MANKIND QUARTERLY 2021 62:1 151-185

More Research Needed: There is a Robust Causal vs. Confounding Problem for Intelligence-associated Polygenic Scores in Context to Admixed American Populations

John G.R. Fuerst*
Cleveland State University, USA
Ulster Institute for Social Research, London, UK
Emil O.W. Kirkegaard, Davide Piffer
Ulster Institute for Social Research, London, UK
* Corresponding author: email j122177@hotmail.com

Amongst admixed American populations, polygenic scores for educational attainment and intelligence (eduPGS), genetic ancestry, and cognitive ability covary. We argue that this plausibly could be due to either confounding or to causally-relevant genetic differences between ancestral groups. It is important to determine which scenario is the case in order to better assess the validity of eduPGS. We investigate the robustness of the confounding vs. causal concern by examining, in detail, the relation between eduPGS, ancestry, and general cognitive ability in East Coast Hispanic and non-Hispanic samples. EduPGS predicted *g* among Hispanics ($B = 0.175$, $N = 506$) and all other groups (European: $B = 0.230$, $N = 4914$; European-African: $B = 0.215$, $N = 228$; African: $B = 0.126$, $N = 2179$) with controls for ancestry. Path analyses revealed that eduPGS, but not skin color, partially statistically explained the association between *g* and European ancestry among both Hispanics and the combined sample. Also, we were unable to account for eduPGS differences between the main ancestral populations using common tests for ascertainment bias and confounding related to population stratification. Overall, our results suggest that eduPGS derived from European samples can be used to predict *g* in American populations. However, owing to the uncertain cause of the ancestry-related differences in eduPGS, it is not yet clear how the effect of ancestry should be handled. We argue that causally-informative research is needed to determine the source of the relation between eduPGS, genetic ancestry, and cognitive ability.

Keywords: Education, Intelligence, Polygenic scores, Hispanics, African-Americans, Philadelphia

General intelligence is perhaps the most powerful variable in social science, as it is not only measurable with relative ease but often strongly predicts numerous academic, economic, occupational, social, and health-related outcomes (Deary, 2010; Gottfredson, 2003). Owing to its relation to general human well-being in contemporary society, understanding the causes of individual differences in general intelligence is of significant social importance. Moreover, the search for the causes of these differences has led many to investigate the environmental and genetic determinants of general intelligence (Plomin et al., 2014). More recently, large-scale genome-wide association studies (GWAS) have been conducted to identify the genetic variants underlying the hereditary contribution to individual differences (see, e.g., Lee et al., 2018; Sniekers et al., 2017). At present, among Europeans, 4-10% of the variance in cognitive ability can be explained by intelligence and educational polygenic scores (eduPGS; Plomin & von Stumm, 2018).

GWAS have mostly been conducted on European-origin populations. Among certain groups (e.g., African Americans), European-based eduPGS have been found to display attenuated predictive accuracy with respect to cognitive ability (Guo, Lin & Harris, 2019; Lasker et al., 2019; Rabinowitz et al., 2019). There are several possible reasons for this attenuation. One explanation appeals to possible lower within-group heritability in non-European groups (e.g., Rabinowitz et al., 2019). This is a theoretically plausible account, since predictive accuracy is a function of heritability (Daetwyler, Villanueva & Woolliams, 2008). However, because the heritability of IQ is similar across racial and ethnic groups in the US (for a meta-analytic review, see Pesta et al., 2020), a more likely possibility is decay of linkage disequilibrium (LD), which results in different correlations between SNPs across different ancestry groups (Zanetti & Weale, 2018).

The significance of LD decay in attenuating transethnic predictive accuracy depends on the specific populations in question, as the impact of LD decay will be modified by ancestry-assortative mating, selection, admixture, genetic drift and other factors, which vary across populations. Although the predictive accuracy of European-derived eduPGS was found to be attenuated among African Americans, this did not seem to be the case in an admixed Brazilian sample (Horta, Hartwig & Victora, 2018). Though, if predictive validity is lower in unadmixed Africans than in Europeans, but an admixed population also has greater variance in polygenic scores than unadmixed Europeans, this greater variance in the admixed population can hide the lower predictive validity in Africans. Nor was it the case for Asian and non-Black Hispanic Americans (Guo, Lin & Harris, 2019). A similar pattern has been discovered for other, medically related PGS, with the accuracy of PGS being attenuated the most for African Americans (Duncan et al., 2019, Figure 2). This is likely because African

Americans are primarily a sub-Saharan African group, and because sub-Saharan Africans are the continental lineage most genetically distant from Europeans (and other major races; Duncan et al., 2019). Generally, the validity of eduPGS has to be tested on a population by population basis and cannot be assumed to generalize.

Moreover, among admixed populations, evaluating the validity of PGS is complicated when ancestry, PGS, and traits of interest covary. In these situations, the covariance could result from a combination of ascertainment bias and confounding related to population stratification, or could be a result of genetic differences causally related to the trait (Lawson et al., 2020). Depending on the scenario, corrections for ancestry may either be accurate, overcorrect, or undercorrect. Thus, it is important to evaluate the reason for the intercorrelations.

In the case of admixed American populations, previous research has shown that ancestry, eduPGS, and cognitive ability scores are inter-correlated (Lasker et al., 2019). This is found when examining self-identified racial/ethnic (SIRE) groups separately. For example, Kirkegaard et al. (2019), Lasker et al. (2019), Warne (2020), and Guo, Lin and Harris (2019) found that cognitive ability was associated with admixture components within ethnic groups (e.g., self-identifying African Americans & Hispanics). These findings were robust to controls for SIRE, despite the fact that, as Fang et al. (2019, p. 764) noted, SIRE "acts as a surrogate to an array of social, cultural, behavioral, and environmental variables" and so "stratifying on SIRE has the potential benefits of reducing heterogeneity of these non-genetic variables and decoupling the correlation between genetic and non-genetic factors."

There are two obvious scenarios which could explain the covariance between ancestry, eduPGS, and cognitive ability. These are depicted in Figure 1. The first, (a), is the confounding scenario. Here, ancestry is associated with causally irrelevant eduPGS-related loci as a result of ascertainment bias and confounding related to population stratification (e.g., Kim et al., 2018; Martin et al., 2017); coincidentally, ancestry is also associated with cognitive ability by way of the environment. The environmental differences cause socioeconomic ones which, in turn, cause cognitive ability ones. In this scenario, eduPGS would have a spurious relation to *g* between groups, despite a causal one within groups. The second, (b), is a causal scenario. In this case, ancestry is associated with causally relevant eduPGS-related loci due to evolutionary history. These genetic differences cause cognitive ability ones which, in turn, lead to socioeconomic

ones. In this case, eduPGS would be a component with constitutive explanatory relevance in the relation between ancestry and *g*.[1]

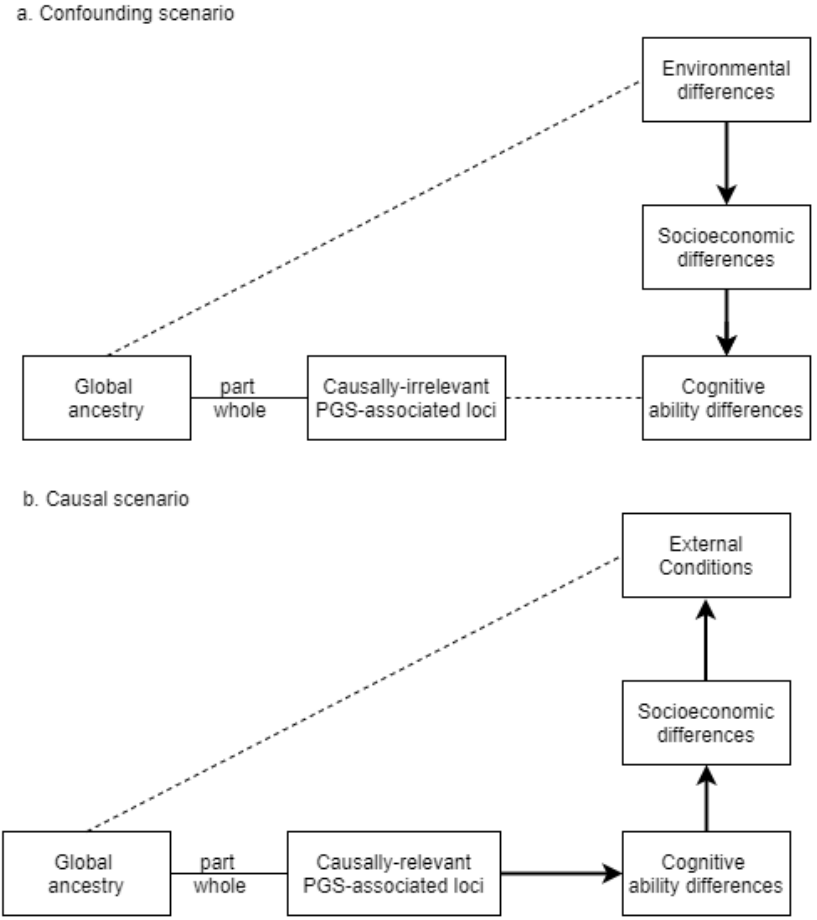

Figure 1. *Theoretical scenarios depicting the relation between ancestry, eduPGS, cognitive ability, and socioeconomic differences.*

[1] Regarding terminology, since the relation between ancestry and eduPGS is better characterized as constitutive rather than causal, eduPGS could be better characterized as being a component with constitutive explanatory relevance (Craver, 2007; Ylikoski, 2013), rather than a mediator as defined by Pearl (2014).

Note: In the confounding scenario (a) global or overall ancestry is correlated with environmental differences (e.g., due to vertical environmentally-transmitted factors or, perhaps, racial phenotype-based discrimination); these environmental differences cause cognitive ones by way of socioeconomic ones; at the same time, global ancestry is correlated with causally-irrelevant cognitive ability differences. In the alternative causal scenario (b), global ancestry is correlated with causally-relevant PGS-associated loci owing to evolutionary divergence in the quantitative genetic trait. These genetic differences cause cognitive ability ones which in turn cause socioeconomic ones, which lead to differences in external conditions. Note, this is not intended as a directed acylic graph.

Determining which of the two models comes closer to reality is important for an accurate interpretation of PGS in admixed American populations (Lawson et al., 2020). We do not attempt to determine which scenario is correct in this paper. Rather, we investigate if there is a robust confounding vs. causal problem with respect to cognitive ability that needs to be solved by future research. This would be the case if the data was found consistent with either a causal or confound scenario as illustrated above.

To do this, we used the Trajectories of Complex Phenotypes sample, which is a large population-representative Philidelphian sample. This sample has a number of advantages over other available ones. First, it is a local sample and so the issue of ancestry-related geographic confounding (Kirkegaard et al., 2019; Lawson et al., 2020) is not of significant concern. Second, the cognitive battery is well designed and allows for a robust examination of psychometric bias between ethnic groups. Third, the heritability of cognitive ability has been previously reported for the two largest ethnic groups (African and European Americans; Mollon et al., 2021); this is important as the predictive validity of PGS depends on the trait heritability (Pesta et al., 2020).

We focus here on the relation between ancestry, eduPGS, and general cognitive ability in the Hispanic sample and compare these results with those from European, biracial European-African, and African American samples. As far as we are aware, no previous research has investigated this issue using a largely Caribbean Hispanic sample. For background, Hispanics residing in Philadelphia are largely a Caribbean-origin, admixed population, with a predominantly Puerto Rican component (69% Puerto Rican; US Census Bureau, 2016). Hispanics, in general, have been found to score significantly below European Americans on measures of general intelligence, with Roth et al. (2017, Table 12) calculating meta-analytic effects of d = -0.65 to -1.04.

Given these typically large differences, we assess measurement invariance to ensure that our measure of cognitive ability functions the same for European

and Hispanic Americans. Next, via regression, we examine whether associations between genetic ancestry and general cognitive ability can be accounted for by either SIRE, skin color, or parental education. Following this, we evaluate the predictive validity of eduPGS. We next examine whether the validity of eduPGS is robust to controls for genetic ancestry and color. Using path modeling, we then explore the extent to which eduPGS can statistically explain the association between European genetic ancestry and general cognitive ability.

After, we apply Jensen's Method of Correlated Vectors (MCV; Jensen, 1998) to examine the relation between eduPGS and *g*-loadings. The expectation is that eduPGS effects will be largest on the most *g*-loaded subtests since genetic effects, in contrast to environmental effects, tend to be *g*-loaded (te Nijenhuis et al., 2019). This phenomenon, of a positive correlation between the vector of *g*-loading and some other vector, is referred to as a "Jensen effect" (Woodley of Menie, Fernandes & Hopkins, 2015). While EduPGS has been found to be differentially associated with subtest scores (de la Fuente et al., 2020; Genç et al., 2021), these samples do not allow for a robust evaluation of whether eduPGS exhibits a Jensen effect owing to a limited number of subtests or small sample sizes. Thus, we evaluate the matter here. We further examine the relation between *g*-loadings, group differences, and ancestry. Since a simple causal scenario would predict a positive manifold of Jensen effects for eduPGS, heritability, ancestry, and group differences, these analyses can show if the data are consistent with such a scenario.

Finally, we examine the eduPGS scores for common forms of ascertainment bias and confounding related to population stratification to see if we can easily account for the effects of population structure. If we can, there may be no confounding vs. causal scenario in need of resolving. To be clear, though, it is outside the scope of this paper to investigate all possible forms of confounding. We therefore restrict consideration to some forms of bias discussed in the literature. The overall goal is to better understand how eduPGS, ancestry, SIRE, color, and cognitive ability are associated with one another and to evaluate whether there is a robust confound vs. causal concern in need of further research.

Materials and Methods

The Trajectories of Complex Phenotypes (TCP) study was conducted by the Center for Applied Genomics at the Children's Hospital of Philadelphia, and the Brain Behavior Laboratory at the University of Pennsylvania. Participants were English-speaking Philadelphians, aged 8-21 years at the time of testing (which was done primarily between 2010 and 2013). Those with severe cognitive or medical impairments were excluded from the final sample.

Cognitive ability

Participants were administered the Penn Computerized Neurocognitive Battery (PCNB; Gur et al., 2010; Moore et al., 2015), a psychometrically robust cognitive battery that incorporates tasks linked to specific brain systems. This is a widely used 1-hour neurocognitive battery, which has been previously validated for this sample (Moore et al., 2015; Satterthwaite et al., 2016), which has an age range of 8 to 22 years. Measurement invariance for this battery was previously found to hold between African and European Americans (He & Li, 2021; Lasker et al., 2019). We created *g*-scores based on ten subtests for which measurement invariance (MI) held between our major SIRE groups (European, Hispanic, and African Americans). Details about the measures, the tests for measurement invariance, and subtest scores by subgroup can be found in Supplementary File 1.

Parental education

Following Lasker et al. (2019), we computed *z*-scores individually for paternal and maternal education and then averaged these. The average score was then *z*-scored again. Paternal and maternal education were the only available measures of socioeconomic status (SES). Parental education is arguably the most relevant socioeconomic related variable as it has the strongest correlation with educational-based PGS (Lee et al., 2018; Figure 4) and as it is as highly correlated with cognitive ability / scholastic achievement as are other commonly used parental-based indicators, such as occupation and income (Sirin, 2005, Table 3). That said, parental education is not a complete measure of SES, and we do not treat it as such.

SIRE

Self-identified race/ethnicity (SIRE) was based on yes/no questions for which the participants could select multiple races or ethnicities out of the following set of choices: Black or African-American; American Indian or Alaskan Native; Asian; European-American; Hispanic/Latino; Native Hawaiian/Pacific Islander; Other; and Not available/Pending validation. Those who identified as European only and did not report being Hispanic were coded as “European American,” and mutatis mutandis for “African” and “European-African”. Those who identified as Hispanic, regardless of self-identified race, were coded as such. Among Hispanics, we also identified individuals by self-identified race: European, African, Other, and a mixed category of all others (e.g., European-African Hispanic).

Genetic Ancestry Percentages

Different arrays covered different variants, so to obtain overlapping sets of single nucleotide polymorphisms (SNPs), we imputed variants with the Michigan Imputation Server (https://imputationserver.sph.umich.edu/index.html). We used this server with the Minimac3 imputation algorithm, 1000G Phase 3 v5 as reference panel, and Eagle v2.3 Phasing. For computational efficiency when calculating admixture percentages, we filtered the 15.5 M variants available to the 6.5 M with a minor allele frequency (MAF) of at least 0.05. For color scores (based on a total possible of 35 variants), we did not filter by MAF, so as not to lose variants. Imputation was done using PLINK v1.90b6.8 (Chang et al., 2015). For individual ancestry estimates, we used ADMIXTURE version 1.3.0 D.H. (Alexander, Novembre & Lange, 2009). Prior, we merged the TCP with 1000 Genomes reference samples: European (British, CEU, Finnish, Spanish, and Tuscan), African (African-American SW, African-Caribbean, Esan, Gambian Mandinka, Luhya, Mende, and Yoruba), Amerindian (Peruvian), and mixed American (Puerto Rican, Colombian, and Mexican-American). We did this to gain a reliable estimate of Amerindian ancestry, since Hispanics in this sample were a three-way admixed population. We then ran ADMIXTURE with k = 3 genetic clusters. The results for the 1000 Genomes reference populations appear in Figure 2, while those for the TCP European, European-African, African, and Hispanic samples appear in Figure 3. Note that the Hispanics in the sample were largely an Afro-European descent group, with an ancestry profile matching the predominantly Puerto Rican origin of the Philadelphia Hispanic population.

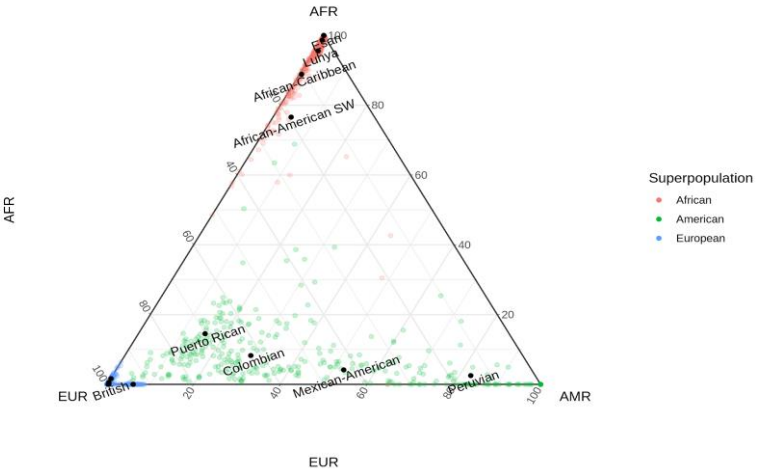

Figure 2. *Admixture ternary plot for 1000 Genomes reference samples.*

Note: Due to clumping, some labels have been removed. Populations are: European (British, CEU, Finnish, Spanish, Tuscan), African (African-American SW, African-Caribbean, Esan, Gambian Mandinka, Luhya, Mende, Yoruba), Amerindian (Peruvian), and mixed American (Puerto Rican, Colombian, and Mexican-American).

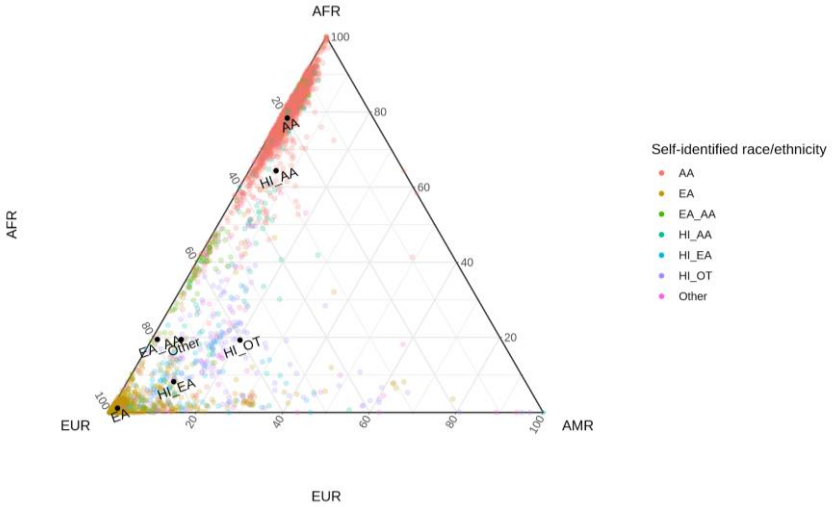

Figure 3. *Admixture ternary plot for TCP self-identified race-ethnicity (SIRE) groups*

Note: EA = non-Hispanic European American, EA_AA = non-Hispanic European-African American, AA = non-Hispanic African American, HI_AA= Hispanic African American, HI_EA= Hispanic European American, HI_OT = Hispanic Other, and Other = any other individual who also identified as Hispanic.

Skin color

Because the data did not include measures of appearance, we followed Lasker et al. (2019) and calculated phenotypic scores from genotypic data. Namely, we imputed phenotype based on genotype using the HIrisPlex-S web application (https://hirisplex.erasmusmc.nl/). Developed for use by the U.S. Department of Justice in forensic investigations, and validated on thousands of people from around the world (Chaitanya et al., 2018), the HIrisPlex-S imputes skin, hair, and eye color probabilities from 41 SNPs (with overlaps: 6 for eye color, 22 for hair color, and 36 for skin color), with a high degree of accuracy. We focus on skin color because this trait is given primacy by proponents of race-associated phenotypic discrimination ("colorism") models (e.g., Dixon & Telles, 2017; Marira & Mitra, 2013), and because the other traits have more missing data, owing to

poor tagging of SNPs in some of the microarrays. A detailed discussion of the color measure and validity are provided in Supplementary File 2.

Cognitive ability-related polygenic scores

Lasker et al. (2019) detailed the rationale for eduPGS selection. Briefly, we validate four overlapping eduPGS from Lee et al. (2018): The eduPGS with all variants trained without the 23andMe cohort (with N = 7,762,369 SNPs in the present dataset), the multi-trait analysis of genome-wide association study (MTAG) eduPGS 10k SNPs (with N = 8,442 SNPs in the present dataset), the MTAG eduPGS lead SNPs (with N = 1,558 SNPs in the present dataset), and finally Lee et al.'s (2018) putatively causal variants (with N = 111 SNPs in the present dataset).

We used the MTAG 10k eduPGS for further analysis, since it had the highest validity in our two largest groups (European and African). This was as predicted, since both the MTAG eduPGS 10k SNPs and MTAG–lead SNPs are predicted to show moderate within-discovery population validity and moderate validity in non-discovery populations. This is because 'lead' or 'clumped' SNPs with greater statistical significance are more likely to be causal or very close to a causal variant, which are also more likely to be transethnically valid (Grinde et al., 2019; Marigorta & Navarro, 2013; Spencer, Cox & Walters, 2014; Wang & Teo, 2015). In contrast, the eduPGS score based on all variants (without 23andMe) is predicted to show high within discovery population validity, but poor validity in non-discovery populations owing to high LD decay bias as a result of including a large number of SNPs, irrespective of their significance (Zanetti & Weale, 2018). And because there are only 111 causal SNPs, the validities are predicted to be low in both discovery and non-discovery populations (e.g., Lasker et al., 2019).

In a supplementary analysis, we additionally examined the effect of computing MTAG 10k eduPGS using population-GWAS versus within-family beta weights for the SNPs. For the population-GWAS weights, we used Lee et al.'s (2018) published MTAG weights computed based on 1.1 million individuals. For the within family weights, we contacted Lee et al. (2018), who provided us with the within family weights from their analyses of 22,000 sibling pairs. For these analyses, we computed eduPGS using the subset of MTAG SNPs for which there were both within-family and population-GWAS weights.

Results

Descriptive statistics

Descriptive statistics for all groups appear in Table 1. Self-identifying European, European-African, African, and Hispanic Americans were ancestrally 98%, 79%, 20%, and 60% European, respectively. Among Hispanics, those who

identified as European were 81% ancestrally European, while those who identified as African were 29% ancestrally European. This difference in ancestry is consistent with those reported previously (e.g., Table S5; Bryc et al., 2015). The Hispanic Other group had the largest percentage of Amerindian ancestry, at 20%, which would be consistent with origins in Central or South America as opposed to the Caribbean (Table S5; Bryc et al., 2015).

Table 1. *Sample characteristics for the participants; mean ± standard deviation, and sample size in italics.*

	Age	% European	% Amerindian	Parental education	Mean *g*	Mean color
European	13.76 ± 3.64 *4937*	0.98 ± 0.07 *4939*	0.01 ± 0.04 *4939*	0.34 ± 0.96 *4909*	0.00 ± 1.01 *4914*	14.70 ± 3.84 *3862*
European-African	13.15 ± 3.58 *232*	0.79 ± 0.28 *232*	0.01 ± 0.03 *232*	0.16 ± 0.95 *230*	-0.14 ± 1.05 *228*	19.18 ± 7.61 *166*
African	14.08 ± 3.75 *2227*	0.20 ± 0.11 *2228*	0.02 ± 0.02 *2228*	-0.57 ± 0.77 *2180*	-1.01 ± 1.07 *217*	30.96 ± 5.87 *557*
Hispanic	13.59 ± 3.86 *514*	0.60 ± 0.29 *515*	0.11 ± 0.15 *515*	-0.32 ± 1.01 *507*	-0.57 ± 1.13 *506*	22.19 ± 7.59 *398*
HI_European	13.62 ± 4.09 *117*	0.81 ± 0.14 *117*	0.11 ± 0.09 *117*	-0.17 ± 1.03 *115*	-0.33 ± 1.17 *115*	19.45 ± 5.60 *80*
HI_Other	13.39 ± 3.94 *121*	0.60 ± 0.16 *122*	0.20 ± 0.18 *122*	-0.53 ± 0.98 *121*	-0.65 ± 1.17 *120*	21.14 ± 6.50 *103*
HI_African	13.88 ± 3.76 *151*	0.29 ± 0.20 *151*	0.06 ± 0.11 *151*	-0.52 ± 0.88 *148*	-0.84 ± 0.98 *146*	28.45 ± 6.78 *116*
Other, not Hispanic	13.42 ± 3.72 *125*	0.77 ± 0.27 *125*	0.09 ± 0.16 *125*	-0.02 ± 1.06 *123*	-0.39 ± 1.13 *125*	18.17 ± 6.31 *99*

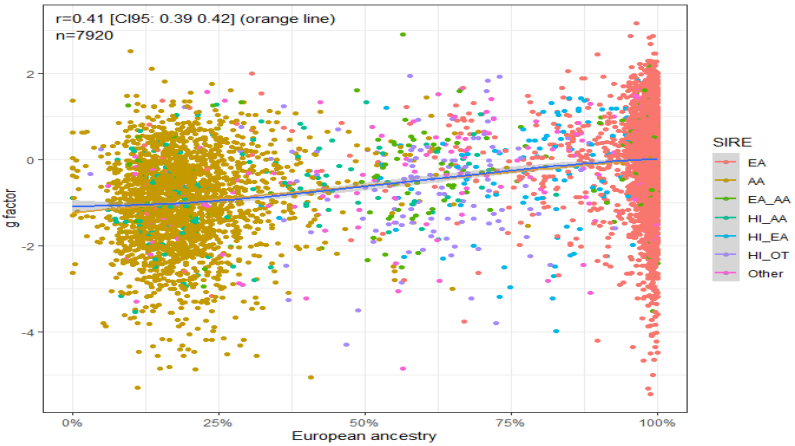

Figure 4. *Regression plot of the relation between g and European genetic ancestry in the combined sample*.

Note: EA = non-Hispanic European American; HI_AA = Hispanic African American, HI_EA = Hispanic European American, HI_OT = Hispanic Other, and Other = any other individuals who also identified as Hispanic.

The association between cognitive ability and European ancestry for the combined sample is depicted in Figure 4. A Bayesian Generalized Additive Model line (blue) is superimposed on a regression line (orange); moreover, self-identified racial groups are color coded. As can be seen, the association is nearly linear. Additionally, European ancestry is positively and significantly associated with *g* in the admixed SIRE groups ($r_{\text{European-African}}$ = .26, r_{African} = .084, r_{Hispanic} = .30), though not for European Americans (r_{European} = .02) among whom there is little non-European admixture.

Regression Analysis for Ancestry as a Predictor of g

Multiple regression analysis is preferable to bivariate analysis, since there is a possibility of confounding with social and environmental factors, particularly ones correlated with SIRE (Fang et al., 2019) and race-associated phenotype (Conley & Fletcher, 2017), and also of a non-independence of admixture components. As such, we relegate the bivariate results to Supplementary File 3.

In the initial regression analysis, shown in Table 2, we deal exclusively with those who identify as Hispanic American. In the first, Model (1a), we include only African and Amerindian ancestry as independent variables. In Model 1b, we add self-identified race to see if ancestry retains predictivity and has independent explanatory power. In Model 2, we approach the issue from the perspective of "colorism" (that is, discrimination based on race-associated phenotype, specifically color). As such, in Model 2a, we include only color. We then add ancestry in 2b to see which variables retain validity. In Model 3 we further add parental education to the model with color and ancestry. Note, all values except genetic ancestry are standardized. This allows the B coefficients for ancestry to be interpreted as a change in a standard deviation of cognitive ability going from 0% to 100% ancestry. Doing so allows results from different samples with different variances in ancestry to be compared on the same metric.

As can be observed, African ancestry retained validity when either SIRE or color were added. In contrast, with the inclusion of SIRE, Amerindian ancestry became a nonsignificant predictor. However, the direction and magnitude of the effect was as previously reported (Kirkegaard et al., 2019). As seen in Model 2a, darker color was a significant predictor of lower cognitive ability, consistent with previously reported findings (e.g., Hu et al., 2019). However, the effect became nonsignificant and the sign reversed with the inclusion of ancestry. This is consistent with the view that the association between color and ability is simply

indexing that between ancestry and ability (e.g., Hu et al., 2019; Lasker et al., 2019). Note, Model 1 and Model 2 above have different numbers, since there were fewer cases with color data; running the results with the same sample sizes (listwise deletion) did not lead to a difference of interpretation.

Table 2. *Regression analysis for ancestry as a predictor of g among Hispanic Americans with controls for skin color (Model 2b), and parental education (Model 2c) added; B coefficients with standard errors in parentheses.*

Predictor	Model 1	Model 1b	Model 2a	Model 2b	Model 3
Intercept	-0.140	0.285	-0.466	-0.037	0.116
	(0.085)	(0.198)	(0.056)	(0.098)	(0.207)
Ancestry: AFR	-1.205***	-1.663***		-1.483***	-1.378***
	(0.167)	(0.259)		(0.235)	(0.307)
Ancestry: AMER	-0.709*	-0.675		-0.859*	-0.624
	(0.337)	(0.361)		(0.407)	(0.432)
SIRE HI_European		-0.408*			-0.176
		(0.195)			(0.205)
SIRE HI_Other		-0.480			-0.111
		(0.173)			(0.181)
SIRE Other w/ HI		-0.366			-0.149
		(0.182)			(0.193)
Skin Color			-0.192**	0.136	0.093
			(0.060)	(0.077)	(0.076)
Parental Education					0.263***
					(0.051)
Adjusted R^2	0.091	0.099	0.024	0.111	0.165
N	506	506	391	391	338

Note: *$p < .05$, **$p < .01$, ***$p < .001$. Model 2a shows the results with color as an alternative predictor. AMR = Amerindian ancestry. AFR = African ancestry. Reference SIRE = HI_African.

Next, we repeat the analysis with the other groups (EA, EA-AA, and AA) added. The results appear in Table 3, and are descriptively similar to those for the Hispanic-only sample. As before, Amerindian ancestry loses validity on inclusion of SIRE. This time, however, the effect size of Amerindian ancestry is close to zero. Statistically, this happens because among Philadelphian Europeans, the association between Amerindian ancestry and ability is non-significantly positive (B = 0.500, S.E. = 0.391, N = 4914, p = .20), while the association between African ancestry and ability is significantly negative (B = -0.616, S.E. = 0.266, N = 4914, $p < .05$). Why this is the case is not clear. In Model 2a, color alone has validity, but it becomes nonsignificant and changes direction on inclusion of ancestry (2b). This remains the case when parental education is added in model 3.

Table 3. *Regression analysis for ancestry as a predictor of g in the combined sample with controls for skin color (Model 2) and parental education (Model 3) added; B coefficient with standard error in parentheses.*

Predictor	Model 1a	Model 1b	Model 2a	Model 2b	Model 3
Intercept	0.009	0.010	-0.291***	0.016	-0.089
	(0.015)	(0.015)	(0.013)	(0.023)	(0.023)
Ancestry: African	-1.274***	-0.988***		-1.262***	-0.785***
	(0.032)	(0.108)		(0.073)	(0.129)
Ancestry: American	-0.532*	0.043		-0.299	0.148
	(0.208)	(0.242)		(0.253)	(0.281)
SIRE: African		-0.248*			-0.107
		(0.088)			(0.095)
SIRE: European/African		0.042			0.044
		(0.072)			(0.081)
SIRE: HI_European		-0.263**			-0.091
		(0.100)			(0.115)
SIRE: HI_African		-0.208			-0.098
		(0.112)			(0.121)
SIRE: HI_Other		-0.482***			-0.127
		(0.109)			(0.115)
SIRE: Other, no Hisp.		-0.097			0.035
		(0.076)			(0.082)
Skin Color			-0.390***	-0.003	0.002
			(0.014)	(0.026)	(0.025)
Parental education					0.331***
					(0.014)
Adjusted R^2	0.165	0.168	0.122	0.164	0.237
N	7920	7920	5991	59991	5940

Note: *$p < .05$, **$p < .01$, ***$p < .001$. Model 1b shows the results with color as an alternative predictor. Reference SIRE = European.

Analyses of the Validity of eduPGS

Again, we relegate discussion of the bivariate results for the four educational PGS to Supplementary File 3. And we used the MTAG 10k eduPGS for further analysis. While this variable did not have the highest validity among Hispanics or among European-Africans, it had the highest predictive validity among the two largest groups, European and African Americans (Supplementary File 3; Table S3-S6). Additionally, Lee et al. (2018; Table 4c) shows that MTAG eduPGS has higher predictive validity than GWAS based eduPGS for cognitive ability in two samples (the National Longitudinal Study of Adolescent to Adult Health and the Health and Retirement Study). Thus we are justified in focusing on the MTAG-based eduPGS.

The relation between this eduPGS and cognitive ability for the four major groups is shown in Figure 5. The slopes (B) and intercepts based on the regression equation were: Hispanic ($B = 0.294$, S.E. = 0.043; Intercept = -0.291;

N = 506), European (*B* = 0.230, S.E. = 0.014; Intercept = -0.005; N = 4914), European-African (*B* = 0.284, 0.058; Intercept = 0.020; N = 228), and African American (*B* = 0.151, S.E. = 0.029; Intercept = -0.747; N = 2179). The slope for the Hispanic (*t*(5416) = 1.415, N.S.) and the European-African sample (*t*(5,138) = 0.905, N.S.) was not significantly steeper than the European one, though the power to detect a significant difference was relatively low. For the African American sample, the slope was significantly flatter (*t*(7,089) = 2.45, $p < 0.05$).

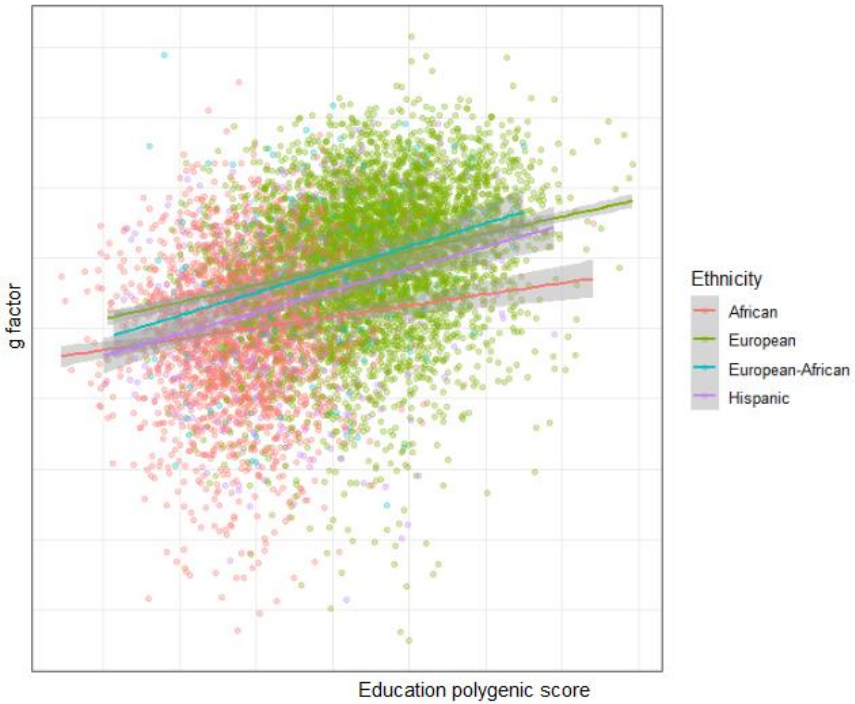

Figure 5. *Regression plot for the predictive validity of MTAG 10k eduPGS with respect to g in the Hispanic (purple), European (green), European-African (blue), and African American (red) samples.*

Because we saw previously that ancestry is a robust predictor of cognitive ability in the admixed samples, we next ran multivariate regression to test to what extent the association between eduPGS and ability may be due to confounding with either global ancestry or skin color. Model 1, Table 4, shows the effect of eduPGS alone for Hispanics. EduPGS is significantly related to *g* (*B* = 0.294, *N* =

506, $p < .001$). When adding ancestry covariates in Model 2, the association attenuates but remains significant ($B = 0.175$, $N = 506$, $p < .001$). Based on the formula provided by Clogg, Petkova and Haritou (1995) for comparing betas in nested models, the *z*-score for the difference was 1.79, which is statistically significant ($p = .037$; one-tailed). Additionally, adding color in Model 3 did not further attenuate this association ($B = 0.194$, $N = 391$, $p < .001$). Note, the association in Model 2 for Hispanics was not substantially different from that in the equivalent model for Europeans (Model $2_{European}$: $B = 0.230$, $N = 4{,}914$).

Table 4. *Regression results for the effect of eduPGS on cognitive ability among Hispanic Americans, B coefficient with standard error in parentheses.*

Predictor	Model 1	Model 2	Model 3
Intercept	-0.291 (0.062)	-0.115 (0.084)	-0.030 (0.096)
eduPGS	0.294*** (0.043)	0.175*** (0.051)	0.194*** (0.051)
Ancestry: African		-0.814*** (0.201)	-0.989*** (0.266)
Ancestry: Amerindian		-0.472 (0.341)	-0.492 (0.412)
Color			0.129 (0.076)
Adjusted R^2	0.084	0.110	0.141
N	506	506	391

Note: *$p <$ .05, **$p <$ .01, ***$p <$ .001

Next we ran the same analysis on the two other heavily admixed groups, European-African and African Americans. Table 5, Model 1 shows the effect of eduPGS alone for European-African Americans ($B = 0.284$, $N = 228$, $p < .001$). Adding ancestry in Model 2 seems to attenuate the association somewhat but the effect remains significant ($B = 0.215$, $N = 228$, $p < .01$). When color was added in Model 3, the association also remained significant ($B = 0.181$, $N = 164$, $p < .05$).

Table 5. *Regression results for the effect of eduPGS on cognitive ability among European-African Americans, B coefficient with standard error in parentheses.*

Predictor	Model 1	Model 2	Model 3
Intercept	0.020 (0.074)	0.026 (0.086)	0.026 (0.124)
eduPGS	0.284*** (0.058)	0.215** (0.072)	0.181* (0.081)
Ancestry: African		-0.521 (0.296)	-0.743 (0.437)
Ancestry: Amerindian		4.247 (2.632)	5.740* (2.765)
Color			0.140 (0.130)
Adjusted R^2	0.091	0.104	0.099
N	228	228	164

Note: *$p <$ .05, **$p <$ .01, ***$p <$ .001

For African Americans, shown in Table 6, the effect of eduPGS on *g* (*B* = 0.151, *N* = 2,179, *p* < .001) again seems to be attenuated slightly but remains significant (*B* = 0.126, *N* = 2,179, *p* < .001) with the inclusion of ancestry in Model 2. As seen in Model 3, adding color did not further attenuate this association (*B* = 0.133, *N* = 1,526, *p* < .001).

Table 6. *Regression results for the effect of eduPGS on cognitive ability among African Americans, B coefficient with standard error in parentheses, B coefficient with standard error in parentheses.*

Predictor	Model 1	Model 2	Model 3
Intercept	-0.747 (0.055)	-0.423 (0.171)	-0.229 (0.204)
eduPGS	0.151*** (0.029)	0.126*** (0.030)	0.133*** (0.037)
Ancestry: African		-0.478* (0.219)	-0.532 (0.278)
Ancestry: Amerindian		0.398 (1.198)	-0.455 (1.691)
Color			-0.066 (0.043)
Adjusted R^2	0.012	0.014	0.021
N	2179	2179	1526

Note: *p < .05, **p < .01, ***p < .001

Finally, for the combined sample, Model 1, Table 7, shows the effect of eduPGS alone. EduPGS is significantly related to *g* (*B* = 0.384, *N* = 7920, *p* < .001). When adding ancestry covariates in Model 2, the association is attenuated but remains significant (*B* = 0.221, *N* = 7,920, *p* < .001). Additionally, adding SIRE and color, in Model 4, did not further substantially attenuate this association (*B* = 0.218, *N* = 5,991, *p* < .001).

Table 7. *Regression results for the effect of eduPGS on cognitive ability for the combined sample.*

Predictor	Model 1	Model 2	Model 3	Model 4
Intercept	-0.322 (0.012)	-0.117 (0.016)	-0.117 (0.017)	-0.144 (0.025)
eduPGS	0.384*** (0.010)	0.221*** (0.013)	0.219*** (0.013)	0.218*** (0.014)
Ancestry: African		-0.795*** (0.042)	-0.511*** (0.110)	-0.501*** (0.136)
Ancestry: Amerindian		-0.251 (0.205)	0.214 (0.238)	0.430 (0.289)
SIRE: AA			-0.249** (0.087)	-0.224* (0.097)
SIRE: EA-AA			0.074 (0.071)	0.048 (0.084)
SIRE: Hispanic AA			-0.183 (0.110)	-0.213 (0.124)

Predictor	Model 1	Model 2	Model 3	Model 4
SIRE: Hispanic EA			-0.192 (0.098)	-0.180 (0.117)
SIRE: Hispanic Other			-0.406* (0.107)	-.309** (0.117)
SIRE: Other			-0.081 (0.074)	-0.048 (0.083)
Color				0.006 (0.026)
Adjusted R^2	0.159	0.195	0.196	0.196
N	7920	7920	7920	5991

Note: *p < .05, **p < .01, ***p < .001. SIRE (self-identified race/ethnicity) categories: AA = African American, EA = European American.

Path analysis

While fitting cross-sectional data to a path model cannot prove the causal assumptions, doing so can provide estimates of the effect magnitudes on the assumption that the model is correct (Bollen & Pearl, 2013). As such, we depict two sets of path model results fit with the **lavaan** R package (Rosseel, 2012). We limited the first path analysis to Hispanics to reduce transethnic bias in eduPGS validity. In the first model, we include European ancestry, color, and eduPGS as covariates. As color scores were not missing at random, we did not impute data, but rather handled missing data with listwise deletion. As with previous analyses, European ancestry was left unstandardized. The path model is shown in Figure 6. The detailed path estimates are shown in Supplementary File 4 (Table S1). In the model, eduPGS partially explained the association between European ancestry and cognitive ability. Independent of eduPGS, European ancestry was also strongly positively associated with cognitive ability. As expected, European ancestry was strongly negatively associated with darker color. However, darker skin color had no significant independent effect on cognitive ability. Moreover, the sign of the beta here was in the "wrong" direction relative to predictions from a colorism model. That is, darker color was unexpectedly associated with higher intelligence when controlling for ancestry. The model indicates that eduPGS is a plausible component; color, in contrast, does not seem to be a plausible mediator.

In Figure 7, we repeat this analysis with the complete sample. Since SIRE had little consistent effect, independent of ancestry, we do not include SIRE variables in the path analysis. The results are comparable, except that the beta for color is β = -.001 instead of β = .117 (both nonsignificant). Detailed results are provided in Table S2 of Supplementary File 4.

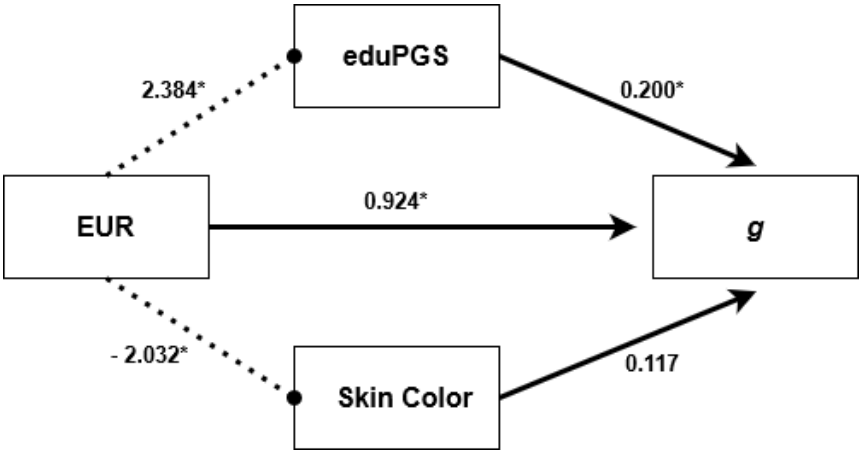

Figure 6. *Path diagram for the relation between European ancestry (EUR), skin color, education polygenic score (eduPGS), and g in the Hispanic American sample. N = 391. *Statistically significant at p < .001. Dashed-lines designate covariance.*

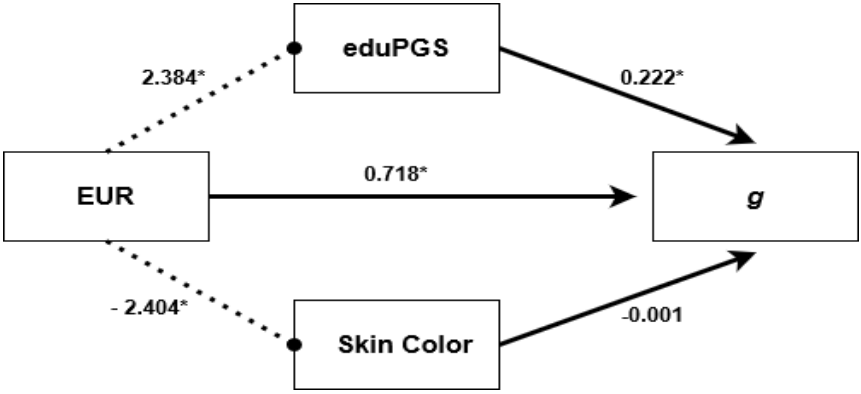

Figure 7. *Path diagram for the relation between European ancestry (EUR), skin color, education polygenic score (eduPGS), and g in the complete sample. N = 5,991. * Statistically significant at p < .001. Dashed lines designate covariance.*

As an alternative model, we include parental education instead of color. Color was dropped as it was not a significant predictor of *g* and because there were limited cases with color scores. While most parents are likely biological parents,

not all are, and so the adolescent's ancestry is not exactly equivalent to the midparent ancestry in this case. As such, we represent the relationship between the adolescent's European ancestry and their parent's education as covariance (indicated by dashed lines). For Hispanics, the path model is shown in Figure 8, and the detailed estimates appear in Table S3 of Supplementary File 4. In the model, eduPGS again partially explained the association between European ancestry and *g*. However, parental education was also a significant predictor. The covariance between parental education and adolescent eduPGS was significant. With the current data, however, it is not possible to disentangle the causal paths between European ancestry, parental education, adolescent eduPGS, and adolescent *g*. This is because adolescent eduPGS approximates biological mid-parent eduPGS. And it is expected that, within populations, mid-parent eduPGS will be causally related to parental educational levels which, in turn, will be genetically correlated with adolescent *g* (see, e.g., Trzaskowski et al., 2014). Nonetheless, this model also indicates that eduPGS plausibly has constitutive explanatory relevance regarding the relation between European ancestry and *g*.

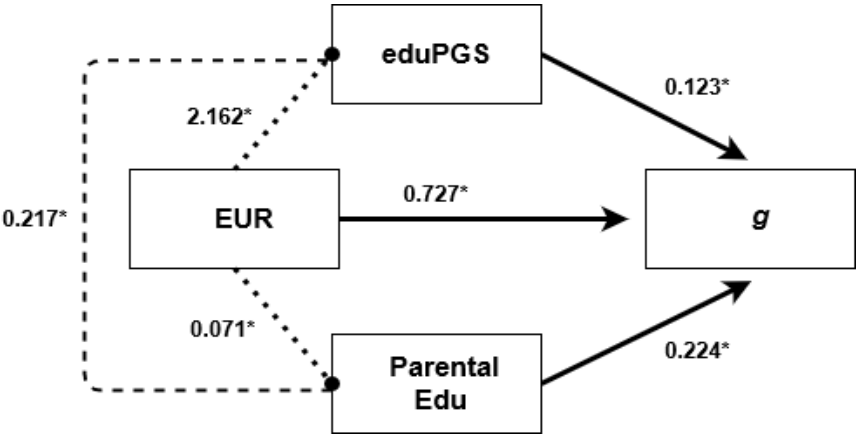

Figure 8. *Path diagram for the relation between European ancestry (EUR), education polygenic score (eduPGS), parental education, and g in the Hispanic American sample. N = 500. *Statistically significant at p < .001. Dashed lines designate covariance.*

Again, we repeat this analysis with the combined sample. The path model is shown in Figure 9, and the estimates appear in Table S4 of Supplementary File 4. The results are comparable.

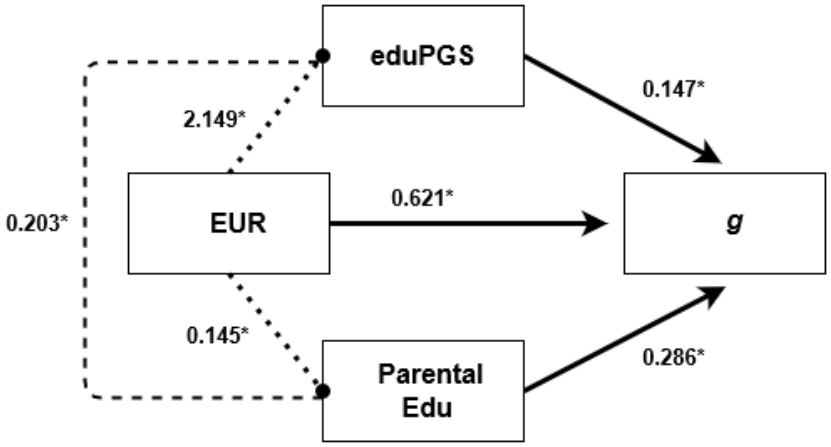

Figure 9. *Path diagram for the relation between European ancestry (EUR), education polygenic score (eduPGS), parental education, and g in the combined sample. N = 7846. * Statistically significant at p < .001. Dashed-lines designate covariance.*

The Spearman-Jensen hypothesis

If the association between eduPGS and cognitive ability is primarily genetic in nature, a Jensen effect between eduPGS effects and subtest *g*-loadings is expected, because genetic effects tend to primarily act through *g* (te Nijenhuis et al., 2019). Moreover, because the weak version of Spearman's hypothesis fits the data well, for both the African-European and Hispanic-European differences, it is expected that the magnitude of the differences on subtests will positively correlate with the subtest *g*-loadings (i.e., there will be a Jensen effect on score differences). Spearman's hypothesis can also be extended to associations with genetic ancestry. If the association between ancestry (not just SIRE) and cognitive ability is primarily due to ancestry-related differences in *g*, a Jensen effect with respect to ancestry would also be expected. The subtest correlations and the subtest group differences used for this analysis are reported in Tables S5-S6 of Supplementary File 4.

Table 8a,b shows the vector correlations (using unrounded vectors). The results based on the ten subtests with measurement invariance appear above the diagonal; while, those for all 15 subtests appear below. As can be seen, all associations are moderately to strongly positive. The strong association between *g*-loadings and eduPGS is consistent with the finding that Lee et al.'s (2018) eduPGS is associated with genetic *g* (de la Fuente, 2020). Consistent with other research, there is a strong Jensen effect on ethnic differences (te Nijenhuis, van

den Hoek & Dragt, 2019) and on ancestry-related differences within ethnic groups (Hu et al., 2019; Lasker et al., 2019). Generally, the effect of eduPGS, like that of ancestry and SIRE group differences, is pronounced on the most *g*-loaded and more heritable subtests, consistent with the predictions of a causal scenario.

Table 8a. *Results from Method of Correlated Vectors analysis. Correlations of subtest scores with g (g loading), with % European ancestry and with education polygenic score (PGS); size of standardized subtest gaps between SIRE groups (AA = African-American, EA = European-American, HI = Hispanic, EAA = EA-AA mixed), and subtest heritability (h^2). N = 10 subtests above diagonal (for the measurement invariant subtests), and 15 below diagonal (for all subtests).*

		g load	Anc. *r*	Anc. *r* (AA)	Anc. *r* (HI)	PGS *r*	PGS *r* (EA)
1.	*g* loading	1	.89	.85	.92	.91	.90
2.	Ancestry *r* (all)	.88	1	.69	.88	.94	.90
3.	Ancestry *r* (AA)	.82	.62	1	.78	.71	.70
4.	Ancestry *r* (HI)	.90	.91	.66	1	.89	.85
5.	PGS *r* (all)	.89	.96	.65	.92	1	.99
6.	PGS *r* (EA)	.89	.94	.65	.89	.99	1
7.	PGS *r* (AA)	.83	.63	.84	.77	.73	.74
8.	PGS *r* (HI)	.76	.80	.53	.90	.86	.82
9.	EA/EAAA gap	.63	.66	.67	.61	.62	.59
10.	EA/AA gap	.92	.98	.69	.92	.97	.95
11.	EA/HI gap	.87	.99	.64	.86	.94	.92
12.	h^2	.43	.40	.37	.42	.56	.59

Table 8b. *Results from Method of Correlated Vectors analysis. Correlations of subtest scores with g (g loading), with % European ancestry and with education polygenic score (PGS); size of standardized subtest gaps between SIRE groups (AA = African-American, EA = European-American, HI = Hispanic, EAA = EA-AA mixed), and subtest heritability (h^2). N = 10 subtests above diagonal (for the measurement invariant subtests), and 15 below diagonal (for all subtests).*

		PGS *r* (AA)	PGS *r* (HI)	EA/EAA gap	EA/AA gap	EA/HI gap	h^2
1.	*g* loading	.90	.83	.81	.94	.85	.37
2.	Ancestry *r* (all)	.73	.84	.85	.97	.99	.33
3.	Ancestry *r* (AA)	.89	.63	.87	.76	.66	.39
4.	Ancestry *r* (HI)	.85	.94	.80	.90	.82	.31
5.	PGS *r* (all)	.84	.87	.81	.96	.92	.55
6.	PGS *r* (EA)	.85	.82	.77	.93	.89	.60
7.	PGS *r* (AA)	1	.78	.78	.82	.68	.60
8.	PGS *r* (HI)	.72	1	.76	.87	.78	.39
9.	EA/EAAA gap	.60	.54	1	.87	.86	.43
10.	EA/AA gap	.72	.84	.68	1	.96	.47
11.	EA/HI gap	.61	.75	.69	.97	1	.38
12.	h^2	.64	.47	.30	.50	.42	1

Bias in Education-related PGS (eduPGS)

Evaluation of bias in the TCP sample

PGS may be biased due to the source population with which they were computed (i.e., ascertainment bias). There are a couple of obvious mechanisms by which this bias can occur. First, European-based eduPGS may be biased against non-Europeans due to the inclusion of European-specific variants (Thomson, 2019). These population-specific variants might have very low frequencies in non-European populations. Second, out-of-African based eduPGS may be biased against African populations due to an overrepresentation of derived (due to new mutations) versus ancestral (shared with other primates) variants in the out-of-Africa populations (Kim et al., 2018; Thomson, 2019). As a robustness check, we investigate both possibilities.

First, we computed MTAG eduPGS excluding variants with minor allele frequency (MAF) < .01 (leaving 7,636 overlapping variants) and < .05 (leaving 7,172 overlapping variants) among African lineages, using the 1000 Genomes reference samples to determine the African MAF. As a result, these eduPGS exclude variants not also present in African populations. As seen in Table 9, this exclusion had no substantive effect on the mean eduPGS differences between groups.

Table 9. *eduPGS for European, Hispanic, and African American participants, using either the complete 10k eduPGS, or the score computed after exclusion of SNPs with minor allele frequencies of < 1% or < 5% in Africans. Standard deviations appear in parentheses.*

SIRE group	N	eduPGS complete	eduPGS MAF >.01	eduPGS MAF >.05
European	4939	0.02 (0.99)	0.02 (0.99)	0.02 (0.99)
European-African	232	-0.57 (1.14)	-0.55 (1.15)	-0.54 (1.14)
African	2228	-1.76 (0.80)	-1.75 (0.82)	-1.71 (0.82)
Hispanic	515	-0.94 (1.12)	-0.94 (1.13)	-0.91 (1.12)
Hisp. – European	117	-0.56 (1.08)	-0.57 (1.08)	-0.55 (1.09)
Hisp. – African	151	-1.60 (0.92)	-1.59 (0.95)	-1.54 (0.95)
Hisp. – Other	122	-0.95 (1.03)	-0.94 (1.05)	-0.90 (1.04)
Other, not Hispanic	125	-0.51 (1.08)	-0.51 (1.08)	-0.49 (1.08)

Kim et al. (2018) demonstrated that when allelic risk scores are based on out-of-Africa populations, African populations show elevated frequencies of disease-associated loci for ancestral (shared with other primates) alleles, and reduced frequencies for derived (due to new mutations after the split with primates) alleles, even when there are no underlying trait differences. They conclude that "systematic allele frequency differences between populations need not be due to any underlying difference in risk" (p. 5) and propose corrections for bias due to ancestral versus derived allele status. To investigate this, we compute eduPGS by derived and ancestral status. To be clear, we computed one PGS with only those SNPs where the enhancing allele is derived, and then another PGS with only those SNPs where the enhancing allele is ancestral. In this case, risk alleles and trait-enhancing alleles are the same thing; medical versus cognitive GWAS studies just use different terminology.

The results are shown in Table 10. As seen, contrary to the findings of Kim et al. (2018), with eduPGS, non-European populations have both lower derived and ancestral eduPGS scores. Moreover, the differences are largest for the derived ones. As a result, when Kim et al.'s (2018, p. 12) correction is applied, the polygenic score gaps change little.

Table 10. *eduPGS for European, European-African, African, and Hispanic American participants, mean ± standard deviation.*

SIRE group	N	eduPGS derived	eduPGS ancestral	eduPGS corrected*
European-American	4939	0.02 ± 0.99	0.01 ± 1.00	0.02
European-African	232	-0.68 ± 1.15	-0.34 ± 1.12	-0.51
African-American	2228	-2.05 ± 0.85	-1.14 ± 0.79	-1.60
Hispanic	515	-1.05 ± 1.21	-0.66 ± 1.02	-0.85
Hisp.: European	117	-0.57 ± 1.07	-0.45 ± 1.05	-0.49
Hisp.: Other	122	-1.01 ± 1.03	-0.72 ± 1.08	-0.85
Hisp.: African	151	-1.84 ± 1.06	-1.06 ± 0.82	-1.45
Other, not Hispanic	125	-0.60 ± 1.20	-0.31 ± 0.97	-0.46

* Corrected following Kim et al.'s (2018) procedure, which involves adjusting the ancestral PGS down by 0.1902% and adjusting the derived PGS up by 0.1082% and then averaging the two. Kim et al. report that, in general, "ancestral risk [i.e., trait enhancing] alleles are found at 9.51% higher frequency in Africa, and derived risk [i.e., trait enhancing] alleles are found at 5.40% lower frequency in Africa" (p. 1). They then applied this general finding, with percentages doubled to account for diploidy, to specific traits. We do the same.

Additionally, we computed the validities to see if the corrections affected these. The validities by eduPGS are reported in Table 11. As seen, the validities for the different eduPGS were approximately the same for a given SIRE group.

Table 11. *Predictive validities of MAF ≥ 0.01, MAF ≥ 0.05, derived, and ancestral eduPGS among European, European-African, Hispanic, African and Hispanic Americans (N = sample size).*

SIRE group	N	MTAG 10k eduPGS	eduPGS MAF ≥ 0.01	eduPGS MAF ≥ 0.05	eduPGS derived	eduPGS ancestral
European	4914	.227	.227	.224	.207	.189
European-African	228	.308	.305	.298	.283	.263
African	2228	.112	.113	.115	.103	.090
Hispanic	515	.293	.288	.287	.292	.225

Finally, it has been argued that the eduPGS differences may be upwardly biased by using the discovery sample-based direction of effects for SNPs (Thompson, 2019). To investigate this we computed the betas for the MTAG SNPs, used as predictors of *g*, for European (*N* = 4,939) and African Americans (*N* = 2,228) separately. The SIRE specific betas are provided in the Supplementary Material. We then examined mean differences and validities for the variants which showed transracially concordant effects across ethnic groups as compared to those which showed discordant effects. African Americans had lower eduPGS than EA based on the concordant SNPs (-3.13 and 0.00, respectively), but slightly higher eduPGS based on the discordant SNPs (0.49 and 0.00, respectively). Moreover, among African Americans, the concordant SNPs showed higher validity than the discordant SNPs for *g*. Cross-validation confirmed that the concordant eduPGS were more predictive than the discordant eduPGS among African Americans. A possible interpretation is that the discordant eduPGS, which are less likely to be causal, contain more LD decay related effects. Regardless, there is no evidence, based on this sample, that the eduPGS differences are being inflated by the inclusion of SNPs with transethnically discordant effects as some have argued (Thompson, 2019).

Evaluation of bias in the 1000 Genomes samples

We additionally ran supplementary analyses which leveraged the 1000 Genomes data to explore the effects of score construction on the population differences. Methods and detailed results are reported in Supplementary File 5. First, we examined the effect of using population-GWAS vs. within-family PGS. In this analysis, we found that the African-European differences based on the within-family PGS were reduced in size but nonetheless large. Next, we examine the effect of using trans-ethnically concordant betas. The results were similar to those reported in our sample. Finally, we examined the predictive validity of different polygenic score constructions across populations. While the expected

associations remained positive, we found that the construction of eduPGS affected the magnitude of correlations, consistent with reports by others (e.g., Berg et al., 2019, Figure 1; Sohail et al., 2019, Figure 4). These results highlight Duncan et al.'s (2019) caution that different eduPGS can give markedly different results, rendering interpretation uncertain.

Discussion

To better understand how eduPGS function in admixed American populations, and to determine if there is a robust confound vs. causal problem, we examined the association between intelligence-related polygenic scores, global ancestry, and general cognitive ability in Hispanic and non-Hispanic European, European-African, and African American samples. In this sample, we were able to confirm full factorial invariance for the cognitive tests in the case of the European-Hispanic and European-African differences. And, previously, full factorial invariance had been found to hold along the full range of European ancestry (Lasker et al., 2019). Thus, the associations found reflect those with latent mental ability.

Among Hispanics and in the combined Hispanic and non-Hispanic American sample, the association between European/African ancestry and cognitive ability was robust to controls for SIRE, color, and parental education. For Amerindian ancestry, in both the Hispanic-only and the combined samples, the association with cognitive ability became nonsignificant when SIRE was added as a covariate. In the Hispanic-only sample, the effect remained directionally consistent with previously reported results (Kirkegaard et al., 2019; Warne, 2020). Statistically, the insignificance, in this case, was a result of the higher standard errors of Amerindian ancestry, as compared to African, which was due to the lower variance in Amerindian ancestry. Similarly, in a small, "non-Hispanic Caucasian" sample, with little variance in ancestry, European ancestry versus (apparently) Mexican/Mexican-American ancestry was only weakly and non-significantly associated with IQ ($r = .09$, $N = 120$; Wang et al., 2016). For the combined sample, however, the effect of Amerindian ancestry turned positive with SIRE controls. Statistically this was due to Amerindian ancestry being slightly positively correlated with general intelligence in the non-Hispanic White sample ($r = .014$; $N = 4914$, *N.S.*) and to the much larger non-Hispanic White than Hispanic sample size. Generally, the relation between cognitive ability and Amerindian ancestry needs to be better explored in predominantly European-Amerindian "Mestizo" samples. The results to date are ambiguous.

Here, explanations for the association between European ancestry and cognitive ability by confounding with either geography or racial phenotype (see, e.g., Conley & Fletcher, 2017) are not viable. This was a local population sample

from the Philadelphia area, so geographic confounding is not a substantial concern. We did find a significant negative association between cognitive ability and darker skin color. However, rather than the association between ancestry and cognitive ability being explained by color, the association between color and cognitive ability was statistically explained by ancestry.

These latter results are key in that there is a large literature on "colorism" which purports to demonstrate color or pigment-based discrimination by showing mere correlations between color and social outcomes (e.g., Marira & Mitra, 2013). Moreover, it has been argued by some that such discrimination might account for potential associations between ancestry and cognitive ability (e.g., Conley & Fletcher, 2017). However, our results concur with the competing distributional model described by Hu et al. (2019, Figure 1), in which the association between color and cognitive ability is a proxy (versus a cause) of that between ancestry and cognitive ability. This finding has implications for genetic research (Lawson et al., 2020), as it suggests that color-based discrimination is not likely an additional source of confounding.

We found that eduPGS was significantly associated with cognitive ability within the European, European-African, African, and Hispanic samples. For the MTAG 10k eduPGS, which we use for further analyses, the associations in the admixed populations were attenuated with the inclusion of ancestry. Regardless, the association remained significant and not substantially different from that in the European sample (B_{European} = .230 vs. $B_{\text{European-African}}$ = .215 and B_{Hispanic} = .175), except in the case of African Americans (B_{African} = .126) where, while statistically significant, the association was less than that for European Americans.

Moreover, path analysis indicated that eduPGS partially statistically explained the association between European ancestry and cognitive ability, whereas color did not. Additionally, the explanatory power of eduPGS was not fully accounted for by parental education, a variable which is both genetically and environmentally correlated with adolescent intelligence within groups and thus may be likewise between groups.

The eduPGS differences may be spurious owing to difficult to control ascertainment bias and confounding related to population stratification. For example, Berg et al. (2019) found attenuated effects for height PGS when applying UK versus pan-European-based PGS to Eurasian samples. The reason seems to be that controls for ancestry components do not always fully capture population structure effects, however, some confounding effects can be avoided by computing PGS using a more homogeneous population and then applying these to the more heterogeneous populations of interest (see: Berg & Coop, 2014). In our case, though, we start with eduPGS based on European origin samples and then apply these to samples of different continental ancestry. As

such, we already incorporated this component of Berg et al.'s (2019) analysis into ours.

We further investigated whether differences might be due to biasing effects of discovery population-specific variants, whether the allele associated with higher values of the trait is derived or ancestral, failure of SNP sign concordance between populations, and discrepancy between population-GWAS and within-family coefficients. The forms of ascertainment bias and confounding related to population stratification addressed by our procedures do not explain the group differences in eduPGS. However, it is beyond the scope of this paper to explore all forms of possible confounding.

The regression and path analysis results raise some possible concerns regarding the use and interpretation of eduPGS in admixed American populations. Ancestry covaries with trait and eduPGS scores. This could either be due to trait-relevant genetic differences between the ancestral groups of the admixed populations or to a mix of both environmental differences and also ascertainment bias and confounding related to population stratification. If the former, controlling for ancestry can attenuate the effect of eduPGS; however, if the latter, leaving ancestry unadjusted can inflate it.

Additionally, MCV indicated that the effect of eduPGS was strongly *g*-loaded, as was the effect of ancestry and group differences. This is consistent with all these effects acting primarily by way of *g*. These results for eduPGS are not self-evident. This is because it has been shown that some of the effect of eduPGS is a shared environmental effect (Domingue & Fletcher, 2019); however, the effect of adoption, a shared environmental effect, exhibits an anti-Jensen effect (te Nijenhuis, Jongeneel-Grimen & Armstrong, 2015). Thus, that eduPGS would act like typical genetic effects, and be most pronounced on *g*-loaded subtests, is not obvious. In addition to eduPGS, both ancestry and group differences exhibited Jensen effects, which is consistent with differences being primarily in *g* (i.e., Spearman's hypothesis). A practical implication of this is that good measures of *g* are needed to capture the full statistical effects of cognitive ability in context to research on eduPGS and related variables. Moreover, the positive manifold of Jensen effects is at least consistent with a causal model. To reconcile it with a confounding model one would need to additionally propose a mechanism by which environmentally induced phenotypic differences produced Jensen effects (e.g., Flynn, 2019).

The differences in eduPGS between ancestral groups are large enough that this is an issue which future research should try to resolve. While the magnitudes of differences for the true causal SNPs are unknown, magnitudes can be calculated for presently known education-related SNPs. We can use the fixation index, a measure of population differentiation, to do this. Supplementary File 6

shows the Fst for the SNPs from Lee et al. (2018) for 1000 Genomes super-populations. The Fst value between Europeans and Africans, which are the two main ancestral groups for the admixed populations here, is .1090 As detailed in Supplementary File 3, for typically reported heritabilities (h^2 = .5; Pesta et al., 2020; Polderman et al., 2015) this magnitude of population differentiation gives medium to large expected mean differences (i.e., when environments are equal), in this case, equivalent to d = 0.68. Thus, if transethnically unbiased eduPGS differences turn out to be commensurate with those based on Lee et al.'s (2018) eduPGS, medium to large phenotypic differences are expected under conditions of environmental equality. Controlling for ancestry might then bias eduPGS effects.

Related to this point, a reader suggested that we should run analyses to detect polygenic selection as done by Berg et al. (2019). However, whether differences are due to drift or selection is not necessarily relevant to whether there are genetic differences between populations. Moreover, given the expected differences discussed above, selection, in the form of stabilizing or convergent selection *between populations* is needed to show trait equality, not trait differences. That is, the evolutionary default or null expectation would be that trait-causing SNP frequency differences will be commensurate with random SNP ones, not that selection acted to homogenize differences in this particular trait (Edelaar & Björklund, 2011; Edge & Rosenberg, 2015; Leinonen et al., 2013; Rosenberg et al., 2019). As Rosenberg et al. note: "[P]henotypic differences among populations are predicted under neutrality to be similar in magnitude to typical genetic differences among populations" (*p*. 30). We are not aware of anyone who has established selection acting between populations such as to make causally-relevant genetic differences (related to education and intelligence) substantially smaller than expected by drift. Whether there was divergent selection is undetermined (e.g., Guo et al., 2018; Racimo, Berg & Pickrell, 2018).

Considering these results, it may be that individual ancestry tracks cognitive ability in admixed populations across the Americas. If so, proper interpretation of the predictive accuracy of eduPGS in American admixed populations will require a better understanding of the causal pathways underwriting this association. We argue that admixture mapping is the appropriate next step (for a rationale, see Kirkegaard et al., 2019). That said, different sub-ancestral components (e.g., North vs. South Amerindian / European, West versus East African) may yield different associations between global ancestry and cognitive ability, so additional global ancestry studies are warranted to better understand the pattern of effects.

Limitations

Owing to the sample sizes of the subgroups, some of the ancillary analyses, such as the comparison of eduPGS validities, were underpowered. Replications should be attempted with larger samples; until then, caution is warranted regarding interpretation. Importantly, we did not attempt to determine the cause of the covariance between eduPGS, ancestry, and *g*. We can only say that the cause is presently undetermined and that it needs to be resolved for a proper interpretation of the predictive accuracy of eduPGS in admixed American populations. While we suggest local admixture mapping as a way to narrow the uncertainty, further research should also explore other forms of confounding.

All supplemental files are available online at www.mankindquarterly.org.

References

Alexander, D.H., Novembre, J. & Lange, K. (2009). Fast model-based estimation of ancestry in unrelated individuals. *Genome Research* 19: 1655-1664.

Berg, J.J. & Coop, G. (2014). A population genetic signal of polygenic adaptation. *PLoS Genetics* 10(8): e1004412.

Berg, J.J., Harpak, A., Sinnott-Armstrong, N., Joergensen, A.M., Mostafavi, H., Field, Y., ... & Coop, G. (2019). Reduced signal for polygenic adaptation of height in UK Biobank. *Elife* 8: e39725.

Bollen, K.A. & Pearl, J. (2013). Eight myths about causality and structural equation models. In: S.L. Morgan (ed.), *Handbook of Causal Analysis for Social Research*, pp. 301-328. Dordrecht: Springer.

Bryc, K., Durand, E.Y., Macpherson, J.M., Reich, D. & Mountain, J.L. (2015). The genetic ancestry of African Americans, Latinos, and European Americans across the United States. *American Journal of Human Genetics* 96: 37-53.

Chaitanya, L., Breslin, K., Zuñiga, S., Wirken, L., Pośpiech, E., Kukla-Bartoszek, M., ... & Kayser, M. (2018). The HIrisPlex-S system for eye, hair and skin colour prediction from DNA: Introduction and forensic developmental validation. *Forensic Science International: Genetics* 35: 123-135.

Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M. & Lee, J.J. (2015). Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* 4(1): 7.

Clogg, C.C., Petkova, E. & Haritou, A. (1995). Statistical methods for comparing regression coefficients between models. *American Journal of Sociology* 100: 1261-1293.

Conley, D. & Fletcher, J. (2017). *The Genome Factor: What the Social Genomics Revolution Reveals about Ourselves, Our History, and the Future*. Princeton, NJ: Princeton Univ. Press.

Craver, C. (2007). Constitutive explanatory relevance. *Journal of Philosophical Research* 32: 3-20.

Daetwyler, H.D., Villanueva, B. & Woolliams, J.A. (2008). Accuracy of predicting the genetic risk of disease using a genome-wide approach. *PLoS One* 3(10): e3395.

de la Fuente, J., Davies, G., Grotzinger, A.D., Tucker-Drob, E.M. & Deary, I.J. (2020). A general dimension of genetic sharing across diverse cognitive traits inferred from molecular data. *Nature Human Behaviour* 5: 49-58.

Deary, I.J. (2010). Cognitive epidemiology: Its rise, its current issues, and its challenges. *Personality and Individual Differences* 49: 337-343.

Dixon, A.R. & Telles, E.E. (2017). Skin color and colorism: Global research, concepts, and measurement. *Annual Review of Sociology* 43: 405-424.

Domingue, B.W. & Fletcher, J. (2019). Separating measured genetic and environmental effects: Evidence linking parental genotype and adopted child outcomes. *Behavior Genetics* 50: 301-309.

Duncan, L., Shen, H., Gelaye, B., Meijsen, J., Ressler, K., Feldman, M., ... & Domingue, B. (2019). Analysis of polygenic risk score usage and performance in diverse human populations. *Nature Communications* 10(1): 3328.

Edelaar, P.I.M. & Björklund, M. (2011). If F_{ST} does not measure neutral genetic differentiation, then comparing it with QST is misleading. Or is it? *Molecular Ecology* 20: 1805-1812.

Edge, M.D. & Rosenberg, N.A. (2015). Implications of the apportionment of human genetic diversity for the apportionment of human phenotypic diversity. *Studies in History and Philosophy of Science Part C: Studies in History and Philosophy of Biological and Biomedical Sciences* 52: 32-45.

Fang, H., Hui, Q., Lynch, J., Honerlaw, J., Assimes, T.L., Huang, J., ... & DuVall, S.L. (2019). Harmonizing genetic ancestry and self-identified race/ethnicity in genome-wide association studies. *American Journal of Human Genetics* 105: 763-772.

Flynn, J.R. (2019). A response to te Nijenhuis et al. (2019). *Journal of Biosocial Science* 51: 913-916.

Genç, E., Schlüter, C., Fraenz, C., Arning, L., Metzen, D., Nguyen, H.P., ... & Ocklenburg, S. (2021). Polygenic scores for cognitive abilities and their association with

different aspects of general intelligence – a deep phenotyping approach. *Molecular Neurobiology*: https://link.springer.com/article/10.1007/s12035-021-02398-7.

Gottfredson, L.S. (2003). *g*, jobs, and life. In: H. Nyborg (ed.), *The Scientific Study of General Intelligence: Tribute to Arthur R. Jensen*, pp. 293-342. New York: Pergamon.

Grinde, K.E., Qi, Q., Thornton, T.A., Liu, S., Shadyab, A.H., Chan, K.H.K., ... & Sofer, T. (2019). Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. *Genetic Epidemiology* 43: 50-62.

Guo, G., Lin, M.J. & Harris, K.M. (2019). Socioeconomic and genomic roots of verbal ability. *bioRxiv*. https://www.biorxiv.org/content/biorxiv/early/2019/02/10/544411.full.pdf

Guo, J., Wu, Y., Zhu, Z., Zheng, Z., Trzaskowski, M., Zeng, J., ... & Yang, J. (2018). Global genetic differentiation of complex traits shaped by natural selection in humans. *Nature Communications* 9: 1865.

Gur, R.C., Richard, J., Hughett, P., Calkins, M.E., Macy, L., Bilker, W.B., ... & Gur, R.E. (2010). A cognitive neuroscience-based computerized battery for efficient measurement of individual differences: Standardization and initial construct validation. *Journal of Neuroscience Methods* 187: 254-262.

He, Q. & Li, J.J. (2021). Factorial invariance in hierarchical factor models of mental disorders in African American and European American youths. *Journal of Child Psychology and Psychiatry* 62: 289-298.

Horta, B.L., Hartwig, F.P. & Victora, C.G. (2018). Breastfeeding and intelligence in adulthood: Due to genetic confounding? *Lancet Global Health* 6(12): e1276-e1277.

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*. Westport, CT: Praeger.

Kim, M.S., Patel, K.P., Teng, A.K., Berens, A.J. & Lachance, J. (2018). Genetic disease risks can be misestimated across global populations. *Genome Biology* 19: 179.

Kirkegaard, E.O., Woodley of Menie, M.A., Williams, R.L., Fuerst, J. & Meisenberg, G. (2019). Biogeographic ancestry, cognitive ability and socioeconomic outcomes. *Psych* 1(1): 1-25.

Lasker, J., Pesta, B.J., Fuerst, J.G. & Kirkegaard, E.O. (2019). Global ancestry and cognitive ability. *Psych* 1(1): 431-459.

Lawson, D.J., Davies, N.M., Haworth, S., Ashraf, B., Howe, L., Crawford, A., ... & Timpson, N.J. (2020). Is population structure in the genetic biobank era irrelevant, a challenge, or an opportunity? *Human Genetics* 139: 23-41.

Lee, J.J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, M., ... & Fontana, M.A. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature Genetics* 50: 1112-1121.

Leinonen, T., McCairns, R.S., O'Hara, R.B. & Merilä, J. (2013). Q_{ST}–F_{ST} comparisons: Evolutionary and ecological insights from genomic heterogeneity. *Nature Reviews Genetics* 14: 179-190.

Marigorta, U.M. & Navarro, A. (2013). High trans-ethnic replicability of GWAS results implies common causal variants. *PLoS Genetics* 9(6): e1003566.

Marira, T.D. & Mitra, P. (2013). Colorism: Ubiquitous yet understudied. *Industrial and Organizational Psychology* 6(1): 103-107.

Martin, A.R., Gignoux, C.R., Walters, R.K., Wojcik, G.L., Neale, B.M., Gravel, S., ... & Kenny, E.E. (2017). Human demographic history impacts genetic risk prediction across diverse populations. *American Journal of Human Genetics* 100: 635-649.

Mollon, J., Knowles, E.E., Mathias, S.R., Gur, R., Peralta, J.M., Weiner, D.J., ... & Glahn, D.C. (2021). Genetic influence on cognitive development between childhood and adulthood. *Molecular Psychiatry* 26: 656-665.

Moore, T.M., Reise, S.P., Gur, R.E., Hakonarson, H. & Gur, R.C. (2015). Psychometric properties of the Penn Computerized Neurocognitive Battery. *Neuropsychology* 29: 235-246.

Pearl, J. (2014). Interpretation and identification of causal mediation. *Psychological Methods* 19: 459-481.

Pesta, B.J., Kirkegaard, E.O., te Nijenhuis, J., Lasker, J. & Fuerst, J.G. (2020). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence* 78: 101408.

Plomin, R. & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics* 19: 148-159.

Plomin, R., DeFries, J.C., Knopik, V.S. & Neiderhiser, J.M. (2014). *Behavioral Genetics*, 6th ed. New York: Worth Publishers.

Polderman, T.J., Benyamin, B., De Leeuw, C.A., Sullivan, P.F., van Bochoven, A., Visscher, P.M. & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics* 47: 702-709.

Rabinowitz, J.A., Kuo, S.I.C., Felder, W., Musci, R.J., Bettencourt, A., Benke, K., ... & Kouzis, A. (2019). Associations between an educational attainment polygenic score with educational attainment in an African American sample. *Genes, Brain and Behavior* 18(5): e12558.

Racimo, F., Berg, J.J. & Pickrell, J.K. (2018). Detecting polygenic adaptation in admixture graphs. *Genetics* 208: 1565-1584.

Rosenberg, N.A., Edge, M.D., Pritchard, J.K. & Feldman, M.W. (2019). Interpreting polygenic scores, polygenic adaptation, and human phenotypic differences. *Evolution, Medicine, and Public Health* 2019: 26-34.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software* 48(2): 1-36. http://www.jstatsoft.org/v48/i02/

Roth, P.L., van Iddekinge, C.H., DeOrtentiis, P.S., Hackney, K.J., Zhang, L. & Buster, M.A. (2017). Hispanic and Asian performance on selection procedures: A narrative and meta-analytic review of 12 common predictors. *Journal of Applied Psychology* 102: 1178-1202.

Satterthwaite, T.D., Connolly, J.J., Ruparel, K., Calkins, M.E., Jackson, C., Elliott, M.A., ... & Qiu, H. (2016). The Philadelphia Neurodevelopmental Cohort: A publicly available resource for the study of normal and abnormal brain development in youth. *Neuroimage* 124: 1115-1119.

Sirin, S.R. (2005). The relationship between socioeconomic status and school outcomes: Meta-analytic review of research. *Review of Educational Research* 75: 417-453.

Sniekers, S., Stringer, S., Watanabe, K., Jansen, P.R., Coleman, J.R., Krapohl, E., ... & Amin, N. (2017). Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence. *Nature Genetics* 49: 1107-1112.

Sohail, M., Maier, R.M., Ganna, A., Bloemendal, A., Martin, A.R., Turchin, M.C., ... & Sunyaev, S.R. (2019). Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. eLife 8: e39702.

Spencer, A.V., Cox, A. & Walters, K. (2014). Comparing the efficacy of SNP filtering methods for identifying a single causal SNP in a known association region. *Annals of Human Genetics* 78: 50-61.

te Nijenhuis, J., Jongeneel-Grimen, B. & Armstrong, E.L. (2015). Are adoption gains on the g factor? A meta-analysis. *Personality and Individual Differences* 73: 56-60.

te Nijenhuis, J., van den Hoek, M. & Dragt, J. (2019). A meta-analysis of Spearman's Hypothesis tested on Latin-American Hispanics, including a new way to correct for imperfectly measuring the construct of *g*. *Psych* 1(1): 101-122.

te Nijenhuis, J., Choi, Y.Y., van den Hoek, M., Valueva, E. & Lee, K.H. (2019). Spearman's hypothesis tested comparing Korean young adults with various other groups of young adults on the items of the Advanced Progressive Matrices. *Journal of Biosocial Science* 51: 875-912.

Thompson, J. (2019, May 2). Piffer kicks against the pricks. Available online: http://www.unz.com/jthompson/piffer-kicks-against-the-pricks/

Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., ... & Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence* 42: 83-88.

US Census Bureau (2016). *American factfinder – Results*. US Department of Commerce, Economics and Statistics Administration. Retrieved from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

Wang, F.L., Pandika, D., Chassin, L., Lee, M. & King, K. (2016). Testing the relations among family disorganization, delay discounting, and adolescent alcohol use: A genetically informed study. *Alcoholism: Clinical and Experimental Research* 40: 846-856.

Wang, X. & Teo, Y.Y. (2015). Trans-ethnic fine-mapping of rare causal variants. In: E. Zeggini & A. Morris (eds.), *Assessing Rare Variation in Complex Traits*, pp. 253-261. New York: Springer.

Warne, R.T. (2020). Continental genetic ancestry source correlates with global cognitive ability score. *Mankind Quarterly* 60: 400-422.

Woodley of Menie, M.A.W., Fernandes, H.B. & Hopkins, W.D. (2015). The more *g*-loaded, the more heritable, evolvable, and phenotypically variable: Homology with humans in chimpanzee cognitive abilities. *Intelligence* 50: 159-163.

Ylikoski, P. (2013). Causal and constitutive explanation compared. *Erkenntnis* 78(2): 277-297.

Zanetti, D. & Weale, M.E. (2018). Transethnic differences in GWAS signals: A simulation study. *Annals of Human Genetics* 82(5): 280-286.