Twin Research and Human Genetics (2022), **25**, 10–23
doi:10.1017/thg.2022.3

CAMBRIDGE UNIVERSITY PRESS

Article

Social Adversity Reduces Polygenic Score Expressivity for General Cognitive Ability, but Not Height

Mateo Peñaherrera-Aguirre[1,*], Michael A. Woodley of Menie[2,*], Matthew A. Sarraf[3] and Kevin M. Beaver[4]

[1]Department of Psychology, University of Arizona, Tucson, AZ, USA, [2]Vrije Universiteit Brussel, Center Leo Apostel for Transdisciplinary Studies, Brussels, Belgium, [3]Independent Researcher, USA and [4]College of Criminology and Criminal Justice, Florida State University, Tallahassee, FL, USA

Abstract

It has been hypothesized that even 'perfect' polygenic scores (PGSs) composed of only causal variants may not be fully portable between different social groups owing to gene-by-environment interactions modifying the expression of relevant variants. The impacts of such interactions involving two forms of social adversity (low socioeconomic status [SES] and discrimination) are examined in relation to the expressivity of a PGS for educational attainment composed of putatively causal variants in a large, representatively sampled and genotyped cohort of US children. A relatively small-magnitude Scarr–Rowe effect is present (SES × PGS_{EDU} predicting General Cognitive Ability [GCA]; $sR = .02$, 95% CI [.00, .04]), as is a distinct discrimination × PGS_{EDU} interaction predicting GCA ($sR = -.02$, 95% CI [−.05, 00]). Both are independent of the confounding main effects of 10 ancestral principal components, PGS_{EDU}, SES, discrimination and interactions among these factors. No sex differences were found. These interactions were examined in relation to phenotypic and genotypic data on height, a prospectively more socially neutral trait. They were absent in both cases. The discrimination × PGS_{EDU} interaction is a co-moderator of the differences posited in modern versions of Spearman's hypothesis (along with shared environmentality), lending support to certain environmental explanations of those differences. Behavior-genetic analysis of self-reported discrimination indicates that it is nonsignificantly heritable ($h^2 = .027$, 95% CI [−.05, .10]), meaning that it is not merely proxying some underlying source of heritable phenotypic variability. This suggests that experiences of discrimination might stem instead from the action of purely social forces.

Keywords: Expressivity; gene-by-environment interaction; general cognitive ability; Scarr-Rowe effect; social adversity

(Received 3 February 2022; accepted 3 February 2022; First Published online 8 April 2022)

It is well known that polygenic scores (PGSs) for educational attainment, and more broadly for other traits, are sensitive to cross-ancestry comparisons, being generally less predictive of relevant trait variance among populations that are ancestrally more distant from the populations in which the PGSs were originally estimated (e.g., Belsky et al., 2018; Duncan et al., 2019; Guo et al., 2019; Lee et al., 2018; Weissbrod et al., 2021). One explanation of this 'portability problem' is that it results from linkage disequilibrium (LD) decay, whereby genetic variants that are noncausal of the relevant phenotype, but are nevertheless in LD with causal variants in the discovery sample, will be 'flagged' as though they were causal. Once the same PGS is estimated in a more distantly related population, the apparent predictivity of the PGS diminishes, as many of its constituent variants are now no longer in LD with the causal variants owing to recombination (Zanetti & Weale, 2018). Rabinowitz et al. (2019) proposed an interesting hypothesis, specifically that some of the PGS portability problems between socially identified racial and/or ethnic (SIRE[1]) groups in the US context, specifically in comparisons involving African American and White groups using PGSs for educational attainment and related phenotypes, might stem from gene-by-environment interactions. They state that '[e]nvironmental experiences such as poverty, racial discrimination and attending under-resourced schools may influence whether genetic propensity for educational attainment confers benefits for achievement and college attendance' (p. 2).[2]

Rabinowitz et al. (2019) offer the Scarr–Rowe effect (or 'hypothesis' or 'interaction') as evidence for the plausibility of their idea. This effect is a gene-by-environment interaction characterized by reductions in the heritability of measures of general cognitive ability (GCA), such as IQ, among those exposed to social adversity associated with low socioeconomic status (SES). This is thought to in turn reduce GCA among individuals from low-SES backgrounds by effectively preventing them from realising their full genetic potential for GCA (Rowe et al., 1999; Scarr-Salapatek, 1971; Turkheimer et al., 2003; Turkheimer et al., 2009). Scarr-Salapatek (1971, see also Scarr, 1981) and others (e.g., Flynn, 2018; Jensen, 1968; Lewontin, 1970, 1976) also considered gene-by-environment interactions involving forms of social adversity specific to SIRE groups to be prospective moderators of heritability, and therefore potential causes of SIRE group differences in cognitive performance means.

One meta-analytic report indicates that the heritability of GCA varies as a function of level of SES, at least in light of apparently robust Scarr–Rowe effects in US cohorts (Tucker-Drob & Bates,

Author for correspondence: Michael A. Woodley of Menie, Email: Michael.Woodley@vub.be

*Indicates joint first authorship, where author order was determined alphabetically.

Cite this article: Peñaherrera-Aguirre M, Woodley of Menie MA, Sarraf MA, and Beaver KM. (2022) Social Adversity Reduces Polygenic Score Expressivity for General Cognitive Ability, but Not Height. *Twin Research and Human Genetics* **25**: 10–23, https://doi.org/10.1017/thg.2022.3

2015).[3] Consistent with this finding, Guo and Stearns (2002) report evidence for the Scarr–Rowe effect in the Add Health dataset; however, they also identified an additional statistically significant negative impact on the proportion of heritable variance associated with Peabody Picture Vocabulary Test (PPVT) performance stemming from participant African American (relative to White) designation, even after controlling for a number of SES variables. In relation to this, they noted that:

> [O]ne assumption for such consideration is that the environments for intellectual development of African American children are still more disadvantaged than those for white children after controlling for the included family SES variables. Recent work has documented the discrimination African Americans face, decades after the civil rights movements . . . , and the legacy discrimination has left hinders the development of the middle-class vocabularies that would allow children to score highly on the PPVT. (p. 905)

A key prediction therefore is that even if hypothetical PGSs could be constructed using only causal variants, such PGSs would still not be fully portable between populations (or even among subgroups of individuals of the same population exposed to different environments), as heritability-attenuating gene-by-environment interactions associated with racial discrimination, poverty and other forms of social adversity would cause them to fail to predict equal amounts of phenotypic variance when examined in the context of differentially socially advantaged groups. Some evidence consistent with this observation comes from two studies (Woodley of Menie et al., 2018; Woodley of Menie et al., 2021) that were able to recover Scarr–Rowe effects using childhood SES as a moderator of the expressivity of educational attainment PGSs on measures of GCA in two large genetically informed US samples (sourced from the Wisconsin Longitudinal and Health and Retirement Studies, respectively). In both cases, the effects were *relatively small* by the standards of psychological science (specifically ≤.l0, the average meta-analytic effect magnitude in psychology being around .20; Gignac & Szodorai, 2016), but they were statistically significant and robust to the use of different sets of controls and measurement models.

In the current study, an attempt is made to test the hypothesis of Rabinowitz et al. (2019) in relation to the impacts of two forms of social adversity, low SES and racial and ethnic discrimination, on the expressivity of a PGS for educational attainment (predicting GCA) using a sample composed of individuals from multiple SIRE groups. The study will also examine whether sources of social adversity associated with SIRE, such as racial and ethnic discrimination, modify PGS expressivity on GCA, independent of the effects of SES, which would effectively replicate Guo and Stearns' (2002) findings of distinct effects of these on the heritability of a cognition measure. The current effort uses individuals sourced from the US Adolescent Brain and Cognitive Development (ABCD) dataset (Jernigan et al., 2018), which is broadly representative of the US population, and contains a variety of SES variables, in addition to (multiple) measures self-reported racial and ethnic discrimination. These data are available for large numbers of (specifically parentally identified) Black, White and Hispanic children (Jernigan et al., 2018). As the sample is genotyped, PGSs predictive of GCA and other traits can also be estimated. As an additional step, in order to reduce confounding effects stemming from cross-ancestry comparisons, the interaction models are estimated using an educational attainment PGS generated via the new 'PolyFun-Pred' (POLYgenic FUNctionally-informed fine mapping PREDiction) method, which uses very large numbers of single nucleotide polymorphisms (SNPs) coupled with fine mapping-based functional area weighting to reduce between-population biases due to LD decay, via the identification of prospectively causal variants (Weissbrod et al., 2021). Furthermore, as a null-test of the model, another PolyFun-Pred-type PGS is estimated for a prospectively more socially neutral trait, specifically height, where gene-by-environment interactions involving social adversity might be substantially weaker to nonexistent, at least in relatively modernized social contexts where serious malnutrition is uncommon.[4]

Interestingly, thus far no research on gene-by-environment interactions has attempted to determine whether discrimination (measured directly, rather than crudely proxied via the use of SIRE, as in Guo & Stearns, 2002) uniquely impacts, through adverse gene-by-environment interactions, GCA (contrast this with the extensive body of research into the adverse effects of low SES on the additive heritability of GCA; e.g., Tucker-Drob & Bates, 2015). This is an interesting oversight, as (within the US context) self-reported experiences of racial discrimination are quite prevalent, especially among minority populations (e.g., Boutwell et al., 2017; Lee et al., 2019). Moreover, there has been considerable academic debate concerning the mechanisms through which discrimination might adversely affect GCA among minority populations in the US (for discussion on the plausibility of skin reflectance or color-based prejudice mediated by socio-structural factors, see Cooper, 2005 and Rowe, 2005).

To bring some clarity to this debate, discrimination itself will be examined using behavior-genetic techniques (on data from the ABCD twin subsample) in order to determine whether it is merely proxying some underlying heritable phenotypic variation (with which it would then be associated), or whether it functions instead as a purely environmental, and therefore almost certainly social, force. In the case of the former, it would be expected that it exhibits significantly nonzero heritability, but in the latter case it would be expected that discrimination is not significantly heritable, being instead a function of shared and/or nonshared environmentality. To the authorship's knowledge, no *direct* behavior-genetic analysis of self-reported discrimination has ever been carried out, despite the fact that such efforts have the potential to substantially clarify the empirical foundations of arguments about the means by which discrimination might affect GCA.

Methods

Cohort

The subjects employed here were sourced from the ABCD (v. 3.01) data release. This large longitudinal study involves collaboration between 21 different sites across the US. ABCD sourced its data from over 11,000 children, aged 9-10 years, who mostly came from a mixture of public and private elementary school contexts. ABCD offers data that are broadly representative of this age range of healthy children from the US via probabilistic sampling and the exclusion of those exhibiting severe medical and psychiatric conditions (Jernigan et al., 2018). The sample employed in the current analysis includes those who fall into the Black, White and Hispanic SIRE categories. Although Black and White SIRE groups are most germane to historical discussion on the impacts of different forms of social adversity on the heritability of GCA (e.g., Guo & Stearns, 2002; Scarr-Salapatek, 1971), those who were identified as ethnically Hispanic were also included, as such individuals are also likely to have experienced higher levels of discrimination, relative to non-Hispanic Whites. In the case of ABCD, an 18-item questionnaire is given to the parents asking 'What race do you consider the child to

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

be? Please check all that apply'. This allowed for participants to be assigned to either the Black (element name: *demo_race_a_p__11*, $N = 1282$), or White (*demo_race_a_p__10*, $N = 5413$) racial categories. The following additional question was asked: 'Do you consider the child Hispanic/Latino/Latina?'. This allowed individuals to be assigned to the Hispanic ethnic category (*demo_ethn_v2*, $N = 1185$). Participants were coded as Hispanic if they were categorized as ethnically Hispanic, irrespective of their racial identification (Black or White). Only non-Hispanic participants were coded as either Black or White for the purposes of the current study. This yielded a total sample size of 7,880 individuals with complete data on all of the following variables.

Variables

Genetic data. ABCD provides microarray data on genotyped individuals, where imputation was performed employing the TOPMed imputation server. The following preimputation steps were utilized (by ABCD). First, PLINK v1.9 (Purcell et al., 2007) was used to calculate allele frequencies. Second, .bim files were checked against two reference datasets, the Haplotype Reference Consortium and the 1000 Genomes project. Third, PLINK v1.9 was used to convert everything to VCF files; and fourth, checkVCF.py was used to determine whether this conversion was successful. ABCD then uploaded the VCF files to the TOPMed Imputation Server, where imputation was performed using Eagle v2.4 phasing and mixed ancestry. Postimputation quality filtering was performed by excluding SNPs with an imputation quality score of $r^2 < .40$ using bcftools (v.1.7.2; Li, 2011). The total number of variants that remained after filtering was 103,382,718 for 11,101 participants.

PGSs. Two PGSs were generated, one for educational attainment (PGS_{EDU}) and a second one for height (PGS_{HEIGHT}). Both were generated using PolyFun-Pred, a novel method for polygenic prediction, developed by Weissbrod et al. (2021), which they describe as follows:

> PolyFun-Pred is a new predictor that leverages genome wide functionally informed fine-mapping . . . to estimate posterior mean causal effects (instead of tagging effects . . .) for all SNPs with European MAF≥0.1% (accounting for MAF-dependent architectures . . . 18 million SNPs in this study) by applying PolyFun + SuSiE35 to European training data across 2,763 overlapping 3Mb loci. Leveraging fine-mapped posterior mean causal effects for cross-population polygenic prediction aims to address LD differences between populations; to our knowledge, the application of PolyFun + SuSiE (or any other fine-mapping method) to polygenic prediction has not previously been explored. (p. 2–3)

Weissbrod et al. (2021) provide methods for estimating PolyPred-Fun type PGSs for a large number of traits with summary statistics, based on the UKBioBank. They make all weighting terms for all analyzed SNPs in the UKBioBank publicly available, which allows for the equivalent variants in ABCD to be identified and weighted accordingly for use in generating the relevant equivalent PGSs in this cohort. The PGSs were scored using PLINK (v1.90b6.17). For educational attainment, there was an average (across participants) of 1,013,790.05 ($N = 7880$; $SE = .002$) variants that overlapped with those identified by Weissbrod et al. (2021). For height, there was an average of 850,474.59 ($N = 7880$; $SE = .002$) overlapping variants.

Ancestral principal components. All SNPs were used (in PLINK) to generate 10 ancestral principal components (PCs). These correspond to genome wide patterns of covariation among the frequencies of alleles, which capture, in the form of distinct dimensions, different ancestral population structures among (in this case) the ABCD participants. These can be used as simple and effective controls for the effects of population stratification on the results of studies employing the outputs of GWASs, such as PGSs in mixed ancestry samples (Price et al., 2006). Population stratification can potentially confound the results of such studies by generating spurious patterns of association between PGSs and their outcomes.

GCA. The ABCD contains data on 11 cognitive ability measures. The first seven comprise the NIH Toolbox® cognitive battery. This battery includes picture vocabulary, flanker, list sorting, card sorting, pattern comparison, picture sequence memory, and oral reading recognition subtests. Also included are the matrix reasoning subtest from the Wechsler Intelligence Scale for Children, the efficiency score from the Little Man test, and the Rey Auditory Verbal Learning (immediate recall and delayed recall memory) Tasks (RAVLT; for details of these, see Luciana et al., 2018 and Thompson et al., 2019).

To construct a GCA factor for the participant pool, we employed unit-weighted factor estimation (Gorsuch, 1983). This form of exploratory factor scoring involves standardizing each participant's score on each subtest and then averaging across them to generate a unit-weighted factor. Unit-weighted estimation has the advantage of potentially yielding latent variables that generalize to a far greater degree across samples that are heterogeneous with respect to sampling error than the results of other dimension-reduction techniques (Gorsuch, 1983). Part-whole correlations between each subtest and the unit-weighted factor yield factor loadings. Squaring the factor loadings and then averaging them allows for the proportion of variance accounted for by that factor to be estimated. In the total participant pool, GCA was found to account for 32.7% of the variance ($N = 7880$). Factor loadings (λ) ranged from .64 (in the case of RAVLT short-term memory and RAVLT long-term memory) to .43 (in the case of the Little Man test). The full set of factor loadings is reported in Table 9 in the Results section. Factorial invariance with respect to both SIRE and sex was determined via the estimation of Tucker's congruence coefficients (CC), which index factor similarity. Based on the unit-weighted factor loadings, it was found that GCA exhibited strong consistency in terms of factor structure across sex and SIRE groups (average CC = .997). Coefficients of .95 or greater indicate virtual identicality among factors (Lorenzo-Seva & Ten Berge, 2006).

The predictive validity of PGS_{EDU} on GCA was examined disaggregated by SIRE in order to determine the patterns of portability (Table 1). Prior to entry into the model, PGS_{EDU} was residualized for the fixed effects of the following confounds: family ID (*rel_family_ID*); within- and between-family singleton, twin, and triplet status (*rel group id* and *rel ingroup_order*); relationship of the participant in their family (*rel_relationship*); and collection site ID (*site id 1*). In the combined sample, these confounds account for 8% of the variance in PGS_{EDU}. Each main effect of PGS_{EDU} on GCA was then estimated with controls for the first 10 ancestral PCs (results for these not shown). This (and subsequent) portability analysis was conducted using SAS v. 9.4.

Running regressions using PGS_{EDU} predicting GCA in each SIRE group separately (controlling for the 10 PCs in each case) indicates that despite the use of a technique that is designed to increase the portability of PGSs derived from one population to others, cross-ancestry comparisons employing this PGS

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

Table 1. Results of running the regression analyses predicting GCA using PGS_{EDU} separately by SIRE group, along with Bonferroni-corrected significances of differences

SIRE group	*N*	β	Lower 95% CI	Upper 95% CI
White (non-Hispanic)	5413	.250	.222	.278
Black (non-Hispanic)	1282	.130	.044	.216
Hispanic	1185	.229	.034	.295
Comparison	**Between β (*z* value)**	**Lower 95% CI**	**Upper 95% CI**	***p* value**
Black-Hispanic	−2.59	−.170	−.020	.0095*
Black-White	−4.10	−.180	−.060	.0000*
Hispanic-White	−.71	−.080	.040	.4753

Note: *Significant based on Bonferroni adjusted $p = .0166$.

Table 2. Results of running the regression analyses predicting height using PGS_{HEIGHT} separately by SIRE group, along with Bonferroni-corrected significances of differences

SIRE group	*N*	β	Lower 95% CI	Upper 95% CI
White (non-Hispanic)	5413	.338	.314	.362
Black (non-Hispanic)	1282	.260	.192	.328
Hispanic	1185	.279	.222	.336
Comparison	**Between β (z-value)**	**Lower 95% CI**	**Upper 95% CI**	***p* value**
Black-Hispanic	−.525	−.090	.052	.5999
Black-White	−2.86	−.133	−.024	.0042*
Hispanic-White	−2.11	−.115	−.004	.0349

still exhibit significant portability problems when the regression parameters among the SIRE groups are compared. Portability between the White and Hispanic groups appears to be high, however, since there is no (Bonferroni-adjusted) significant difference in the variance in GCA explained by the PGS between the two groups.

Height. Three height measurements were collected per participant (*anthro_1_height_in* to *anthro_3_height_in*, all measurements are in inches). The average of all three measures (where available) was used to assign a phenotypic height value to each participant. As with PGS_{EDU}, the portability of PGS_{HEIGHT} was examined in relation to the three SIRE groups (results presented in Table 2).

As with PGS_{EDU}, prior to entry into the model, PGS_{HEIGHT} was residualized for the fixed effects of the following confounds: family ID; within- and between-family singleton, twin, and triplet status; relationship of the participant in their family; and collection site ID. In the combined sample, these confounds account for 2% of the variance in PGS_{HEIGHT}. Each main effect of PGS_{HEIGHT} on GCA was then estimated with controls for the first 10 ancestral PCs (results for these not shown).

Running regressions using PGS_{HEIGHT} predicting height in each SIRE group separately (controlling for the 10 PCs in each case) indicates that cross-ancestry comparisons are associated with fewer significant portability problems (relative to PGS_{EDU}), with PGS_{HEIGHT} accounting for similar amounts of variance in phenotypic height in all three SIRE groups. Only the Black-White portability difference reached (Bonferroni-adjusted) significance.

Table 3. Unit-weighted factor loadings of six SES indicators. All λ values are statistically significant

Element name	SES indicator	Variable type	SES λ (95% CI)
family_income_dfct1 to family_income_dfct7	Financial adversity (reversed)	Continuous	.614 (.601, .627)
mother_education, father_education	Education attainment	Continuous	.568 (.553, .582)
dem_12	Parents marital status	Binary	.696 (.685, .707)
nsc_p_ss_mean_3_items	Neighborhood safety	Continuous	.663 (.651, .675)
Empcur	Employment status	Binary	.534 (.519, .549)
sub_income	Income	Continuous	.757 (.748, .766)

Socioeconomic status. Six SES measures were chosen to capture a broad range of environments relevant to participant exposure to childhood economic and social adversity. These include seven items assessing financial adversity in different contexts, which were reverse-scored and then summed into an index score. Additionally included is a measure of neighborhood safety (a three-item index asking about different aspects of neighborhood safety, scaled 1–5), parental marital status (recoded such that 1 = *married*, and 0 = *any other arrangement*) and employment status (1 = *employed*, 0 = *not currently employed*), parental educational attainment (both parents, measures highest level of educational attainment achieved rescaled in order to generate scores ranging from 0 to 18 approximate years of attained education as follows: Never attended/kindergarten only = 0; 1st grade = 1; 2nd grade = 2; 3rd grade = 3; 4th grade = 4; 5th grade = 5; 6th grade = 6; 7th grade = 7; 8th grade = 8; 9th grade = 9; 10th grade = 10; 11th grade = 11; 12th grade; High school graduate, GED or equivalent diploma = 12; Associate degree: occupational program, associate degree: academic program = 14; Bachelor's degree = 16; Master's degree, professional school and doctoral degree = 18), and family income (parentally reported total dollar amount income over the past 12 months, recoded using the lowest reported amount within a range of earnings [except for 1, where the low end was $0, so $500 was subtracted from the high end] as follows: 1 = $4500, 2 = $5000, 3 = $12,000, 4 = $16,000, 5 = $25,000, 6 = $35,000, 7 = $50,000, 8 = $75,000, 9 = $100,000, and 10 = $200,000). A general SES dimension was extracted from among these items using unit-weighted estimation. All variables along with their element names and associated factor loadings (ranging from .534 to .757) are listed in Table 3. CCs revealed virtual identicality in factor structures across sexes and all three SIRE groups (average CC = .996).

Discrimination. ABCD administered a variety of items at the year 1 follow-up mark to determine participant self-reported experience of various forms of discrimination. Six of these items clearly tap some aspect of racial, ethnic, or national-origin discrimination. These are listed in Table 4 along with their variable codes, whether they are continuous (Likert) or binary. Also displayed are the results of unit-weighted estimation, coupled with the use of the

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

Table 4. Unit-weighted factor loadings of six discrimination indicators. All λ values are statistically significant

Element name	Question	Variable type	Discrimination λ (95% CI)
dim_yesno_q1	In the past 12 months, have you felt discriminated against: because of your race, ethnicity, or color? Definition of ethnicity: groups of people who have the same customs, or origin.	Binary	.595 (.581, .609)
dim_yesno_q2	In the past 12 months, have you felt discriminated against: because you are (or your family is) from another country?	Binary	.435 (.417, .452)
dim_matrix_q1	How often do the following people treat you unfairly or negatively because of your ethnic background? Teachers.	Continuous	.668 (.656, .680)
dim_matrix_q2	How often do the following people treat you unfairly or negatively because of your ethnic background? Other adults outside school.	Continuous	.701 (.690, .712)
dim_matrix_q3	How often do the following people treat you unfairly or negatively because of your ethnic background? Other students.	Continuous	.745 (.736, .754)
dim_matrix_q4	I feel that others behave in an unfair or negative way toward my ethnic group.	Continuous	.705 (.694, .715)

Table 5. Self-reported double log discrimination means disaggregated by SIRE group. Bonferroni-adjusted multiple comparisons examining the difference between SIRE groups

SIRE group	*N*	Mean	*SD*	*SE*
Black	1282	.512	1.308	.037
Hispanic	1185	.183	1.161	.034
White	5413	−.161	.813	.011
Comparison	**Cohen's *d***	**Lower 95% CI**	**Upper 95% CI**	***p* value**
Black-Hispanic	.265	.186	.345	<.0001*
Black-White	−.725	−.787	−.663	<.0001*
Hispanic-White	−.388	−.452	−.325	<.0001*

Note: *Bonferroni adjusted $p = .0166$.

mice package (van Buuren & Groothuis-Oudshoorn, 2011) in R with a maximum of 50 iterations to account for missingness, which was used to extract a common discrimination dimension from among the six items (factor loadings ranged from .435 to .745). CCs revealed virtual identicality in factor structures across sexes and the three SIRE groups (average CC = .998).

The discrimination variable was found to be strongly positively skewed, with large numbers of individuals having reported no experience of discrimination (skewness = 3.390). As gene-by-environment interaction effects involving continuous measures are (partially) scale dependent, meaning that they can result from nonnormality and can be attenuated once appropriate transformations are made to the data (Martin, 2000), the discrimination measure was double log-transformed, which reduced its skewness to 1.815 (bringing it into the −2 to +2 range considered generally acceptable; e.g., George & Mallery, 2010). The SES factor was also examined for problematic skewness, but none was found (skewness = −1.034).

In order to determine whether discrimination was functioning as expected, means of (double log-transformed) discrimination along with standard errors and (multiple-comparison-corrected) significances of the differences among the means were estimated for all SIRE groups. These are presented in Table 5.

As anticipated, discrimination means varied significantly between SIRE groups, such that Black participants reported experiencing discrimination to a significantly greater degree than did Hispanic participants, who in turn reported experiencing significantly greater discrimination than White participants.

Measurement model

General linear model. In the current analysis, GCA is used as the dependent variable. All GLMs were conducted in *UniMult 2* (for documentation on an earlier version of this software, see Gorsuch, 1991) with the Type-I sum of squares procedure, which uses hierarchical partitioning of variance to estimate the effects of independent variables based on their hypothesized sequence of impacts on the dependent variable. In order to accurately report standardized effect sizes using this GLM approach, semipartial regression coefficients are estimated. These are presented along with 95% CIs, *F*-ratio test statistics, and significance levels. Hierarchical (rather than simultaneous) estimation of effects in models containing interaction terms is theoretically reasonable, as relevant interaction terms should be shown to be independent of main effects in addition to confounding interaction terms, after the estimation of the former (Nelder, 1994; Rodriguez et al., 1995). Consistent with this, in population genetics, it is standard to partition phenotypic trait variance as follows:

$$V_P = V_G + V_E + V_{GE}$$

Where V_P is the phenotypic trait variance, V_G is the variance associated with all genetic influences, V_E is the variance associated with all environmental influences, and V_{GE} is the variance associated with gene-by-environment interactions (Singh & Singh, 2018).

In line with the above equation, the predictors of GCA are entered into the model as follows. First, the main effects of genetic confounds and influences are estimated in the following sequence: (1) the main effects of the 10 ancestral PCs, (2) the main effect of (residualised) PGS_{EDU}. Then the main effects of environmental influences are estimated in the following sequence: (3) the main effect of SES, (4) the main effect of discrimination. Finally, the interactions between the various genetic and environmental factors are estimated in the following sequence: (5) the interactions between the ancestral PCs and PGS_{EDU}, (6) the interactions between the ancestral PCs and SES, (7) the Scarr–Rowe effect, which is operationalised as the interaction between PGS_{EDU} and SES, (8) the interactions between the ancestral PCs and discrimination, and finally (9) the discrimination × PGS_{EDU} interaction. The model is run separately on the combined sample, and for each sex, in order to determine whether sex differences in the interactions of interest are present. The same analysis (usnig the combined sample) is conducted using height as the criterion

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

variable and PGS_{HEIGHT} instead of PGS_{EDU}. This yielded three different GLM models in total. In all cases, only the main effects of the PGSs, SES, and discrimination, and the SES × PGS and discrimination × PGS interactions are reported. Effect sizes associated with ancestry and other interactions are not reported.

Finally, all variables were standardized prior to entry into the regression (see Woodley of Menie et al., 2021 for the use of the same approach to estimating the Scarr–Rowe effect using PGS data in the Health and Retirement Study).

CPEM. A second method is also used to test the robustness of the interactions of interest, specifically the Continuous Parameter Estimation Model (CPEM; Gorsuch, 2005). This method uses the dot-product of the participant's standardized dependent and the independent variable to generate a continuous parameter estimate (CPE) of the covariance among these, which can then be correlated with another variable in order to examine potential moderation effects. The method has the advantage of utilizing fewer model degrees of freedom than the more conventional approach to estimating two-way interactions — as the interaction term (the CPE) can be directly regressed against its moderator, without the requirement for estimating main effects. The major disadvantage to using this method is that the resultant effect size will be confounded with unmodeled effects, thus CPEM effect sizes tend to be larger than those generated using two-way interaction models. This approach has been used (along with a two-way interaction model) to test for the presence of the Scarr–Rowe effect in an analysis employing data from the Wisconsin Longitudinal Study (Woodley of Menie et al., 2018). In this study, a CPE was generated via the dot-product of the participant's standardized PGS_{EDU} and their IQ scores. This was then regressed against a composite measure of their childhood socioeconomic conditions. The resultant effect size was positive, meaning that as childhood SES improved, the participant's PGS covaried more strongly with their IQ scores — suggesting increased expressivity of the former onto the latter in response to improved SES. Here, CPEs will be constructed using the participant's PGS_{EDU} along with their GCA scores, yielding two separate effect sizes, one with SES as the criterion and another with discrimination as the criterion. If the latter effect is present, it is expected that the resultant effect size will be negative in sign — meaning that PGS_{EDU} is less expressive on GCA when self-reported discrimination is high.

Behavior-genetic analyses. ABCD contains data on both monozygotic (MZ) and dizygotic (DZ) twins, along with full siblings (with ages for correction), covering all SIRE groups currently considered. These will be used to estimate the additive heritability (A) and shared (C) and nonshared (E) environmentality of self-reported discrimination. This is to test whether discrimination is a proxy measure for heritable phenotypic variation of some sort, which could occur for any number of possible reasons, or whether it is purely environmental. If discrimination is associated with some underlying phenotype (a major candidate being skin reflectance; Cooper, 2005; Rowe, 2005), which is the actual factor that discriminatory behavior targets, then it might be expected that participant experiences of discrimination will exhibit a heritability >0%, as it will reflect (by statistical association) the heritability of this underlying trait. In the case of skin color, there is evidence that heritability is very high. Paik et al. (2011), for example, found in one human population that a constitutive skin color measure exhibited an additive heritability (h^2) of .82. As a reference trait, the heritability of GCA will also be estimated. These analyses are conducted using the behavior genetics R packages *lavaan 0.6–9* (Rosseel, 2012) and *pacman 5.1* (Rinker & Kurkiewicz, 2017).

Co-moderation analysis. An additional analysis is conducted in order to determine whether and how the GCA loadings among cognitive ability subtests moderate the effect magnitudes of (any) Scarr–Rowe and discrimination × PGS_{EDU} interactions on the cognitive ability measures. For this analysis, the GLM model is estimated using each cognitive ability subtest separately (yielding 11 potential Scarr–Rowe and discrimination × PGS_{EDU} interactions). Unit-weighted estimation is then used to composite these effect-size vectors into a common factor along with the vectors of the subtest GCA loadings, White–Black–Hispanic performance differences (expressed as *r*-statistics with weighted averaging) for each subtest, PGS_{EDU}-by-subtest associations, and subtest additivity (A), shared environmentality (C), and nonshared environmentality (E) components estimated using the twin (plus full siblings) subset (also rescaled as *r*-statistics). This configuration allows for a determination of whether or not gene-by-environment interactions might contribute to the differences posited by modern versions of Spearman's hypothesis, which hold that the magnitude of the differences in ability means between SIRE groups is positively moderated by GCA[5] (Jensen, 1980, 1998; Spearman, 1927; see also the more contemporary work of Frisby & Beaujean, 2015; te Nijenhuis & van den Hoek, 2016; te Nijenhuis et al., 2019).

The results of vector correlation analyses involving clustering among multiple correlated vectors have been offered as evidence for the so-called hereditarian hypothesis[6] on the basis that the magnitudes of the impacts of 'genetic' factors (such as inbreeding depression) on IQ battery subtests have been found to cluster along with the vectors of subtest GCA loadings and (specifically) White-Black mean performance differences, whereas the vectors of probably largely, possibly entirely, environmental effects, such as the Flynn effect (the secular increase in IQ test scores across decades), do not cluster with these effects (Rushton, 1999). In an item-level analysis of Raven's Progressive Matrices 'puzzles', Rushton et al. (2007) observed that in comparisons involving multiple SIRE groups, the group difference magnitudes in performance across items were correlated (positively) with their heritabilities, but not with their environmentalities (after controlling for item reliability and pass-rate variance). This finding has been interpreted as offering additional evidence for the hereditarian hypothesis (see also discussion in Rushton & Jensen [2010] and Warne [2021]). The hereditarian interpretation of results such as these has been critiqued, however (Flynn, 1999; Nisbett, 2009; Wicherts & Johnson, 2009, for discussion on the problem of factorial identification in the results of vector correlation analyses see Ashton & Lee, [2005]).

Utilizing a similar subtest-level co-moderation approach to Rushton (1999), and incorporating behavior-genetic variance components (A, C, and E), hereditarian predictions can be easily tested, as (on Rushton et al.'s [2007] assumptions) it would be expected that Scarr–Rowe and discrimination × PGS_{EDU} interactions, because they are thought to represent environmental influences on gene expression, should be *more* pronounced on subtests that are *less* GCA loaded, *less* additively heritable, *less* strongly associated with PGS_{EDU} (this essentially being a weaker measure of subtest heritability), and *less* predictive of SIRE group differences (these four vectors should exhibit strong, positive intercorrelations by contrast), and correspondingly *more* strongly associated with various forms of environmentality. Deviations

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

Table 6. General linear models predicting GCA using PGS_{EDU} residuals, SES, double log-transformed discrimination, and the corresponding interactions on GCA after controlling for the influence of ancestral principal components and confounding interactions (corresponding effect sizes not shown). All variables are standardized prior to regression. Results are for the combined sample, males, and females. Also presented are the results of two CPEM analyses, one examining the Scarr–Rowe effect, and a second examining the discrimination × PGS_{EDU} interaction

Combined sample					
Variables	Effect size (*sR*)	95% CI	*F* value	*df1/df2*	*p* value
Z-Residualized PGS_{EDU}	.19	.17, .21	356.61	1/7,834	<.0001
Z-Socioeconomic status	.18	.16, .20	323.54	1/7,834	<.0001
Z-Discrimination	−.08	−.10, −.06	61.21	1/7,834	<.0001
Z-Residualized PGS_{EDU}* Z-Socioeconomic status	.02	.00, .04	4.17	1/7,834	.04
Z-Residualized PGS_{EDU}* Z-Discrimination	−.02	−.05, .00	5.55	1/7,834	.02
Omnibus	R^2	95% CI	*F* value	*df1/df2*	*p* value
For all independent variables	.19	.18, .20	41.88	45/7,834	<.0001
Variables (predicting SES)	Effect size *(sR)*	95% CI	*F* value	*df1/df2*	*p* value
CPE[Z-Residualized $PGS_{EDU\ NET\ OF\ PCs}$ * Z-GCA]	.04	.02, .06	14.49	1/7,878	.0001
Variables (predicting discrimination)	Effect size *(sR)*	95% CI	*F* value	*df1/df2*	*p* value
CPE[Z-Residualized $PGS_{EDUNET\ OF\ PCs}$ * Z-GCA]	−.03	−.05, −.01	8.15	1/7,878	.004
Female subgroup					
Variables	Effect size *(sR)*	95% CI	*F* value	*df1/df2*	*p* value
Z-Residualized PGS_{EDU}	.23	.21, .28	13.47	1/3,702	<.0001
Z-Socioeconomic status	.20	.21, .28	12.98	1/3,702	<.0001
Z-Log discrimination	−.07	−.13, −.06	−5.46	1/3,702	<.0001
Z-Residualized PGS_{EDU}* Z-Socioeconomic status	.03	.00, .06	4.38	1/3,702	.04
Z-Residualized PGS_{EDU}* Z-Discrimination	−.02	−.05, .01	2.47	1/3,702	.12
Omnibus	R^2	95% CI	*F* value	*df1/df2*	*p* value
For all independent variables	.22	.20, .24	23.96	45/3,702	<.0001
Male subgroup					
Variables	Effect size *(sR)*	95% CI	*t* value	*df1/df2*	*p* value
Z-Residualized PGS_{EDU}	.16	.13, .19	127.16	1/4,086	<.0001
Z-Socioeconomic status	.17	.14, .20	140.50	1/4,086	<.0001
Z-Discrimination	−.08	−.12, −.05	5.27	1/4,086	<.0001
Z-Residualized PGS_{EDU}* Z-Socioeconomic status	.01	−.02, .04	.58	1/4,086	.44
Z-Residualized PGS_{EDU}* Z-Discrimination	−.02	−.05, .01	2.68	1/4,086	.10
Omnibus	R^2	95% CI	*F* value	*df1/df2*	*p* value
For all independent variables	.18	.16, .19	19.78	45/4,086	<.0001

from this pattern would be problematic for the hereditarian model, since, if found consistently, they would indicate failure of one of its key lines of supporting evidence (Warne, 2021).

Results

PGS_{EDU} GLMs

Table 6 presents the results of the three GLMs (involving PGS_{EDU}; combined sample, males, and females) and the CPEM analyses.

In the main analysis, the overall model fit (based on the Multiple *R*) is significant, with all independent variables accounting for 19% of the variance in GCA. PGS_{EDU} was a significant predictor of GCA (independent of the PCs). SES and discrimination also had significant independent main effects on GCA in the theoretically anticipated directions. The model also revealed a significant Scarr–Rowe effect (specifically a positively signed interaction between PGS_{EDU} and SES; graphed in Figure 1a). An independent, significant discrimination × PGS_{EDU} interaction is also present in these data (graphed in Figure 1b).

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

Table 7. General linear model evaluating the influence of PGS_{HEIGHT} residuals, SES, double log-transformed discrimination, and the corresponding interactions on height controlling for ancestral principal components and confounding interactions (results not shown). All variables are standardised prior to regression

Variables	Effect size *(sR)*	95% CI	*F* value	*df1/df2*	*p* value
Z-Residualized PGS_{HEIGHT}	.32	.30, .34	900.96	1/7,834	<.0001
Z-Socioeconomic status	.01	−.01, .03	1.27	1/7,834	.26
Z-Discrimination	−.03	−.06, .01	10.62	1/7,834	.001
Z-Residualized PGS_{HEIGHT} * Z-Socioeconomic status	−.01	−.04, .01	2.06	1/7,834	.15
Z-Residualized PGS_{HEIGHT} * Z-Discrimination	.01	−.02, .03	.34	1/7,834	.56
Omnibus	R^2	95% CI	*F* value	*df1/df2*	*p* value
For all independent variables	.12	.11, .14	23.71	45/7,834	<.0001

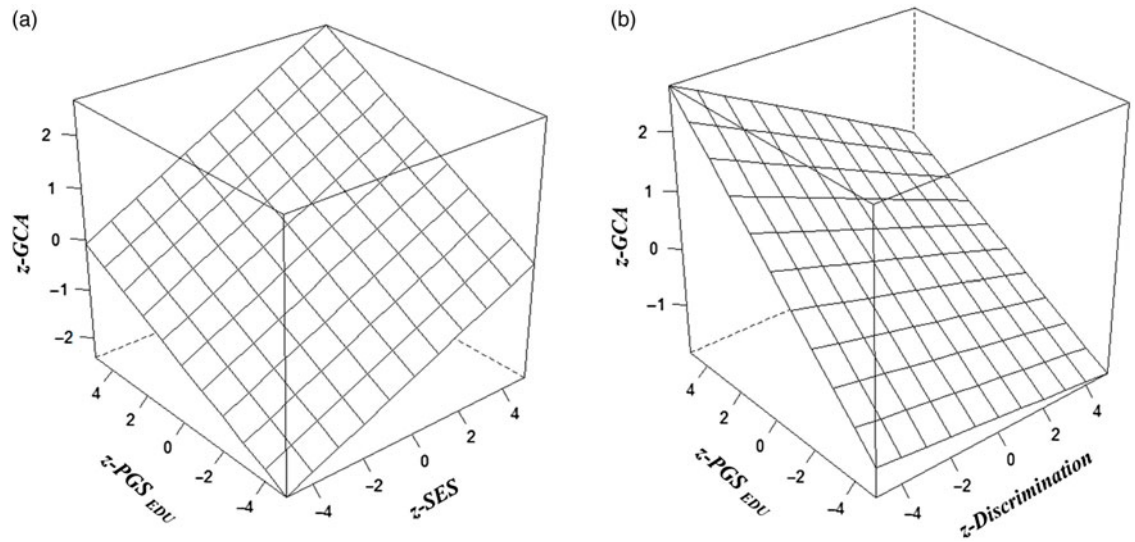

Fig. 1 (a). Regression plane plot visualising the interactions between PGS_{EDU} and SES on GCA scores (the Scarr-Rowe effect) **(b).** Regression plane plot visualising the interactions between PGS_{EDU} and discrimination on GCA scores (the discrimination×PGS_{EDU} interaction).

The CPEM analyses revealed (1) a significant Scarr–Rowe effect and (2) a significant discrimination × PGS_{EDU} interaction (Table 6). In terms of the sex-disaggregated samples, the analysis revealed a significant Scarr–Rowe effect in the female subgroup but not in the male subgroup. The models did not detect discrimination×PGS_{EDU} interactions in either the female or male subgroups (Table 6). There were no significant sex differences associated with either effect (Scarr–Rowe effect: $z = .89$, $p = .3752$; discrimination × PGS_{EDU} interaction: $z = .00$, $p = 1.000$).

It should be noted that the effect sizes associated with the interaction terms are *relatively small* in magnitude (≤.10; Gignac & Szodorai, 2016), although this is anticipated given that gene-by-environment interactions are expected to be much smaller than main effects (McGue & Carey, 2017). Moreover, these estimates are very similar in terms of magnitude to those found in PGS-based studies of the Scarr–Rowe effect in other US samples, based on the use of both two-way interaction and CPEM modeling approaches (Woodley of Menie et al., 2018; Woodley of Menie et al., 2021), suggesting cross-study replicative consistency.

PGS_{HEIGHT} GLM

The results of the GLM (involving PGS_{HEIGHT}) are presented in Table 7.

As with the previous models, the overall model fit (based on the Multiple *R*) is significant, with all independent variables accounting for 12% of the variance in height. PGS_{HEIGHT} was a significant and positive predictor of height (independent of the PCs). Although there are no significant main effects of SES, discrimination had a relatively small magnitude negative effect on height. The model did not detect significant interactions between either PGS_{HEIGHT} and SES, or between PGS_{HEIGHT} and discrimination.

Heritability Analysis of Discrimination and GCA

The results of the heritability analyses of discrimination and GCA are presented in Table 8.

The analyses indicate that GCA exhibits a (statistically significant) additive heritability of 49%. This is in line with previously reported heritability values for this trait in cohorts aged around 10 years (Bouchard, 2013). There were also significant contributions stemming from shared and also nonshared environmentality (with the contribution of the latter being substantially greater, which is also consistent with the literature; Bouchard, 2004). By contrast, discrimination exhibits no significant additive heritability, but is associated with significant shared and nonshared environmentality. This is inconsistent with theories positing that

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

Table 8. Variance component analyses estimating the proportion of additive genetic variance (A^2), shared environmental variance (C^2), and unshared environmental variance (E^2) associated with discrimination and GCA along with 95% CIs

Group	Phenotype	N_{DZ} + sibs	N_{MZ}	A^2 (95% CI)	C^2 (95% CI)	E^2 (95% CI)
Full sample	Self-reported discrimination	420	252	.027 (−.048, .102)	.117 (.042, .190)*	.855 (.834, .874)*
Full sample	GCA	420	252	.488 (.429, .543)*	.250 (.178, .319)*	.263 (.192, .332)*

Note: *$p < .05$.

discrimination should proxy underlying phenotypes (such as skin reflectance or color, which, as was previously noted, appears to be very highly heritable), and indicates that participant experience of discrimination is purely a function of the action of both shared and (to a much greater degree) nonshared environmental factors.

Moderation Analysis

Table 9 lists the vectors of the subtest GCA loadings along with the discrimination × PGS_{EDU} interaction vector (these were rescaled positively by multiplying each one by −1 in order to make the results of the vector correlation analyses more intuitive), the magnitude of the weighted averaged White-Black-Hispanic (SIRE) performance differences for each subtest, which were rescaled as *r* statistics in order to make them equivalent to the other effect sizes, the magnitude of the PGS_{EDU}-by-subtest association, the Scarr–Rowe effect vector, and the vectors of additivity (A), shared environmentality (C), and nonshared environmentality (E) estimated for each subtest (also rescaled as *r*-statistics). The table also lists the correlations among these vectors, and the results of the unit-weighted co-moderation (multivector) analysis.

Consistent with extant meta-analytic work on the relevant SIRE group comparisons (te Nijenhuis & van den Hoek, 2016; te Nijenhuis et al., 2019), the weighted (accounting for the dissimilar sample sizes) vector of the mean differences among the three SIRE groups was positively and significantly correlated with the vector of GCA factor loadings ($r = .68$), indicating that Spearman's hypothesis holds in this cohort. There was a nonsignificant positive association between the vector of SIRE group differences and the subtest A vector ($r = .19$); by contrast, there was a strong, significant positive association between the vector of these differences and the PGS_{EDU}-by-subtest association vector ($r = .96$). The latter vector correlated significantly and positively with the vector of GCA loadings ($r = .66$). The PGS_{EDU}-by-subtest association vector is apparently only weakly (and nonsignificantly) proxying the A vector ($r = .19$). A significant negative association between the subtest E and the SIRE group differences vectors was also present ($r = -.63$). Unsurprisingly, both subtest C and E vectors were significantly negatively associated with the subtest A vector ($r = -.75$ and $-.79$, respectively). The E vector was also significantly negatively correlated with the PGS_{EDU}-by-subtest association vector ($r = -.67$). The discrimination × PGS_{EDU} and Scarr–Rowe interaction vectors were positively and negatively, but nonsignificantly, associated with the subtest GCA-loading vector ($r = .26$ and $-.50$, respectively).

The multivector unit-weighted factor loaded positively and significantly onto the GCA subtest loading vector ($\lambda = .69$), the discrimination × PGS_{EDU} interaction vector ($\lambda = .65$), the SIRE group difference vector ($\lambda = .76$), the PGS_{EDU}-by-subtest association vector ($\lambda = .72$), and the subtest C vector ($\lambda = .75$). In the case of the Scarr–Rowe effect vector, the loading was close to zero ($\lambda = .06$). Nonsignificant negative loadings were found in the case of the subtest A and E vectors ($\lambda = -.25$ and $-.26$, respectively). The multivector factor accounted for 34% of the variance across vectors.

Discussion

The presence of both the Scarr–Rowe and discrimination × PGS_{EDU} interactions is evidenced using two methods, a GLM with a large number of controls and a more straightforward moderation analysis involving CPEM. In both cases, the former effect positively, and the latter negatively, predicts participant GCA, consistent with expectations that these sources of social adversity might contribute to reductions in the portability of PGSs for GCA in comparisons involving differentially socially advantaged groups.

The effects are found with the use of a $PGS_{EDU,}$ constructed using the PolyFun-Pred method, which attempts to increase the predictive validity of PGSs between different ancestral groups via incorporation of variants that have a higher probability of being causal. The fact that (in particular) a discrimination × PGS_{EDU} interaction can be recovered despite these precautions reduces the likelihood that this result is purely a function of LD decay stemming from cross-ancestry comparisons involving different SIRE groups self-reporting different levels of discrimination, although the PGS still seems to have significantly lower portability as a predictor of GCA in the Black subsample compared to the White and Hispanic ones (Table 1). It seems likely that some of the difference in portability still stems from residual LD decay; however, it should also be kept in mind that this pattern of findings is very much in line with the expectation that reduced PGS portability between SIRE groups should be present owing to gene-by-environment interactions, even if hypothetically 'perfect' PGSs composed of only causal variants were to be used (Rabinowitz et al., 2019). This interpretation is strengthened when considered in relation to PGS_{HEIGHT}, which exhibits much greater between-SIRE group portability (Table 2), in addition to which interactions between SES, discrimination, and PGS_{HEIGHT} are unambiguously absent when used to predict participant height (as is a main effect of SES; however, a relatively small main effect of discrimination is present in the theoretically expected direction). This suggests that among more socially neutral traits, such as height (relative to GCA), the forms of social adversity considered here have no (apparent) effects on PGS_{HEIGHT} expressivity. By contrast, both low SES and high discrimination reduce PGS_{EDU} expressivity on GCA. No sex differences were present for either interaction, but the subgroup analyses may have been underpowered to detect these.

The finding of the Scarr–Rowe effect in the current work is inconsistent with the outcome of one relatively recent study (Figlio et al., 2017), which failed to detect the effect using a very large sample of young (born in the 1990s and 2000s) US (specifically Floridian) twins and siblings. One possible reason for this discrepancy is that ABCD samples more broadly with respect to the US population than did the study of Figlio et al. (2017), thus their findings may have been confounded by regional factors that

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

Table 9. Loading of GCA on subtests along with (sign reversed) discrimination×PGS_{EDU} interactions, (weighted) SIRE group difference in subtest score means (rescaled as *r* values), Scarr-Rowe effects, and A, C, and E variance components transformed as r values. The vector correlations among these are reported along with the results of the multivector co-moderation analysis

Cognitive ability	GCA λ (95% CI; 1)	Discrimination × PGS_{EDU} interaction (2)	SIRE Difference; r (3)	PGS_{EDU} (4)	A (*r*) (5)	C (*r*) (6)	E (*r*) (7)	Scarr-Rowe effect (8)
Picture vocabulary	.59* (.58, .60)	.02 (−.00, .04)	.29* (.27, .31)	.20* (.18, .22)	.45* (.39, .51)	.55* (.50, .60)	.70* (.66, .74)	.01 (−.01, .03)
Flanker	.54* (.53, .56)	.03* (.01, .05)	.12* (.10, .14)	.05* (.03, .07)	.37* (.30, .43)	.26* (.19, .33)	.89* (.87, .90)	.02* (.00, .04)
List sorting	.63* (.62, .64)	.03* (.01, .05)	.20* (.18, .22)	.14* (.12, .17)	.41* (.35, .47)	.44* (.38, .50)	.80* (.77, .83)	.02* (.00, .05)
Card sorting	.59* (.58, .60)	.00 (−.02, .02)	.15* (.13, .17)	.08* (.05, .10)	.60* (.55, .65)	.00 (−.08, .08)	.80* (.77, .83)	.00 (−.02, .02)
Pattern comparison	.47* (.45, .49)	.01 (−.02, .03)	.08* (.06, .10)	.04* (.02, .06)	.35* (.28, .41)	.38* (.31, .44)	.85* (.83, .87)	.04* (.02, .06)
Picture sequence memory	.56* (.55, 57)	.02* (.00, .04)	.16* (.14, .18)	.08* (.06, .11)	.69* (.65, .73)	.00 (−.08, .08)	.72* (.68, .75)	.01 (−.01, .03)
Oral reading recognition	.60* (.59, .61)	.00 (−.02, .02)	.20* (.18, .22)	.16* (.14, .19)	.83* (.81, .85)	.00 (−.08, .08)	.56* (.51,.61)	.01 (−.01, .04)
Matrix test	.57* (.56, .58)	.00 (−.02, .03)	.19* (.17, .21)	.15* (.13, .17)	.41* (.35, .47)	.44* (.38, .50)	.80* (.77, .83)	.01 (−.01, .03)
Little Man Test	.43* (.41, .45)	.00 (−.02, .03)	.09* (.07, .11)	.05* (.03, .07)	.48* (.42, .54)	.00 (−.08, .08)	.88* (.86, .90)	.01 (−.02, .03)
Ravlt sd memory	.64* (.63, .65)	.02* (.00, .05)	.18* (.16, .20)	.12* (.10, .14)	.42* (.36, .48)	.41* (.35, .74)	.81* (.78, .83)	.01 (−.01, .03)
Ravlt ld memory	.64* (.63, .65)	.01 (−.01, .04)	.18* (.16, .20)	.12* (.10, .14)	.39* (.32, .45)	.48* (.42, .54)	.78* (.75, .81)	.00 (−.03, .02)
1	1.00	.26 (−.40, .74)	.68* (.67, .69)	.66* (.10, 90)	.10 (−.53, .66)	.38 (−.28, .80)	−.43 (−.82, .23)	−.50 (−.85, .14)
2		1.00	.18 (−.47, .70)	.02 (−.59, .61)	−.36 (−.79, .31)	.41 (−.25, .81)	.21 (−.45, .72)	.32 (−.35, .77)
3			1.00	.96* (.85, 99)	.19 (−.46, .71)	.42 (−.24, .81)	−.63* (−.89, −.05)	−.43 (−.82, .23)
4				1.00	.19 (−.46, .71)	.44 (−.22, .82)	−.67* (−.91, −.12)	−.37 (−.79, .30)
5					1.00	−.75* (−.93, −.27)	−.79* (−.94, −.36)	−.37 (−.79, .30)
6						1.00	.19 (−.46, .71)	.22 (−.44, .72)
7							1.00	.31 (−.36, .77)
8								1.00
Multivector loading	.69* (.15, .91)	.65* (.08, .90)	.76* (.30, .93)	.72* (.21, .92)	−.25 (−.74, .41)	.75* (.27, .93)	−.26 (−.74, .40)	.06 (−.56, .64)

Note: A, Additivity; C, shared environmentality; E, nonshared environmentality.
*$p \leq .05$,

do not generalize to the larger population of the US. Alternatively, as twin-based methods rely on phenotype-only resemblance among individuals of differing zygosity, there may be confounding phenotype-dependent effects in such models that can only be resolved with reference to direct measures of the relevant genotypes and their interactions. Figlio et al. (2017) were unable to determine zygosity in their same-sex twin sample,[7] which further complicates interpretations of their results.

Interestingly, discrimination (unlike GCA) appears to show virtually no influence from additive genetic factors. This strongly militates against the idea that, for someone to report having experienced it, a heritable phenotypic locus of some sort is required in order to act as a basis for, for example, discriminative social sorting. The prime candidate for such a locus is skin reflectance or color (Cooper, 2005; Rowe, 2005). As there is evidence that this phenotype is very substantially heritable, where self-reported discrimination is strongly a function of skin color, some significant (association-based) heritability of self-reported discrimination would reasonably be expected. These findings indicate that discrimination acts primarily through shared and nonshared environmental factors. This means that cognitively impairing forms of discrimination are likely associated with the *purely* socially constructed aspects of SIRE (those that are wholly independent of heritable ancestry-related phenotypes such as skin reflectance or color). It has been found that 'race' as a concept may, to a substantial degree, be a byproduct of social coalitional categorization, the significance of which can be 'erased' once alternative social cues are presented that more accurately map onto relevant coalitional structure (Kurzban et al., 2001; for a meta-analysis of 'erasing race' effects, see Woodley of Menie et al., 2020); thus, there is likely much about 'race' and related phenomena that exists purely in the psychological (and also sociological) realm and that is wholly divorced from outward markers of biogeographic ancestry. Elucidating the processes that go into the construction and persistence of such 'social forces' goes well beyond the current study.[8]

It was found that a multivector composed of all vectors loads positively and significantly onto the discrimination × PGS_{EDU} interaction vector, subtest GCA-loading vector, SIRE mean-differences vector, the PGS_{EDU}-by-subtest association vector, and the subtest C vector. The exceptions were the subtest-level Scarr–Rowe effect and the subtest A and E vectors, on which the multivector loaded nonsignificantly in all cases (Table 9). This indicates that the discrimination × PGS_{EDU} interaction is a potential contributing factor to the differences posited in modern versions of Spearman's hypothesis. More broadly, these findings — in particular, the finding that the A vector is nonsignificantly negatively related, whereas the C vector is significantly positively related, to the multivector — runs contrary to the expectation of proponents of the hereditarian hypothesis, which predicts that vectors of strongly genetic factors and GCA loadings should cluster, and should be independent of vectors involving strongly environmental factors (e.g., Rushton, 1999; Warne, 2021). The finding that the PGS_{EDU}-by-subtest association vector is positively and significantly associated with the multivector might, by contrast, be taken to evidence the hereditarian hypothesis (on the basis that this counts as a 'genetic effect', as per Rushton, 1999); however, this interpretation is confounded by the aforesaid positive co-moderation effects associated with the discrimination × PGS_{EDU} interaction and C vectors, both of which indicate environmental contributions. A plausible hypothesis is that where polygenic influences on subtest scores are generally higher, there is simply *greater* opportunity for (in particular) shared environmental factors to contribute via gene-by-environment interactions (involving discrimination) to group differences in GCA. It is notable in this regard also that the multivector actually loads *negatively* onto the A vector (which is the stronger measure of heritability). The finding that the A vector is a nonsignificant correlate of the SIRE differences vector further conflicts with the results of other studies that have used this parameter in vector correlation analyses to support the hereditarian position (Rushton et al., 2007). On this basis, Rushton's (1999) argument may not hold true, and therefore, should be treated more cautiously.

It should also be noted that the Scarr–Rowe effect is apparently not contributing to SIRE group differences, as this vector is nonsignificantly (and negatively) correlated with the vectors of both GCA loadings and SIRE group differences. This finding reinforces the argument that the term 'Scarr-Rowe effect' (and related terms) should be used exclusively to describe influences on GCA stemming from the action of purely SES-related social factors that are not intrinsically coupled with SIRE and associated phenomena (such as the differences posited in modern versions of Spearman's hypothesis and racial/ethnic discrimination; Giangrande & Turkheimer, 2021). The possibility that the Scarr–Rowe effect is not associated with cognitive performance at the level of GCA is furthermore consistent with a prediction of Woodley of Menie et al. (2018), who noted that:

> [I]f it is found that [GCA] loading negatively moderates ability measures' sensitivity to the Scarr-Rowe effect, then the [GCA] loading of tests might be an important factor to control for in future meta-analyses. Moreover, it suggests that the Scarr-Rowe effect may help increase our understanding of the Flynn effect (which also occurs to the greatest extent on the least [GCA]-loaded abilities . . .), as reductions in the strength of the former effect may be a driver of the latter effect. (p. 500)

Critics of the current findings might object that the effect sizes of the interactions are relatively small (by the standards of findings in the psychological sciences); however, gene-by-environment interactions involving *specific* environmental measures are theoretically expected to be of relatively small magnitude when compared with main effects (McGue & Carey, 2017). Just as '[a] typical human behavioral trait is associated with very many genetic variants, each of which accounts for a very small percentage of the behavioral variability' (the so-called fourth law of behavior genetics; Chabris et al., 2015, p. 304), it may furthermore be the case that a typical human behavioral trait is *also* associated with very many gene-by-environment interactions, each of which accounts for a very small percentage of the behavioral variability (a possibility that McGue & Carey, 2017, p. 43 suggest). Moreover, the relatively small magnitudes of these effects may also in part be a function of the imperfect validity and reliability of the various constructs employed in these analyses, the effect of which would be expected to attenuate the associated effect sizes (see Schmidt & Hunter, 2015).

A final objection might be made on the grounds that the population assessed here is quite young and that the impact of the gene-by-environment interactions detected might be transient, eventually exhibiting the fadeout effect. This relates to the tendency for early-life environmental factors that raise (and presumably also those that lower) IQ to no longer exert these effects as individuals age. It has been hypothesized that this might reflect the action of the Wilson effect, the tendency for the additive heritability of IQ to increase over (much of) the life course (Bouchard, 2013). While there are indications that the Scarr–Rowe effect 'fades' with age (Gottschling et al., 2019), it does not appear to fade to zero.

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

Possibly the most compelling evidence of this has been found in genetically informed studies employing cohorts that range from mid to late age in the Wisconsin Longitudinal and Health and Retirement Studies, where in both cases the Scarr–Rowe effect was detected specifically using various measures of participant childhood SES (Woodley of Menie et al., 2018, Woodley of Menie et al., 2021). On this basis, it might be reasonable to expect that the discrimination × PGS_{EDU} interaction might not 'fade out' completely either, unless its apparently greater affinity for GCA makes it more sensitive to fading out than the Scarr–Rowe effect — continued longitudinal data collection on the ABCD cohort could help to resolve this uncertainty in the future. It might also be the case that the discrimination × PGS_{EDU} interaction was generally stronger historically, as reductions in discrimination have been posited as a potential cause of (specifically) Black-White attainment gap closure on certain cognitive ability measures in the US (Rindermann & Thompson, 2013, p. 828).

In light of the theorizing of McGue and Carey (2017) discussed above, it is possible that there are other as yet unknown gene-by-environment interactions that, together with the discrimination × PGS_{EDU} interaction, could jointly substantially account for the SIRE group differences posited by modern versions of Spearman's hypothesis, which would obviously provide a strong evidential basis for environmental as opposed to hereditarian accounts of Spearman's hypothesis. A great deal of future research is required to adequately investigate this possibility.[9]

Acknowledgments. All data were sourced from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), accessioned in the NIMH Data Archive (NDA). ABCD is supported by the National Institutes of Health and federal partners under the following award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123 and U24DA041147.

The full list of supporters is available at https://abcdstudy.org/federal-partners.html.

Participating sites and study investigators are listed at https://abcdstudy.org/consortium_members/.

ABCD consortium investigators did not participate in the analysis or writing of this report. The ABCD data used in this paper were sourced from Version 3.01. Raw data are available at https://nda.nih.gov/edit_collection.html?id=2573, and instructions on how to access these data can be found at https://nda.nih.gov/training/modules/study.html. Additional support for this work came from supplements to U24DA041123 and U24DA041147, the National Science Foundation (NSF 2028680) and Children and Screens: Institute of Digital Media and Child Development Inc.

Conflict of Interest. The authors declare no conflict of interest.

Financial Support. The authors received no financial support.

Notes

1 SIRE sometimes denotes only self-identified race and/or ethnicity. Studies employing data on race and/or ethnicity to which we refer sometimes use other-identified, therefore not just self-identified, race and ethnicity data. The more encompassing term 'sociallyidentified race and/or ethnicity' is therefore preferred here, as this captures data on race and ethnicity regardless of whether self- or other-identification (or both) was used.

2 See Duncan et al. (2019, p. 7) for similar arguments concerning the role of SES and racial discrimination in confounding genetically informed studies involving socially non-neutral traits.

3 The Scarr–Rowe effect appears to be absent in the populations of other developed (e.g., Europe and Australia) and certain developing (e.g., Nigeria) regions or nations (Bates et al., 2016; Tucker-Drob & Bates, 2015; Hur & Bates, 2019), possibly reflecting the influence of greater equalization with respect to either positive or negative environments. These might include more extensive (relative to the US) welfare access in the case of the former, and higher levels of extreme poverty in the case of the latter.

4 This would be expected on the basis that height is likely not as important for attaining better life outcomes overall, relative to IQ, which partly reflects an individual's success in acquiring cognitive capital. On this basis, it is reasonable to hypothesize that height is not likely to be as negatively impacted as IQ through the action of social adversity. For this to hold, IQ/GCA should simply be *more* predictive of certain important positive life outcomes (such as income) than height. This can be demonstrated empirically using the NLSY'79, where regressing participant GCA scores (extracted from the subtests of the Armed Services Vocational Aptitude Battery), height, and sex against family income yields the following results: GCA $\beta = .465$, $p < .0001$, height $\beta = .028$, $p = .064$, and sex $\beta = .055$, $p < .0001$ ($N = 6310$). Independent of GCA and sex, height is not predictive of family income, consistent with the expectation that it is a more neutral trait with respect to socially significant life outcomes. It should be noted that our findings are consistent with claims of Duncan et al. (2019, p. 7), who also argued that height is less likely to be influenced by or confounded with sources of social adversity than traits like cognitive ability.

5 Spearman's hypothesis is sometimes said to come in a strong and weak form: on the strong form, mean SIRE group differences in cognitive test scores are entirely due to GCA differences; on the weak form, mean SIRE group differences in cognitive test scores are only partly due to GCA differences (see Frisby & Beaujean, 2015). Throughout this paper, where we refer to Spearman's hypothesis, we are only referring to the weak form.

6 This hypothesis is often defined in different ways in this context. Sometimes it seems to mean the view that genetic differences between SIRE groups account for at least 50% of cognitive test score differences between them. Sometimes a stronger view is taken where an explanation positing any genetic contribution to such between-SIRE group differences (i.e., >0% of differences attributed to genetic causes) is a hereditarian hypothesis. For recent discussion of the hereditarian hypothesis, see Warne (2021).

7 They ran their model on the assumption that all mixed-sex twins were dizygotic (in extraordinarily rare cases, monozygotic twin pairs can be mixed sex), and that their same-sex twin sample was 50% monozygotic and 50% dizygotic.

8 An example of potentially relevant research is the social–psychological study of Salvatore and Shelton (2007), who found that having (Black and White) subjects encounter (simulated) racial prejudice in an experimental context (they were tasked with reviewing job files in which an evaluator made either nonprejudicial, ambiguously prejudicial, or blatantly prejudicial judgements about a candidate) impaired their cognitive performance on the Stroop test. However, Black and White performance impairments were associated with different forms of perceived prejudice, with ambiguous prejudice impairing the performance of the former to a greater degree and blatant prejudice impairing the performance of the latter to a greater degree.

9 The existence of effects on GCA PGS expressivity stemming from discrimination potentially casts doubt on the conclusions drawn from the results of global admixture analyses, specifically in instances where biogeographic-ancestry (BGA)-informative genetic markers are used to predict variation in phenotypes such as GCA, and where such influences are used to support hereditarian hypotheses on the origin of SIRE group differences in such phenotypes (for a review of such studies, see Warne, 2021). This is because BGA will correlate with exposure to adverse environmental factors such as discrimination, the epigenetic influences of which on salient gene expression will be 'folded' into BGA measures, potentially confounding them as predictors of GCA in a way that, absent knowledge of the precise patterns of epigenetic marking in each genome, would make the relevant genetic and environmental influences extremely difficult to disentangle.

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

References

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

Bates, T. C., Hansell, N. K., Martin, N. G., & Wright, M. J. (2016). When does socioeconomic status (SES) moderate the heritability of IQ? No evidence for $g \times$ SES interaction for IQ in a representative sample of 1176 Australian adolescent twin pairs. *Intelligence*, *56*, 10–15

Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., Conley, D., Fletcher, J. M., Freese, J., Herd, P., Moffitt, T. E., Poulton, R., Sicinski, K., Wertz, J., & Harris, K. M. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences, USA*, *115*, E7275–E7284.

Bouchard Jr., T. J. (2004). Genetic influence on human psychological traits. *Current Directions in Psychological Science*, *13*, 148–151.

Bouchard Jr., T. J. (2013). The Wilson Effect: the increase in heritability of IQ with age. *Twin Research & Human Genetics*, *16*, 923–993.

Boutwell, B., Nedelec, J. L., Winegard, B., Shakelford, T., Beaver, K. M., Vaughn, M., Barnes, J. C., & Wright, J. P. (2017). The prevalence of discrimination across racial groups in contemporary America: Results from a nationally representative sample of adults. *PLoS One*, *12*, e0183356.

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., & Laibson, D. I. (2015). The fourth law of behavior genetics. *Current Directions in Psychological Science*, *24*, 304–312.

Cooper, R. S. (2005). Race and IQ: Molecular genetics as *Deus ex Machina*. *American Psychologist*, *60,* 71–76.

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

Figlio, D. N., Freese, J., Karbownik, K., & Roth, J. (2017). Socioeconomic status and genetic influences on cognitive development. *Proceedings of the National Academy of Sciences, USA*, *114*, 13441–13446.

Flynn, J. R. (1999). Evidence against Rushton: the genetic loading of WISC-R subtests and the causes of between-group IQ differences. *Personality & Individual Differences*, *26*, 373–379.

Flynn, J. R. (2018). Reflections about intelligence over 40 years. *Intelligence*, *70*, 73–83.

Frisby, C. L., & Beaujean, A. A. (2015). Testing Spearman's hypotheses using a bi-factor model with WAIS-IV/WMS-IV standardization data. *Intelligence*, *51*, 79–97.

George, D., & Mallery, M. (2010). *SPSS for Windows step by step: A simple guide and reference, 17.0 update* (10th ed.). Pearson.

Giangrande, E. J., & Turkheimer, E. (2021). Race, ethnicity, and the Scarr-Rowe hypothesis: a cautionary example of fringe science entering the mainstream. *Perspectives on Psychological Science*. Advance online publication. doi: 1.1177/17456916211017498.

Gignac, G. E., & Szodorai, E. T. (2016). Effect size guidelines for individual differences researchers. *Personality & Individual Differences*, *102*, 74–78.

Gorsuch, R. L. (1983). *Factor analysis*. L. Erlbaum.

Gorsuch, R. L. (1991). *Unimult: For Univariate and Multivariate Data Analysis*. UniMult.

Gorsuch, R. L. (2005). Continuous parameter estimation model: expanding the standard statistical paradigm. *Journal of the Science Faculty of Chiang Mai University*, *32*, 11–21.

Gottschling, J., Hahn, E., Beam, H. E., Spinath, F. M., Carroll, S., & Turkheimer, E. (2019). Socioeconomic status amplifies genetic effects in middle childhood in a large German twin sample. *Intelligence*, *72*, 20–27.

Guo, G., & Stearns, E. (2002). The social influences on the realization of genetic potential for intellectual development. *Social Forces*, *80*, 881–910.

Guo, G., Lin, M. J., & Harris, K. M. (2019). Socioeconomic and genomic roots of verbal ability. *bioRxiv*, 544411.

Hur, Y.-M., & Bates, T. (2019). Genetic and environmental influences on cognitive abilities in extreme poverty. *Twin Research & Human Genetics*, *22*, 297–301.

Jensen, A. R. (1968). Social class, race, and genetics: implications for education. *American Educational Research Journal*, *5*, 1–42.

Jensen, A. R. (1980). *Bias in mental testing*. Free Press.

Jensen, A. R. (1998). *The g factor: The science of mental ability*. Praeger.

Jernigan, T. L., Brown, S. A., & Dowling, G. J. (2018). The adolescent brain and cognitive development study. *Journal of Research on Adolescence*, *28*, 154–156.

Kurzban, R., Tooby, J., & Cosmides, L. (2001). Can race be erased? Coalitional computation and social categorization. *Proceedings of the National Academy of Sciences USA*, *98*, 15387–15392.

Lee, J. J., Wedow, R., Okbay, A., Kong, O., Maghzian, M., Nguyen-Viet, T. A., Bowers, P., Sidorenko, J., Karlsson Linnér, R., Fontana, M. A., Kundu, T., Lee, C., Li, H., Li, R., Royer, R., Timshel, P. N., Walters, R. K., Willoughby, E. A., … Cesarini, D. (2018). Gene discovery and polygenic prediction from a 1.1-million-person GWAS of educational attainment. *Nature Genetics*, *50*, 1112–1121.

Lee, R. T., Perez, A. D., Boykin, C. M., & Mendoza-Denton, R. (2019). On the prevalence of racial discrimination in the United States. *PLoS One*, *14*, e02110698.

Lewontin, R. C. (1970). Race and intelligence. *Science & Public Affairs: Bulletin of theAtomic Scientists*, *24*, 2–8.

Lewontin, R. C. (1976). Further remarks on race and the genetics of intelligence. In N. J. Block & G. Dworkin (Eds.), *The IQ controversy* (pp. 107–112). Pantheon.

Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*, 2987–2993.

Lorenzo-Seva, U., & Ten Berge, J. M. (2006). Tucker's congruence coefficient as a meaningful index of factor similarity. *Methodology*, *2*, 57–64.

Luciana, M., Bjork, J. M., Nagel, B. J., Barch, D. M., Gonzalez, R., Nixon, S. J., & Banich, M. T. (2018). Adolescent neurocognitive development and impacts of substance use: overview of the adolescent brain cognition development (ABCD) baseline neurocognition battery. *Developmental Cognitive Neuroscience*, *32*, 67–79.

Martin, N. (2000). Gene-environment interaction and twin studies. In T. D. Spector, H. Sneider, & A. J. MacGregor (Eds.), *Advances in twin and sib-pair analysis* (pp. 143–150). Greenwich Medical Media.

McGue, M., & Carey, B. E. (2017). Gene-environment interaction in the behavioral sciences: Findings, challenges, and prospects. In P. H. Tolan, & B. L. Leventhal (Eds.), *Gene-environment transactions in developmental psychopathology* (pp. 35–57). Springer International Publishing.

Nelder, J. A. (1994). The statistics of linear models: back to basics. *Statistics & Computing*, *4*, 221–234.

Nisbett, R. E. (2009). *Intelligence and how to get it: Why schools and cultures count*. Norton.

Paik, S. H., Kim, H.-J., Son, H.-Y., Lee, S., Im, S.-W., Yeon, J. H., Jo, S. J., Eun, H. C., Seo, J. S., Kwon, O. S., & Kim, J.-Il. (2011). Gene mapping study for constitutive skin color in an isolated Mongolian population. *Experimental & Molecular Medicine*, *44*, 241–249.

Price, A. L., Patterson N. J., Plenge R. M., Weinblatt M. E., Shadick N. A., & Reich D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, *38*, 904–909.

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: a toolset for whole-genome association and population-based linkage analysis. *American Journal of Human Genetics*, *81*, 559–575.

Rabinowitz, J. A., Kuo, S. I. C., Felder, W., Musci, R. J., Bettencourt, A., Benke, K., Sisto, D. Y., Smail, E., Uhl, G., Maher, B. S., Kouzis, A., & Ialongo, N. S. (2019). Associations between an educational attainment polygenic score with educational attainment in an African American sample. *Genes, Brain & Behavior*, *18*, e12558.

Rindermann, H., & Thompson, J. (2013). Ability rise in NAEP and narrowing ethnic gaps? *Intelligence*, *41*, 821–831.

Rinker, T. W. & Kurkiewicz, D. (2017). pacman: Package Management for R. version 0.5.0. http://github.com/trinker/pacman

Rodriguez, R., Tobias, R., & Wolfinger, R. (1995). Comment on JA Nelder 'The statistics of linear models: Back to basics'. *Statistics & Computing*, *5*, 97–101.

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press

Rosseel, Y. (2012). lavaan: an R package for structural equation modelling. *Journal of Statistical Software*, *48*, 1–36.

Rowe, D. C. (2005). Under the skin: on the impartial treatment of genetic and environmental hypotheses of racial differences. *American Psychologist*, *60*, 163–173.

Rowe, D. C., Jacobson, K. C., & van den Oord, E. J. (1999). Genetic and environmental influences on vocabulary IQ: parental education level as a moderator. *Child Development*, *70,* 1151–1162.

Rushton, J. P. (1999). Secular gains not related to the *g* factor and inbreeding depression unlike Black-White differences: a reply to Flynn. *Personality & Individual Differences*, *26*, 381–389.

Rushton, J. P., Bons, T. A., Vernon, P. A., & Čvorović, J. (2007). Genetic and environmental contributions to population group differences on the Raven's Progressive Matrices estimated from twins reared together and apart. *Proceedings of the Royal Society B: Biological Sciences*, *274*, 1773–1777.

Rushton, J. P., & Jensen, A. R. (2010). The rise and fall of the Flynn Effect as a reason to expect a narrowing of the Black-White IQ gap. *Intelligence*, *38*, 213–219.

Salvatore, J., & Shelton, J. N. (2007). Cognitive costs of exposure to racial prejudice. *Psychological Science*, *18*, 810–815.

Scarr, S. (1981). *Race, social class, and individual differences in IQ*. L. Erlbaum.

Scarr-Salapatek, S. (1971). Race, social class, and IQ. *Science*, *174*, 1285–1295.

Schmidt, F. L., & Hunter, J. E. (2015). *Methods of meta-analysis: Correcting error and bias in research findings* (3rd ed.). Sage.

Singh, V., & Singh, K. (2018). Additive genetic variance. In J. Vonk & T. K. Shackelford (Eds.), *Encyclopedia of Animal Cognition and Behavior*. Springer.

Spearman, C. (1927). *The abilities of man.* Macmillan.

te Nijenhuis, J., & van den Hoek, M. (2016). Spearman's Hypothesis tested on black adults: a meta-analysis. *Journal of Intelligence*, *4*, 6.

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*, 101–122.

Thompson, W. K., Barch, D. M., Bjork, J. M., Gonzalez, R., Nagel, B.J., Nixon, S. J., & Luciana, M. (2019). The structure of cognition in 9 and 10 year-old children and associations with problem behaviors: findings from the ABCD study's baseline neurocognitive battery. *Developmental Cognitive Neuroscience*, *36*, 100606.

Tucker-Drob, E. M., & Bates, T. C. (2015). Large cross-national differences in gene × socioeconomic status interaction on intelligence. *Psychological Science*, *27*, 138–149.

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic status modifies heritability of IQ in young children. *Psychological Science*, *14*, 623–628.

Turkheimer, E., Harden, K. P., D'Onofrio, B., & Gottesman, I. I. (2009). The Scarr-Rowe interaction between measured socioeconomic status and the heritability of cognitive ability. In K. McCartney & R. A. Weinberg (Eds.), *Modern pioneers in psychological science: An APS-Psychology Press series. Experience and development: A festschrift in honor of Sandra Wood Scarr* (pp. 81–97). Psychology Press.

van Buuren, S., & Groothuis-Oudshoorn, K. (2011). mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, *45*, 1–67.

Warne, R. T. (2021). Between-group mean differences in intelligence in the United States are >0% genetically caused: five converging lines of evidence. *The American Journal of Psychology*, *134*, 480–501.

Weissbrod, O., Kanai, M., Shi, H., Gazal, S., Peyrot, W.J., Khera, A. V., Okada, Y., Biobank Japan Project, Martin, A. R., Finucane, H., & Price, A. L. (2021). Leveraging fine-mapping and non-European training data to improve cross-population polygenic risk scores. *medRxiv*. doi: 1.1101/2021.01.19.21249483.

Wicherts, J. M., & Johnson, W. (2009). Group differences in the heritability of items and test scores. *Proceedings of the Royal Society B: Biological Sciences*, *276*, 2675–2683.

Woodley of Menie, M. A., Heeney, M. D., Peñaherrera-Aguirre, M., Sarraf, M. A., Banner, R., & Rindermann, H. (2020). A meta-analysis of the "erasing race" effect in the US, and some theoretical considerations. *Frontiers in Psychology*, *11*, 1635.

Woodley of Menie, M. A., Pallesen, J., & Sarraf, M. A. (2018). Evidence for the Scarr-Rowe effect on genetic expressivity in a large US sample. *Twin Research & Human Genetics*, *21*, 495–501.

Woodley of Menie, M. A., Peñaherrera-Aguirre, M., Dunkel, C., & Sarraf, M. A. (2021). Evidence for the Scarr-Rowe effect on genetic expressivity in the Health and Retirement Study. *Twin Research & Human Genetics*, *24*, 110–115.

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

https://doi.org/10.1017/thg.2022.3 Published online by Cambridge University Press