory II*. A measure of narrative episodic memory. Participants listen to a short story and recall its details after a delay. Scored as the number of story units accurately reproduced.
3. *East Boston Memory Test (EBMT): Delayed Story Recall*. A measure of narrative episodic memory. Participants recall elements of a brief story after a delay. Scored as the number of story elements correctly reproduced.
4. *Mini Mental State Examination (MMSE): Delayed Recall*. A measure of short-term verbal memory. Participants recall three words presented earlier in the examination. Scored as the number of words correctly recalled (0–3).
5. *CERAD Immediate Construction Praxis*. A measure of visuospatial and constructional ability. Participants copy four geometric figures of increasing complexity. Scored as the accuracy of reproduction.
6. *Number Series*. A measure of fluid reasoning and quantitative ability. Participants identify the next number in a logical sequence. Scored as the number of items correctly solved.

7. *Raven's Matrices*. A measure of fluid reasoning and nonverbal abstract reasoning. Participants identify the missing element completing a visual pattern. Scored as the number of items correctly solved.
8. *Backwards Spelling*. A measure of attention and working memory. Participants spell a common word in reverse order. Scored as the accuracy of the response.
9. *Trail Making Test A*. A measure of processing speed and visual attention. Participants connect numbered circles in ascending order as quickly as possible. Scored as time to completion, with lower scores indicating better performance.
10. *Trail Making Test B*. A measure of executive function and cognitive flexibility. Participants alternate between connecting numbers and letters in ascending order. Scored as time to completion; the added switching demand separates executive control from the pure speed component captured by Part A.
11. *Backwards Counting*. A measure of attention and working memory. Participants count backward from a specified number. Scored as the accuracy and extent of counting.
12. *Symbol Digit Modalities Test*. A measure of processing speed and sustained attention. Participants use a coding key to match symbols to digits, completing as many items as possible within a time limit. Scored as the number of correct substitutions.
13. *Letter Cancellation*. A measure of selective attention and processing speed. Participants scan rows of letters and mark each instance of a target letter. Scored as the accuracy and speed of detection.
14. *Animal Naming*. A measure of verbal fluency and semantic memory. Participants generate as many animal names as possible within a fixed period. Scored as the number of unique valid animals named.
15. *Information*. A measure of object naming and general knowledge. Participants name or identify nine items (including the current president, common objects, and body parts), yielding a composite score of 0–9. Scores were recoded into a three-level ordinal variable: 1 (≤ 7 correct), 2 (8 correct), 3 (9 correct).
16. *MMSE: Orientation*. A measure of orientation to time and place. Participants answer questions about the current date, day, season, and location. Scored as the number of items correctly answered.

Regarding test selection, we followed the framework of Gross et al.^[15] Only two orientation items were available (orientation to time and orientation to place) both of which exhibited ceiling effects; these were combined into a single four-level ordinal MMSE Orientation variable. Most binary language items were

similarly skewed, with most respondents passing; the nine information and knowledge items were therefore combined and recoded as a three-level ordinal variable. Other items (e.g., R1SENTEN and R1DRAW) had very high pass rates and could not be meaningfully combined with other items. Where both immediate and delayed recall versions were available, we retained only the delayed recall measure to avoid redundancy given our modest sample size. CERAD Delayed Construction Praxis cross-loaded on both memory and visuospatial reasoning factors in exploratory analyses and was therefore dropped in favor of CERAD Immediate Construction Praxis, which loaded cleanly onto the visuospatial reasoning factor.

2.2.2. General intelligence (g) and g-scores

Exploratory factor analysis and scree plot analysis indicated that a three-factor solution best fit the combined sample, while retaining enough indicators per factor for model identification. To examine the relationship between European–African ancestry and latent *g*, factor scores were extracted from the MGCEFA model fit to the African ancestry sample and regressed on the European–African principal component.

2.2.3. Age and Sex

Age was drawn from the R1AGE variable, which records age at the time of the HCAP interview; sex was derived from the GENDER variable in the tracker file.

2.2.4. Probability of Dementia

In addition to excluding individuals older than 90, we excluded those whose estimated probability of dementia exceeded 50%, based on the PrDem variable derived by Hudomiet et al.^[17].

2.2.5. Self-identified race/ethnicity and US Born status

We excluded individuals who did not identify as non-Hispanic White or non-Hispanic African American, as well as those who reported not being born in the United States, using the RACE and USBORN variables from the tracker file.

2.2.6. Genetic Ancestry

Principal components (PCs) of genetic ancestry were computed from genome-wide SNP data^[16]. The European–African ancestry axis was identified as the PC most strongly correlated with ancestry-informative polygenic scores (PGS). Ware et al.^[16] reported mean PGS differences between the European and African ancestry groups, with some exceeding 10 *d*, yielding very high correlations with the European–African ancestry PC ($r > .95$), and allowing for identification.

2.2.7. Genetic *g*

Polygenic scores for educational attainment^[18] (EA3, 23andMe excluded) and g ^[19] were computed separately within European-ancestry and African-ancestry participants^[16]. We constructed a composite polygenic score for general cognitive ability (polygenic *g*) using SMTpred^[20], with weights derived from published GWAS parameters ($N_{EA} = 766,345$; $N_{cog} = 300,486$; $h^2_{EA} = .12$; $h^2_{cog} = .25$; $r_g = .73$).

2.3. Procedure

We followed these analytic procedures:

2.3.1. Age and sex regressed variables

All cognitive measures were adjusted for age in 2016 (modeled as a smooth function) and sex using generalized additive models (GAMs) via the *mgcv* package in R^{[21][22]}. Adjustment was performed in the combined sample prior to group-specific analyses, given the limited size of the African ancestry subsample ($N = 341$). Approximately continuous measures were adjusted using Gaussian GAMs. Four measures with severe ceiling effects and only three or four response categories (i.e., Information, Backwards Spelling, MMSE Orientation, and MMSE Delayed Recall) were adjusted using ordered categorical GAMs (family = *ocat*), which respect ordinal structure during covariate adjustment. Residuals were scaled and used as adjusted scores. After adjustment, MMSE Delayed Recall became approximately normal with five observed response categories and was subsequently treated as a continuous variable. The remaining three measures retained their ordinal structure and were specified as ordered indicators in all confirmatory factor analysis (CFA) models (using the *ordered* argument in *lavaan*), ensuring appropriate handling at both the adjustment and measurement modeling stages.

2.3.2. Bifactor Model Specification

In *lavaan*^[23], a bifactor model was specified with four mutually orthogonal factors: a general factor (g) loading on all 16 measures, and three orthogonal group factors: Memory (CERAD Delayed Recall, WMS Logical Memory II, EBMT Delayed Story Recall, MMSE Delayed Recall), Visuospatial Reasoning (CERAD Immediate Construction Praxis, Raven's Matrices, Number Series, Backwards Spelling), and Speed/Attention (Trail Making A, Trail Making B, Symbol Digit, Letter Cancellation, Backwards Counting). Animal Naming, Information, and MMSE Orientation loaded exclusively on g , as they did not cohere with any group factor or form a coherent fourth factor among themselves. This factor structure was guided by the three-factor EFA solution and is depicted in Figure 1. All models were estimated on age- and sex-residualized scores using diagonally weighted least squares (DWLS) with pairwise deletion for missing data.

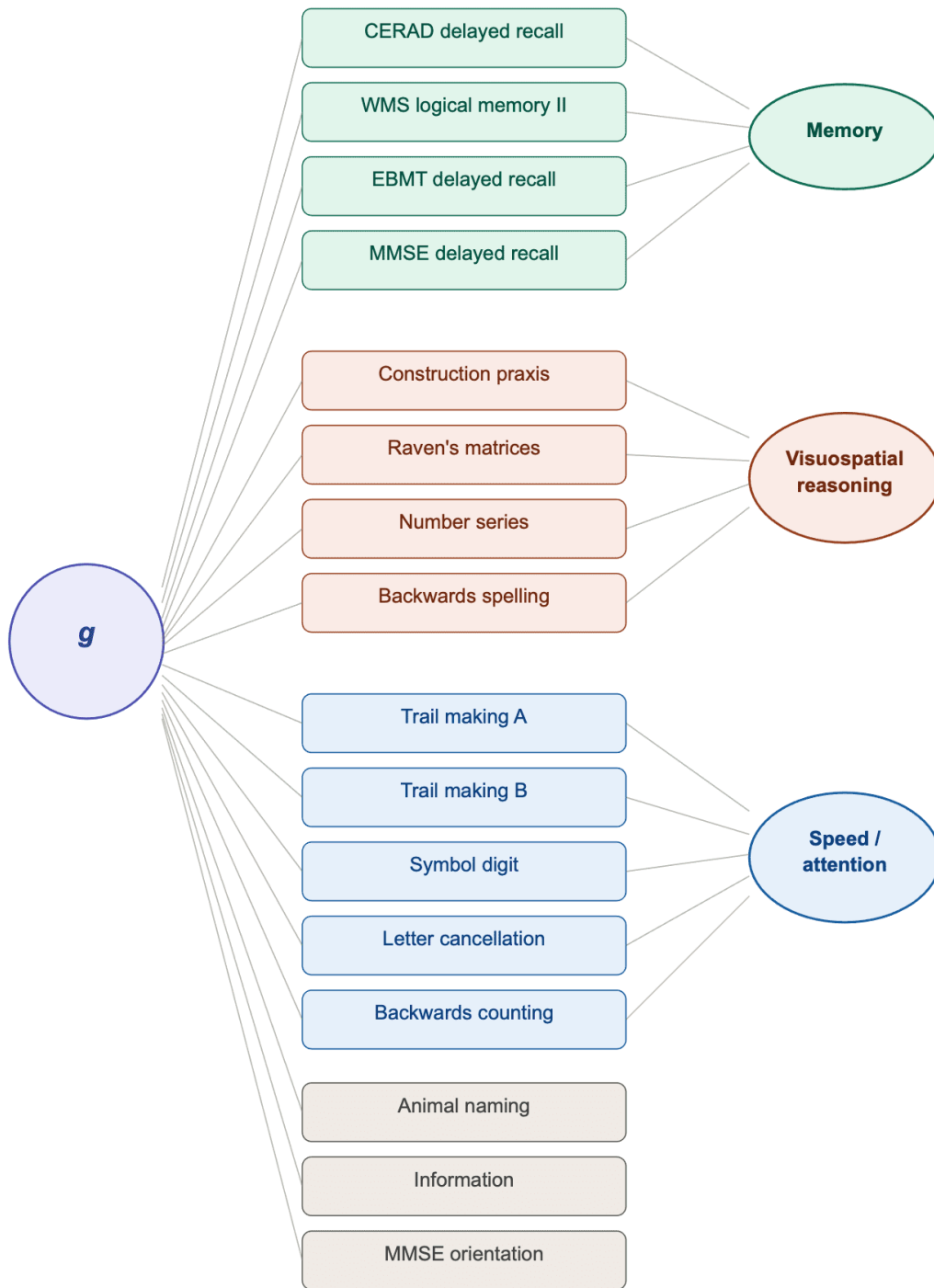


Figure 1. HCAP three-factor model. Note: The *g* factor on the left loads on all 16 measures. The three group factors on the right load only on their respective indicators. The three gray measures at the bottom (Animal Naming, Information, MMSE Orientation) load exclusively on *g*.

2.3.3. Measurement invariance and SH testing

Measurement invariance was tested sequentially: configural, metric (equal loadings), scalar (equal loadings and intercepts or thresholds), and strict (equal loadings, intercepts, and residual variances). Tests were conducted with respect to both the European–African group comparison and the African ancestry subsample split at the median. A median split was used rather than a continuous specification to maintain parallelism with the between-group analyses and because the African ancestry subsample was too small to support LSEM.

Model comparison used scaled fit indices (Comparative Fit Index [CFI], Tucker Lewis Index [TLI], Root Mean Square Error of Approximation [RMSEA], and Standard Root Mean Residual [SRMR]), with $\Delta\text{CFI} \leq .010$ serving as the primary criterion, supplemented by $\Delta\text{RMSEA} \leq .015$ or $\Delta\text{SRMR} \leq .030$ (metric invariance) or $\Delta\text{SRMR} \leq .010$ (scalar or strict invariance), following Chen^[24]. SRMR is treated as a secondary check rather than a co-criterion, as Chen^[24] demonstrated that it tends to over-reject invariant models, particularly under residual variance constraints. Scaled CFI, TLI, and RMSEA are reported because the DWLS estimator produces a mean-and-variance adjusted test statistic; the scaled versions account for non-normality and the categorical nature of ordinal indicators, yielding more accurate fit assessment than their unscaled counterparts.

Once strict invariance was established, SH was tested by comparing nested models that constrained latent means in different ways:

- a. strong SH: only the g mean free across groups;
- b. weak SH: g and all three group factor means free across groups;
- c. contra-SH: only group factor means free across groups;
- d. no difference: all latent means constrained to zero.

2.3.4. MIMIC model

Within the African ancestry subsample, African ancestry was examined as a continuous predictor of latent factors using a Multiple Indicators Multiple Causes (MIMIC) model. We compared a model in which the European–African ancestry PC predicted all latent factors to one in which it predicted g alone, to evaluate whether ancestry effects operated exclusively through g or also through the group factors.

2.3.5. MCV

We computed g -loadings using two approaches: Confirmatory Factor Analysis (CFA)-based loadings extracted from the between-group strict invariance model and Schmid-Leiman g -loadings from EFA. A three-factor solution was selected on the basis of BIC and scree plot inspection; the Schmid-Leiman orthogonalization was applied using *minres* extraction in the pooled sample using the *schmid()* function in the *psych* package^[25].

Although parallel analysis suggested five factors, inspection of the four- and five-factor solutions revealed that the additional factors were trivially small with no substantive indicator loadings, indicating overextraction, a known artifact of parallel analysis in large samples. The three-factor solution closely replicated the theoretically derived bifactor CFA structure.

For the African ancestry subsample, we did not reach a clear BIC minimum, likely reflecting insufficient power for stable factor enumeration. We therefore computed Schmid-Leiman solutions for both three and five factors and included both sets of g -loadings in subsequent correlated vectors analyses. We note, however, that the scree plot favoured a three-factor solution while parallel analysis suggested five. g -loading vectors were highly stable across methods ($r = .866-.985$), and correlated vectors results were consistent across all operationalizations.

For the EFA analyses, age- and sex-residualization was performed on the combined sample prior to group-specific analyses, given the limited size of the African ancestry subsample. The three variables with ordinal response categories (Information, Backwards Spelling, and MMSE Orientation) were treated as continuous in the EFA, given that polychoric estimation can be unstable for ordinal variables with unequal category frequencies after residualization.

Cohen's d for each subtest was computed from age- and sex-residualized scores using a pooled standard deviation across the European and African ancestry samples. The effect of ancestry was computed as the correlation between age- and sex-residualized subtest scores and the European–African ancestry PC within the African ancestry sample.

The effect of polygenic g was computed as the correlation between age- and sex-residualized subtest scores and polygenic g scores residualized on the first 10 ancestry PCs, computed separately within the European and African ancestry samples and then combined meta-analytically using an inverse-variance-weighted Fisher z -transformation.

Cultural load ratings for each subtest were computed following Kan^[26], who distinguished two operationalizations. The first, cultural loading 1, drawn from Jensen^[27], defines culture-reduced tests as those that are nonlanguage, nonscholastic, and require no specific prior knowledge. Under this operationalization, memory tests are coded as culture-reduced, because the relevant stimuli (e.g., word lists, story passages) are provided during the test itself. The second operationalization, cultural loading 2, is based on the degree to which subtest stimuli required alteration when adapting the battery across cultures and languages. Under this criterion, memory tests are coded as culture-loaded, because word lists and story passages required replacement across HCAP country adaptations^[15]. The two operationalizations therefore differ specifically in how memory tests are classified, with all other subtests coded identically. For both, cultural load was treated as a binary variable (0 = culture-reduced, 1 = culture-loaded).

When correlating vectors, we used Pearson's *r*. Correction for subtest reliability was not performed, as reliability estimates were unavailable for all 16 measures.

2.3.6. Regression plot

Latent *g* factor scores were extracted from the baseline single-group measurement model for the African ancestry sample using the `lavPredict()` function in *lavaan*, and standardized. The European–African ancestry PC was likewise standardized. A generalized additive model was fitted with standardized latent *g* as the outcome and the standardized ancestry PC as a smooth predictor term, using the `gam()` function in *mgcv*^{[21][22]}. The fitted relationship was visualized using *visreg*^[28].

3. Results

3.1. African and European Ancestry Groups

As seen in Table 2, the bifactor model fit well at the configural level (scaled CFI = .957, RMSEA = .046; for good model fit CFI should be $\geq .95$ and RMSEA $\leq .06$, see: ^[29]). Fit indices improved or remained stable from configural through metric invariance. There was a meaningful decrement in CFI moving from metric to scalar ($\Delta\text{CFI} = -.010$), which was alleviated by freeing the intercept for Letter Cancellation. The strict invariance model, which incorporated this modification, fit well.

From the strict invariance model, latent mean differences (African ancestry relative to European ancestry, in Cohen's *d* units) were: *g* = -1.18 ($p < .001$), Memory = 0.12 ($p = .232$), Visuospatial Reasoning = -0.88 ($p <$

.001), Speed/Attention = -0.56 ($p < .001$). Group differences were predominantly concentrated on g, which accounted for 52.5% of the total absolute latent mean difference.

SH models were estimated against the strict invariance model. Weak SH, which frees latent means on g and all three group factors, produced fit equivalent to the strict model ($\Delta CFI = .000$). The contra-SH model, with g constrained to zero while freeing group factor means, was rejected based on a substantial CFI decrement ($\Delta CFI = -.025$). Formal model comparisons using the Satorra (2000) scaled χ^2 difference test confirmed that weak SH fit significantly better than strong SH ($\Delta\chi^2(3) = 49.99$, $p < .001$) and contra-SH ($\Delta\chi^2(1) = 183.56$, $p < .001$). Together, these results support weak SH according to which the dominant locus of the between-group difference is therefore g, though residual differences on one or more group factors cannot be ruled out.

Level	χ^2 scaled	df scaled	CFI scaled	TLI scaled	RMSEA scaled	SRMR
Baseline	504.721	91	0.957	0.943	0.046	0.033
Configural	564.328	182	0.956	0.941	0.044	0.036
Metric	516.766	207	0.964	0.958	0.037	0.043
Scalar	613.747	220	0.954	0.950	0.041	0.044
Partial scalar	557.026	219	0.961	0.957	0.038	0.043
Strict	595.634	232	0.958	0.956	0.038	0.044
Variances	524.855	236	0.966	0.966	0.034	0.046
Weak SH	595.634	232	0.958	0.956	0.038	0.044
Strong SH	644.554	235	0.952	0.951	0.040	0.044
Contra SH	810.443	233	0.933	0.931	0.048	0.046
No difference	2084.825	236	0.785	0.782	0.085	0.045

Table 2. Measurement Invariance Fit Indices for the European-African Ancestry Comparison

$N_{\text{European Ancestry}} = 1805$ and $N_{\text{African Ancestry}} = 341$.

3.2. African Ancestry within the African Ancestry Groups

The bifactor model encountered convergence problems in this subsample because g absorbed the variance of the Visuospatial Reasoning group factor. All Visuospatial Reasoning group factor loadings were therefore fixed to zero, effectively reducing the model to a two-group-factor-plus- g specification. As seen in Table 3, with this constraint, the bifactor model fit well at the configural level (scaled CFI = .963, RMSEA = .038), and fit indices improved or remained stable through strict invariance. There was a meaningful decrement in CFI moving from strict to the latent variance model ($\Delta\text{CFI} = -.022$).

From the strict invariance model, latent mean differences (higher vs. lower European ancestry subgroup, in Cohen's d units) were $g = 0.309$ ($p = .026$), Memory = -0.265 ($p = .146$), and Speed/Attention = 0.098 ($p = .577$). Only the g difference reached statistical significance. The proportion of total absolute latent mean difference attributable to g was 62.7%, suggesting a pattern consistent with weak SH.

SH models were estimated against the strict invariance model as the baseline. By fit index criteria, neither the strong SH model ($\Delta\text{CFI} = -.001$) nor the contra-SH model ($\Delta\text{CFI} = -.007$) differed meaningfully from weak SH. Formal model comparisons using the Satorra (2000) scaled χ^2 difference test confirmed that weak SH fit significantly better than contra-SH ($\Delta\chi^2(1) = 4.62$, $p < .05$) but not significantly better than strong SH. These results tentatively support weak SH within the African ancestry subsample, though the smaller sample size limits the strength of this conclusion.

To further evaluate whether ancestry effects were concentrated on g or distributed across group factors, a MIMIC model constraining ancestry effects to g only was compared against an unconstrained model allowing effects on all factors. The models did not differ significantly ($\Delta\chi^2(3) = 4.75$, $p = .19$), providing no evidence that ancestry effects extend beyond g ; however, the analysis lacked the power to confirm that effects were exclusive to g .

Level	χ^2 scaled	df scaled	CFI scaled	TLI scaled	RMSEA scaled	SRMR
Baseline	146.502	95	0.963	0.954	0.040	0.045
Configural	237.381	190	0.963	0.953	0.038	0.057
Metric	225.265	211	0.989	0.987	0.020	0.065
Scalar	242.003	224	0.986	0.985	0.022	0.063
Strict	258.715	237	0.983	0.983	0.023	0.063
Variances	291.194	241	0.961	0.961	0.035	0.079
Weak SH	258.715	237	0.983	0.983	0.023	0.063
Strong SH	262.921	240	0.982	0.982	0.024	0.063
Contra SH	268.321	238	0.976	0.976	0.027	0.063
No difference	282.646	241	0.968	0.968	0.032	0.063

Table 3. Measurement Invariance Fit Indices Across African Ancestry within the African-Ancestry Sample

$N_{\text{Ancestry}_1} = 170$ and $N_{\text{Ancestry}_2} = 171$.

3.3. Method of correlated vectors

Four g-loading vectors were used: (1) MGCEFA g-factor loadings from the strict between-group model, (2) EFA Schmid-Leiman g-loadings from the combined European- and African-ancestry sample, and (3–4) Schmid-Leiman g-loadings from the African-ancestry sample alone using three- and five-factor solutions. These were correlated across the 16 subtests with (5) subtest correlations with the ancestry PC, (6) subtest correlations with polygenic g, (7) the European–African ancestry group mean difference (d), and (8–9) the two cultural load measures. Results are shown in Table 4.

The g-loading, ancestry group difference, within-ancestry effect, and polygenic g vectors exhibited a positive manifold, with all intercorrelations among these vectors being positive. Correlations between the g-loading vectors and the within-African-ancestry effect vector ranged from $r_s = .50$ to $.65$. The partial correlation between the ancestry effect vector and the polygenic g vector, controlling for CFA-derived g-loadings, was $r = .13$ ($p = .64$), suggesting that the covariation between ancestry effects and

polygenic g effects across subtests was largely attributable to shared g -saturation. Correlations between all other vectors and both cultural load measures were near zero or negative.

	[1]	[2]	[3]	[4]	[5]	[6]	[7]	[8]
[1] CFA g loading	1.000							
[2] EFA g loading (EA)	0.865	1.000						
[3] EFA-3 g loading (A)	0.883	0.916	1.000					
[4] EFA-5 g loading (A)	0.917	0.930	0.990	1.000				
[5] Ancestry group (EA) d	0.763	0.846	0.707	0.767	1.000			
[6] rancestry PC (A)	0.576	0.650	0.500	0.511	0.722	1.000		
[7] $r_{\text{polygenic } g}$	0.552	0.626	0.425	0.483	0.636	0.407	1.000	
[8] Cultural load 1	0.008	-0.370	-0.318	-0.246	-0.173	-0.236	0.023	1.000
[9] Cultural load 2	-0.506	-0.646	-0.578	-0.571	-0.545	-0.662	-0.080	0.595

Table 4. *Correlation Matrix of Vectors (N = 16 subtests)*

3.3.1. Association with g compared within and between groups.

The negative association between the African ancestry principal component and latent g was assessed for linearity using a generalized additive model and a quadratic term test. Both indicated a linear relationship across the full ancestry range (GAM edf = 1; quadratic term: $\beta = -0.03$, $F(1, 338) = 0.48$, $p = .487$). The regression plot, which uses the g scores from the baseline African ancestry group model, is shown in Figure 2.

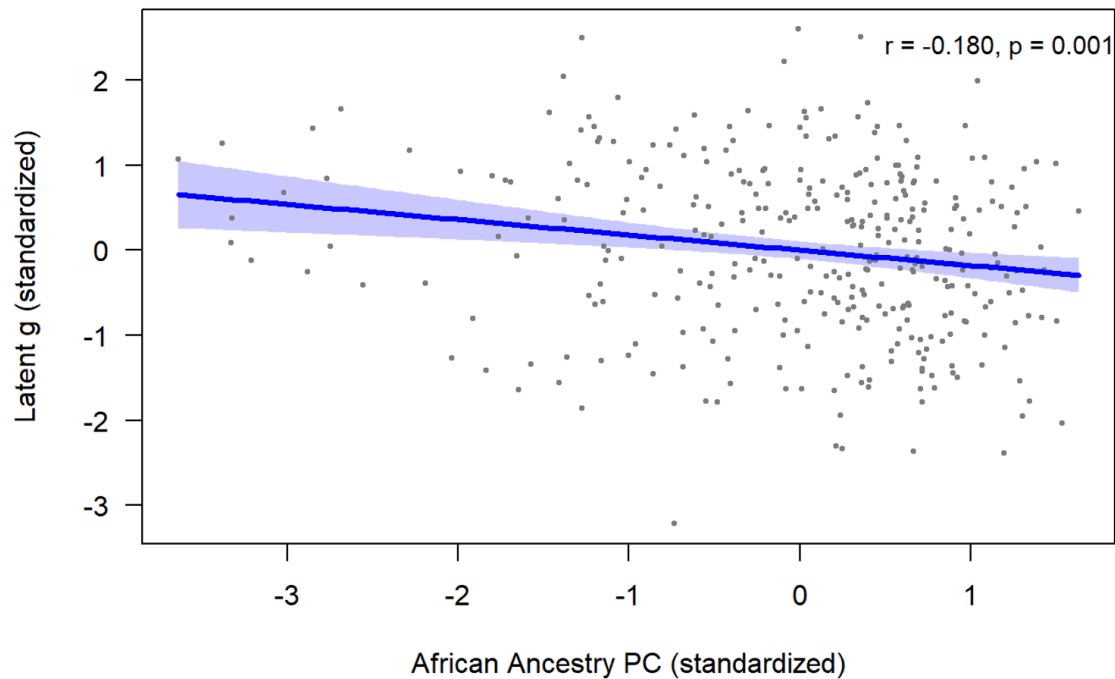


Figure 2. *g*-scores by African Genetic Ancestry in the African Ancestry Subsample. *Note:* The plotted correlation ($r = -.180$) reflects the bivariate association between the ancestry PC and latent *g* extracted from the baseline measurement model; the structural path coefficient from the MIMIC bifactor model, which partials out group factor variance, was $\beta = -.219$.

To express the within-group ancestry–*g* association as a coefficient representing a full 0%–100% shift in African ancestry (*b*), we estimated the *SD* of European ancestry proportions in the HRS African American sample using bootstrap confidence intervals reported by Baharian et al.^[30] (African ancestry: $M = 82.1\%$, 95% CI [81.55, 82.66]; European ancestry: $M = 16.7\%$, 95% CI [16.16, 17.27]; $N = 1,501$). The population *SD* was estimated from the confidence interval width using the standard back-calculation formula, yielding $SD = 10.97\%$. Ancestry proportions were modelled using a beta distribution ($\alpha = 9.21$, $\beta = 2.01$), which respects the bounded nature of ancestry proportions (constrained to 0–1) and captures the left skew expected in this population. After applying the ± 2 *SD* filter on the European–African ancestry PC, which excludes individuals below approximately 60% African ancestry, the beta distribution model estimates that approximately 66 cases (4.4%) would be trimmed, leaving an estimated effective sample of approximately 1,435 with a truncated mean African ancestry of 83.4% ($SD = 9.25\%$). Given imprecision in

these estimates, we adopt a conservative SD of 10% for rescaling, which slightly inflates b relative to the 10.97% estimate.

The unstandardized path coefficient from the ancestry PC to latent g in the unconstrained MIMIC model was -0.094 ($SE = 0.031$, $p = .002$), with a residual g variance of 0.176. Expressing this in standardized g units yields $\beta = -0.094 / \sqrt{0.176} = -0.224$. Dividing by the SD of ancestry proportions (0.10) rescales the coefficient to represent a full 0%–100% shift, giving $b = -2.24$ SD units on g . For comparison, the between-group difference in latent g was $d = 1.18$, computed using pooled within-group variance, for groups differing by approximately 81.6 percentage points in mean African ancestry (African ancestry group $M \approx 82.1\%$ vs. European ancestry group $M \approx 0.5\%$ ^[31]). Assuming linearity, extrapolating to a full 0%–100% shift gives a predicted between-group d of $1.18 / 0.816 = 1.45$.

Thus, the within-African-ancestry effect ($b = -2.24$) is substantially larger than the between-group difference extrapolated to the same scale ($d = 1.45$). This is inconsistent with the ARM prediction that ancestry-associated differences in g scale proportionally with ancestry proportion across levels of analysis. It is also inconsistent, however, with the view that genetic ancestry is unrelated to g within admixed populations.

4. Discussion

In this sample of older Americans, weak SH was well supported for the between-group European–African comparison using MGCFA. For the within-group ancestry analysis, MGCFA yielded less decisive evidence. The contra-SH model fit statistically significantly worse than weak SH, but the practical difference in fit was modest by conventional criteria, limiting the strength of this conclusion. Nonetheless, correlated vectors analyses revealed positive correlations between the ancestry effect vector and g -loadings ($r = .50$ – $.65$), consistent with weak Spearman’s hypothesis. Across both comparisons, g accounted for approximately 53–63% of the total absolute latent mean difference, as estimated from the strict invariance models.

When rescaled to a common ancestry range, the implied magnitude of within-group ancestry effects on g ($b = 2.24$) exceeded the extrapolated between-group difference ($d = 1.45$); both figures represent a full 0%–100% shift in African ancestry. Memory showed a non-significant positive association with African ancestry after controlling for g , a direction consistent with the positive between-group Memory difference ($d = 0.12$) and with results reported by Frisby and Beaujean^[32].

A growing number of studies have found that African, European, and Amerindian genetic ancestry predicts cognitive ability within admixed populations in the United States^{[9][33]}. These findings are consistent with vertical transmission models, in which pre-existing group differences in cognitive ability are passed down across generations through genetic, cultural, or other pathways. Identifying which dimensions of cognitive ability underlie the ancestry-cognition association is theoretically important, as different transmission mechanisms may produce distinct profiles of subtest associations^{[1][34]}. For example, adoption gains, a direct measure of vertical cultural transmission, appear to not be *g*-loaded^{[35][36]}.

In our sample, phenotypic *g*-loadings, ancestry associations, and polygenic *g* effects exhibited a positive manifold, suggesting that ancestry-associated cognitive differences are concentrated on *g* rather than on specific abilities. Cultural load, as operationalized by Kan^[26], did not follow this pattern. Kan^[26] and Kan et al.^[37] reported that cultural loadings, *g*-loadings, and heritability correlate positively across subtests of traditional intelligence tests. This discrepancy may reflect the nature of the HCAP battery: designed as part of a family of international studies assessing cognitive aging rather than as a formal intelligence test, its measures may not map onto cultural load distinctions in the same way that traditional IQ subtests do.

These results add to a growing body of evidence suggesting that the choice of cognitive measures matters when conducting admixture regression analyses. Some admixture regression studies have relied on cognitive composites that are likely weak proxies for *g*. For example, Llibre-Guerra et al.^[38] assessed cognitive performance in a Cuban sample using a composite based on delayed recall, verbal fluency, and a dementia screening tool. The admixture sample also included all identified dementia cases (and constituted 37.3% of this sample), raising concerns about measurement non-invariance across dementia and non-dementia cases and about the psychometric coherence of the resulting composite. In the unadjusted model, not controlling for antecedents such as education, the ancestry effect in that study appeared to be approximately 0.35 SD per full African ancestry shift, considerably smaller than the 2.24 SD observed in the present study. While multiple factors likely contribute, including the different population contexts, sample composition, and the cognitive battery, one contributing factor is likely the use of cognitive measures that are relatively weak proxies for *g*. We suggest that admixture regression studies explicitly consider the *g*-saturation of their cognitive measures in addition to other measurement properties.

This study has several limitations, most notably the small sample sizes. Spearman's hypothesis for genetic ancestry within SIRE groups will need to be tested in larger samples. Our primary aim was to assess whether ancestry-associated cognitive differences exhibit a Jensen effect; accordingly, we did not explore potential mediators, such as parental SES or other background factors, that might statistically account for ancestry–cognitive ability associations, as would be expected if intergenerational transmission pathways partly account for the observed effects. The restriction of the sample to older adults (aged 65+) also limits generalizability to younger populations, and the unavailability of subtest reliability estimates precluded correction for attenuation in the correlated vectors analyses.

Statements and Declarations

Data Availability

This study is a secondary analysis of data from the Health and Retirement Study (HRS). HRS is sponsored by the National Institute on Aging (U01AG009740) and conducted by the University of Michigan. Both HCAP and genetic data were accessed under HRS's sensitive data-use agreement: <https://hrsdata.isr.umich.edu/data-products/sensitive-health>. The analysis used de-identified data and did not constitute human-subjects research requiring additional IRB review.

References

1. ^a ^b Spearman C (1927). *The Abilities of Man: Their Nature and Measurement*. Macmillan.
2. [^] Jensen AR (1992). "Spearman's Hypothesis: Methodology and Evidence." *Multivar Behav Res.* 27(2):225–233.
3. [^] Dolan CV (2000). "Investigating Spearman's Hypothesis by Means of Multi-Group Confirmatory Factor Analysis." *Multivar Behav Res.* 35(1):21–50.
4. ^a ^b ^c ^d ^e Nijenhuis J, Choi YY, van den Hoek M, Valueva E, Lee KH (2019). "Spearman's Hypothesis Tested Comparing Korean Young Adults With Various Other Groups of Young Adults on the Items of the Advanced Progressive Matrices." *J Biosoc Sci.* 51(6):875–912.
5. [^] ^e Nijenhuis J, van den Hoek M (2016). "Spearman's Hypothesis Tested on Black Adults: A Meta-Analysis." *J Intell.* 4(2):6.
6. ^a ^b ^c ^d ^e Lasker J (2019). "A Surfeit of Evidence on Spearman's Hypothesis." 20th Annual Meeting of the International Society for Intelligence Research.

7. [△]Rushton JP (1998). "The "Jensen Effect" and the "Spearman-Jensen Hypothesis" of Black-White IQ Differences." *Intelligence*. 26(3):217-225.
8. [△]Scarr S, Pakstis AJ, Katz SH, Barker WB (1977). "Absence of a Relationship Between Degree of White Ancestry and Intellectual Skills Within a Black Population." *Hum Genet*. 39(1):69–86.
9. ^{a, b, c, d}Hu M, Lasker J, Kirkegaard EOW, Fuerst JG (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.
10. [△]Hu M (2022). "More Evidence of an Association Between European Ancestry and g Among African Americans: An Analysis of a Nationally Representative Sample of American Youth." *Mankind Q*. 62(3).
11. ^{a, b, c}Fuerst J, Kirkegaard EOW, Piffer D (2024a). "More Research Needed: There Is a Robust Causal Vs. Confounding Problem for Intelligence-Associated Polygenic Scores in Context to Admixed American Populations." In Connor G, Fuerst JGR (Eds.), *Studying Correlations Between Genetic Variation and Test Score Gaps*. pp. 146–173. Cambridge: Cambridge Scholars Publishing.
12. [△]Fuerst J, Shibaev V, Kirkegaard EOW (2024b). "A Genetic Hypothesis for American Race/Ethnic Differences in Mean g: A Reply to Warne (2021) With Fifteen New Empirical Tests Using the ABCD Dataset." In Connor G, Fuerst JGR (Eds.), *Studying Correlations Between Genetic Variation and Test Score Gaps*. pp. 276–332. Cambridge: Cambridge Scholars Publishing.
13. ^{a, b, c}Shibaev V, Fuerst J (2026). "Admixture Regression, Cognitive Ability, and Educational Attainment: A Social Science Application Using the Add Health Survey." *Qeios*. doi:[10.32388/EO7DND](https://doi.org/10.32388/EO7DND).
14. [△]Lasker J, Pesta BJ, Fuerst JGR, Kirkegaard EOW (2024). "Global Ancestry and Cognitive Ability." In Connor G, Fuerst JGR (Eds.), *Studying Correlations Between Genetic Variation and Test Score Gaps*. pp. 115–145. Cambridge: Cambridge Scholars Publishing.
15. ^{a, b, c}Gross AL, Li C, Briceño EM (2023). "Harmonisation of Later-Life Cognitive Function Across National Contexts: Results From the Harmonized Cognitive Assessment Protocols." *Lancet Healthy Longev*. 4(10):e573–e583.
16. ^{a, b, c, d}Ware E, Hornish U, Nolte M, Faul J (2024). "HRS Polygenic Scores – Release 5: 2006–2012 Genetic Data, Sensitive Health Data." Survey Research Center, University of Michigan.
17. [△]Hudomiet P, Hurd MD, Rohwedder S (2022). "Trends in Inequalities in the Prevalence of Dementia in the U.S." *Proc Natl Acad Sci U S A*. 119(46):e2212205119.
18. [△]Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, Cesarini D (2018). "Gene Discovery and Polygenic Prediction From a Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals."

Nat Genet. 50:1112–1121.

19. [△]Davies G, Lam M, Harris SE, Trampush JW, Luciano M, Hill WD, Deary IJ (2018). "Study of 300,486 Individuals Identifies 148 Independent Genetic Loci Influencing General Cognitive Function." *Nat Commun.* 9(1):2098.
20. [△]Turley P, Walters RK, Maghzian O, Okbay A, Lee JJ, Fontana MA, Benjamin DJ (2018). "Multi-Trait Analysis of Genome-Wide Association Summary Statistics Using MTAG." *Nat Genet.* 50(2):229–237.
21. [△][▷]Wood S, Wood MS (2015). "Package 'mgcv'." *R Package Version.* 1(29):729.
22. [△][▷]Wood SN, Pya N, Säfken B (2016). "Smoothing Parameter and Model Selection for General Smooth Models." *J Am Stat Assoc.* 111(516):1548–1563.
23. [△]Rosseel Y (2012). "lavaan: An R Package for Structural Equation Modeling." *J Stat Softw.* 48(2):1–36.
24. [△][▷]Chen FF (2007). "Sensitivity of Goodness of Fit Indexes to Lack of Measurement Invariance." *Struct Equ Model.* 14(3):464–504.
25. [△]Revelle W (2024). "psych: Procedures for Psychological, Psychometric, and Personality Research (R package version 2.4.3)." Northwestern University. <https://CRAN.R-project.org/package=psych>.
26. [△][▷][⊗]Kan K-J (2011). "The Nature of Nurture: The Role of Gene-Environment Interplay in the Development of Intelligence." [Doctoral dissertation, University of Amsterdam].
27. [△]Jensen AR (1980). *Bias in Mental Testing*. Free Press.
28. [△]Breheny P, Burchett W (2017). "Visualization of Regression Models Using Visreg." *R J.* 9(2):56–71.
29. [△]Hu LT, Bentler PM (1999). "Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives." *Struct Equ Model.* 6:1–55.
30. [△]Baharian S, Barakatt M, Gignoux CR, Shringarpure S, Errington J, Blot WJ, Bustamante CD, Kenny EE, Williams SM, Aldrich MC, Gravel S (2016). "The Great Migration and African-American Genomic Diversity." *PLoS Genet.* 12(5):e1006059.
31. [△]Nagar SD, Conley AB, Jordan IK (2020). "Population Structure and Pharmacogenomic Risk Stratification in the United States." *BMC Biol.* 18(1):140.
32. [△]Frisby CL, Beaujean AA (2015). "Testing Spearman's Hypotheses Using a Bi-Factor Model With WAIS-IV/WMS-IV Standardization Data." *Intelligence.* 51:79–97.
33. [△]Fuerst J, Hu M (2023). "Deep Roots of Admixture-Related Cognitive Differences in the USA?" *Qeios.* 5(7). doi:10.32388/CCN648.8.

34. [△]Woodley of Menie MA, Peñaherrera-Aguirre M, Sarraf MA (2025). "The Reversal of Spearman's Hypotheses in Incarcerated Populations and the Role of Non-Shared Environmentality." *Intell Cognit Abil.* 1(1):86-95.
35. [△]Jensen AR (1997). "Adoption Data and Two g-Related Hypotheses." *Intelligence.* 25(1):1-6.
36. [△]te Nijenhuis J, Jongeneel-Grimen B, Armstrong EL (2015). "Are Adoption Gains on the g Factor? A Meta-Analysis." *Pers Individ Dif.* 73:56-60.
37. [△]Kan KJ, Wicherts JM, Dolan CV, van der Maas HL (2013). "On the Nature and Nurture of Intelligence and Specific Cognitive Abilities: The More Heritable, the More Culture Dependent." *Psychol Sci.* 24(12):2420-2428.
38. [△]Llibre-Guerra JJ, Li Y, Allen IE, Llibre-Guerra JC, Rodríguez Salgado AM, Peñalver AI, Llibre-Rodríguez JJ (2022). "Race, Genetic Admixture, and Cognitive Performance in the Cuban Population." *J Gerontol A Biol Sci Med Sci.* 77(2):331-338.

Supplementary data: available at <https://doi.org/10.32388/0S5GAA>

Declarations

Funding: No specific funding was received for this work.

Potential competing interests: No potential competing interests to declare.