Wolfram, T., Moore, S., Li, J. H., Anomaly, J., Davidson, I., & Christensen, M. (2026). Interpreting Polygenic Prediction of Cognitive Ability: Evidence for Direct, Reliable, and Portable Genetic Effects. *Intelligence & Cognitive Abilities*, *2*(1), 1–19.
https://doi.org/10.65550/001c.158459

Research Articles

Interpreting Polygenic Prediction of Cognitive Ability: Evidence for Direct, Reliable, and Portable Genetic Effects

Tobias Wolfram[1], Spencer Moore[1], Jeremiah H. Li[1], Jonathan Anomaly[1], Ivan Davidson[1], Michael Christensen[1]

[1] Herasight Research

Keywords: cognitive ability, intelligence, polygenic prediction, within-family genetic effect, polygenic score, gene-environment interaction

https://doi.org/10.65550/001c.158459

Intelligence & Cognitive Abilities

Vol. 2, Issue 1, 2026

The interpretation of polygenic scores (PGS) for general cognitive ability (GCA) remains contested, with concerns about indirect genetic effects, environmental confounding, cross-ancestry portability, and the gap between PGS prediction and twin heritability estimates. Relying on a newly constructed PGS using within-family designs in two independent sibling cohorts (UK Biobank, N=4,642 pairs; ABCD, N=736 pairs), we demonstrate that direct genetic effects account for the large majority of PGS prediction (within-family attenuation $\delta/\beta \approx 0.88$). Correcting for measurement error in brief cognitive assessments, the within-family association with latent general ability is approximately 0.45, substantially higher than observed-scale estimates. Cross-ancestry portability follows theoretical expectations (66% effect retention in African Americans). Within families, higher PGS predicts greater educational attainment, occupational status, and reduced cardiometabolic disease risk, with no evidence for gene-environment interactions or substantial adverse pleiotropy. These findings replicate using a benchmark predictor based on publicly available data, confirming they reflect properties of cognitive genetic architecture rather than idiosyncrasies of a particular score.

Editor's Note

All three reviewers initially recommended a revision that required full algorithm details for an important finding reported in this paper. The Action Editor (RH) encouraged resubmission with these details, noting that transparency would ultimately strengthen the paper and the science, outweighing short-term commercial drawbacks. The authors affirmatively addressed all other reviewer comments and suggestions but withheld algorithm details for proprietary reasons, citing common practice in scientific publishing (see Authors' Statement). Two reviewers accepted this view; one strongly recommended rejection if the details were not included. We reviewed policies from other publishers on this issue, which typically defer to editorial judgment. Top journals have published papers omitting proprietary details, especially algorithms, balancing transparency with intellectual property protection. Without the algorithm details, the reported polygenic score variance predictions cannot be independently verified, but other key findings in the report can. In our view, the paper retains scientific merit and we accepted it. This is not an endorsement but reflects confidence that such work will inspire further research. An Associate Editor at ICA (TB) consults for the authors' company but had no involvement in processing of or decisions about this manuscript and he was not informed of any reviews until after the final decision.—TC.

Author's Statement

This manuscript's scientific contribution centers on what rigorous validation reveals about the genetic architecture of cognitive ability, specifically, that direct genetic effects account for the large majority of PGS prediction, that measurement error substantially attenuates reported heritability estimates, and that cognitive PGS show predominantly beneficial associations with life and health outcomes. The validation methods (within-family designs, reliability-aware scaling, and convergent robustness estimators) are described and reproducible for any researcher with approved access to UK Biobank and ABCD. The specific score construction pipeline and scoring weights underlie an actively maintained commercial product and cannot be fully disclosed. We therefore describe score construction at a conceptual level and provide a benchmark analysis based on publicly available data (Supplement III), demonstrating that our conclusions generalize beyond the proprietary predictor.

Introduction

General cognitive ability (GCA) is a highly heritable trait (Polderman et al., 2015) consistently associated with diverse life outcomes (Strenze, 2007), ranging from educational and occupational attainment to physical (Calvin et al., 2017; Stern, 2012; Whalley, 2001) and mental health (Koenen et al., 2009; Zammit et al., 2004). Strengthening

its genetic measurement is therefore highly relevant for health-related research (Abdellaoui et al., 2025).

Genetically, GCA is influenced by thousands of variants of small effect (Hill et al., 2019), in addition to high-effect rare variants (Chen et al., 2023). The growth of GWAS sample sizes for cognition-related phenotypes from under 10000 individuals (Butcher et al., 2008) to over 1 million participants (Lee et al., 2018) has enabled polygenic scores (PGS) with increasing predictive power. Recent genome-wide association studies have identified hundreds of genome-wide significant loci (Davies et al., 2018; Savage et al., 2018; Van Den Berg et al., 2025).

However, the interpretation of these scores remains contested. Influential analyses suggest that a substantial fraction of PGS prediction reflects indirect genetic effects rather than direct causal pathways (Young, 2019); reported SNP-heritabilities appear far below twin-based estimates, and concerns persist about cross-ancestry portability (Martin et al., 2019), gene–environment interactions, and antagonistic pleiotropy.

Using a new, optimized polygenic score (Methods), we address these challenges by isolating direct genetic effects from confounding and correcting for measurement error in brief cognitive assessments. We show that our analyses replicate when using a publicly available benchmark predictor, demonstrating that our findings reflect properties of cognitive genetic architecture rather than idiosyncrasies of a particular score.

Our analysis focuses on three critical domains of validity. First, we address the attenuation of genetic effects by measurement error. We apply psychometric reliability corrections and latent variable modeling to estimate the association between the PGS and latent general ability (g), rather than noisy observed scores. We demonstrate through simulation and three convergent estimators that this correction recovers the true latent association.

Second, we examine the persistence of predictive power within families to isolate direct genetic effects from population stratification (Young, 2019). We extend this analysis beyond cognitive test scores to evaluate whether the PGS predicts real-world outcomes, including educational attainment, occupational status, and cardiometabolic health, within sibling pairs, thereby establishing its practical utility and assess the potential for antagonistic pleiotropy by conducting a broad screen of psychiatric and health outcomes to test for "off-target" effects. Finally, we systematically characterize the boundary conditions of polygenic prediction. We evaluate the portability of the score across diverse ancestry groups (Martin et al., 2019) in the ABCD cohort to quantify the expected decay in predictive power. Furthermore, we test for gene-environment interactions (GxE) with key socioeconomic factors, such as parental education and family income, to determine whether genetic effects on cognition are additive or dependent on environmental context.

By integrating these approaches, we demonstrate that when measurement error is properly modeled, polygenic scores achieve substantial predictive accuracy for latent general ability ($r \approx 0.5$) that is robust to family-level confounding ($\delta/\beta \approx 0.88$) and broadly applicable across additive environmental contexts. These results suggest that previous concerns about the limited utility of cognitive PGS due to environmental confounding, missing heritability or context-dependence may be overstated when appropriate validation methods are employed.

Within-family Validation

PGS prediction ability among unrelated individuals can be higher than among individuals within the same family. Within families, genetic variation arises solely from random Mendelian segregation and recombination events. In contrast, predictions among unrelated individuals can also capture gene-environment correlations induced by population stratification, indirect genetic effects (the influence of alleles in relatives on the focal individual via the family environment), and correlations with other genetic factors not directly measured by the score. Such correlations often result from assortative mating and other forms of non-random mating (Kong et al., 2018; Young et al., 2022). We therefore validated our GCA PGS in two independent hold-out samples, examining both its prediction ability among unrelated individuals and within-family (Table 1 and Figure 1).

In a sample of 4642 European ancestry sibling pairs from UK Biobank with complete fluid intelligence/verbal-numerical reasoning (VNR)[1] data, we tested associations with VNR after controlling for age, sex and their interaction, as well as 10 genetic principal components[2]. The PGS coefficient (with residualized phenotype and PGS standardized to have variance 1, corresponding to the correlation coefficient with the residualized phenotype[3]) without family fixed-effects was 0.406 (SE 0.009), corresponding to an R^2 of 16.4% (95% confidence interval: $[15.1\%, 17.9\%]$). After adding family-fixed effects, we estimated the standard-

1 UK Biobank labels this measure 'fluid intelligence'; however, the items primarily assess verbal and numerical reasoning, so we refer to it as verbal–numerical reasoning (VNR) throughout, following existing literature (Lyall et al., 2016).

2 Repeating the UKB analyses with 40 genetic PCs (instead of 10) yielded materially identical results (β=0.4047, SE=0.0095, δ=0.3472, SE=0.0181.

3 We residualize phenotypes on standard covariates to remove non-genetic and ancestry-related structure prior to standardization; coefficients therefore reflect standardized associations between the PGS and covariate-adjusted phenotypes. As a robustness check, we additionally residualized the PGS on the same covariates (yielding strict partial correlations, i.e., residualizing both phenotype and PGS on covariates) and obtained essentially unchanged estimates: UKB - δ = 0.3468 (0.0213), β = 0.4062 (0.00874); ABCD - δ = 0.3637 (0.0541), β = 0.3836 (0.0231).

Table 1. Within-family validation results.

Dataset	Phenotype	N (sib pairs)	δ (direct effect)	β (pop effect)	δ/β	Reliability	δ_g	β_g
UK Biobank	Fluid Intelligence	4,642	0.355	0.406	0.876	0.607	**0.456**	0.521
			(0.0218)	(0.009)	(0.050)		**(0.028)**	(0.011)
ABCD	GCA factor score	736	0.363	0.380	0.959	0.803	**0.408**	0.425
			(0.055)	(0.023)	(0.134)		**(0.062)**	(0.026)

This table gives standardized coefficients when predicting the phenotype using our general cognitive ability (GCA) PGS without controlling for family fixed-effects (β, population effect) and when adding family-fixed effects (δ, direct effect). Phenotypes were residualized on age, sex, their interaction, and 10 genetic principal components. The coefficients are given for phenotype and PGS standardized to have variance 1 so correspond to (semi-partial) correlation coefficients, with standard errors in brackets. Standard errors for the effect-reduction ratio δ/β were calculated using sibling pair clustered bootstrap. We also give estimates adjusted for the reliability, ρ, of the phenotype taken as a measure of latent — rather than noisily measured — GCA (often called 'g'). The adjustment is performed by first computing the measurement reliability ρ, which is given by Fawns-Ritchie & Deary (Fawns-Ritchie & Deary, 2020) for Fluid Intelligence in UKB (test-retest reliability $r_{\text{test-retest}}$), and based on our own computations for the measures of general cognitive ability we constructed for ABCD using McDonald's ω. The error-corrected direct and population effects are given by $\delta_g{=}\delta/\sqrt{\rho}$ and $\beta_g{=}\beta/\sqrt{\rho}$, respectively.

ized direct effect of the PGS to be 0.355 (SE 0.0218)[4], implying only slight attenuation within-family (δ/β=0.876, SE=0.050), similar to that observed in the largest meta-analysis of such effects (Okbay et al., 2022) , but in contrast to more pronounced effect reduction recently reported in another dataset (Lin et al., 2025).[5]

We also validated in the ABCD Study (Volkow et al., 2018), a broadly representative sample (ages 9–10) from across the United States. We constructed a general cognitive ability factor from ten assessments including NIH Toolbox measures (Weintraub et al., 2013) (Picture Vocabulary, Flanker, List Sorting, Card Sorting, Pattern Comparison, Picture Sequence Memory, Oral Reading Recognition), WISC-V Matrix Reasoning (Wechsler, 2014), Little Man (Luciana et al., 2018), and Rey Auditory Verbal Learning Test trials (Rey, 1958), using the same validation methodology as in UK Biobank. Each score was first residualized for age, sex, and their interaction, then inverse rank-normal (IRNT) transformed. This measure of GCA showed good psychometric properties (McDonald's ω=0.803) and substantial heritability (h^2=0.64, SE=0.089) based on ACE decomposition in twin and sibling pairs ($N_{MZ} = 242$ and $N_{DZ}/Sib = 736$ European ancestry pairs with complete data). Validating in this group of siblings and dizygotic twins, we obtained similar results to our analysis of VNR in UK Biobank (Table 1).

As a comparative metric of what an estimator based on published results might be able to achieve, we leveraged summary statistics from Savage et al. (2018), the largest published GCA GWAS to date, and used them to build an LDpred2 (Privé et al., 2020)-based polygenic score, which we validated similarly to our own predictor. The results are given in Supplement III and Table S13.

Considering Measurement Reliability

Genetic effects operate on latent general cognitive ability—often referred to as g—rather than directly on the test scores obtained from any specific psychometric instrument (Panizzon et al., 2014). Because psychometric tests are inherently noisy proxies, observed scores contain measurement error that attenuates estimates of both SNP-based heritability and polygenic score associations. From a practical standpoint, it is unrealistic to assume that a brief, 13-item cognitive assessment like the UK Biobank fluid intelligence/VNR test (which takes only a few minutes to complete) measures general cognitive ability as reliably as a comprehensive, multi-hour assessment such as the Wechsler Adult Intelligence Scale battery. Although both tests aim to target the same underlying trait, the extensive item coverage of the Wechsler scales ensures substantially lower measurement error.

Applying a reliability correction thus places shorter and more extensive psychometric assessments on a common latent-trait scale, enabling accurate validation of genetic prediction and meaningful comparisons across measurement instruments. Furthermore, for applications such as predicting the differences in cognitive ability of embryos, the appropriate scale is the latent ability scale, not the observed scale, since differences in genotype at conception affect latent ability, not merely the scores on a test with less than perfect reliability.

To quantify and correct for this attenuation, we employ standard approaches from classical test theory. We assume the measured phenotype (g_{obs}) reflects an underlying latent ability (g) with added measurement error. Reliability (ρ) is defined as the proportion of variance in the observed measure due to variation in the latent trait:

4 As controlling for principal components technically could be considered as conditioning on a collider in the context of a within-family model, we conducted a robustness check where we did not consider PCs, which did not affect the estimate ($\Delta\delta_g$=0.0003).

5 We conducted a further robustness check, using the within-family modeling framework of snipar, which boosts power for estimation of direct genetic effects via Mendelian imputation of parental genotypes (Young et al., 2022) (see Materials & Methods). We similarly found minimal and not significantly different attenuation within-family using this method (ratio =0.853, SE 0.039).

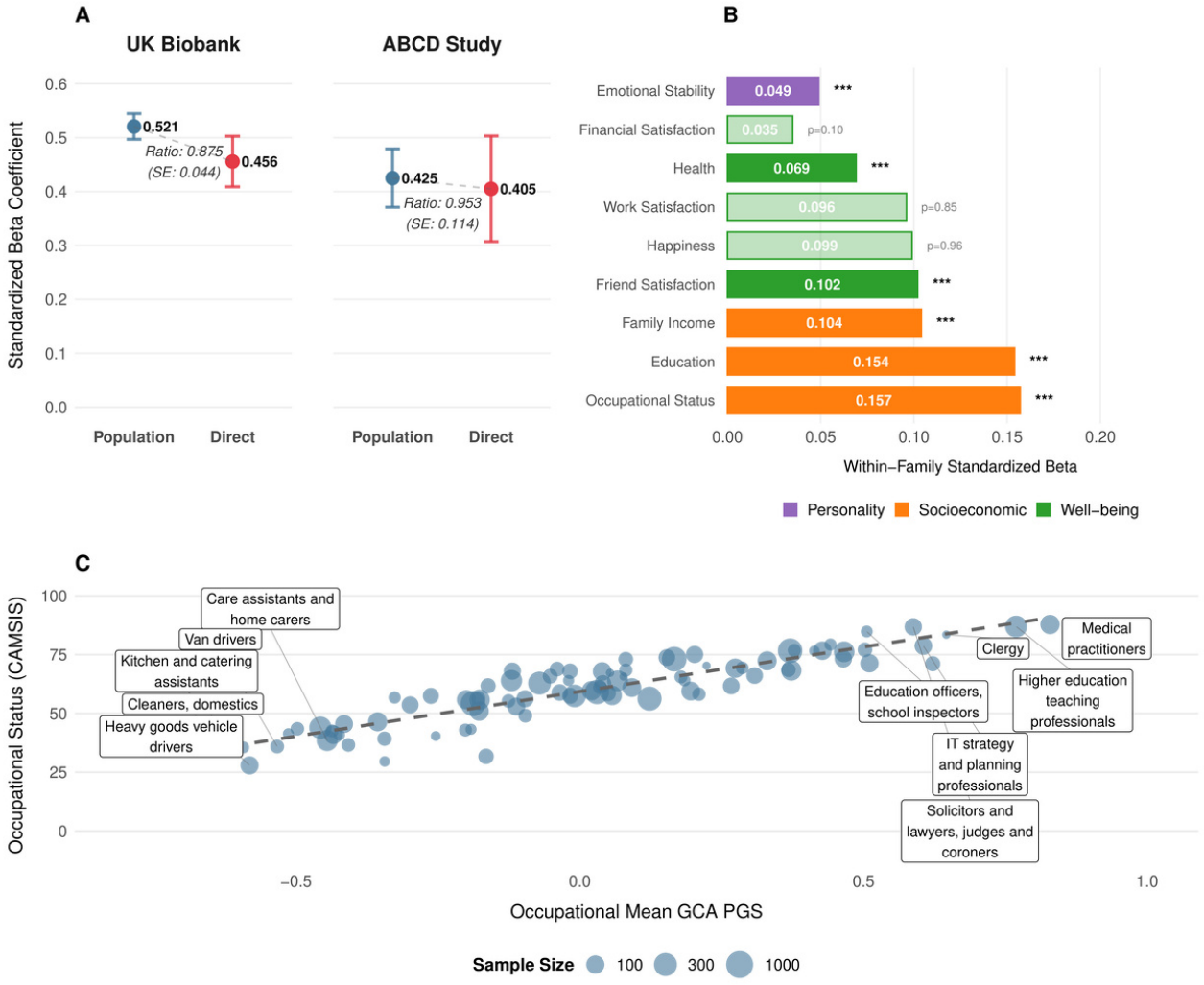


Figure 1. Cognitive Ability Polygenic Scores: Performance and Real-World Relevance.

(A) Polygenic score validation in UK Biobank and ABCD sibling cohorts. 'Population' refers to the score's association without family fixed effects and 'direct' refers to the score's within-family effect (with family-fixed effects). Estimates are given for PGS and phenotype standardized to have variance 1, after the phenotype was residualized on age, sex, their interaction, and 10 principal components. We display here effects on latent general cognitive ability ('*g*'), i.e. after accounting for the reliability of the psychometric measure in each cohort (Table 1). ***(B)*** Within-family polygenic score associations with various life outcomes. *** p < 0.0056 (Bonferroni corrected); * p < 0.05. ***(C)*** Relationship between mean occupational GCA polygenic score and CAMSIS occupational status in the UK Biobank. Point size is proportional to the number of individuals in each occupation.

$$\rho = \frac{Var(g)}{Var\left(g_{\text{obs}}\right)}$$

Under the standard assumption that measurement error is independent of genotype, genetic variance remains unchanged when transitioning from the latent trait to the observed measure, but total phenotypic variance increases, inflating the denominator of variance ratios.

If the SNP-heritability of the latent trait is $h^2_{\text{SNP},g}$, then the SNP-heritability estimated from the observed measure ($h^2_{\text{SNP},g_{\text{obs}}}$) is reduced proportionally to the reliability:

$$h^2_{\text{SNP},g_{\text{obs}}} = \rho \cdot h^2_{\text{SNP},g} \Rightarrow h^2_{\text{SNP},g} = \frac{h^2_{\text{SNP},g_{\text{obs}}}}{\rho}$$

We verified this relationship through simulations (Figure 2A). Specifically, we simulated a phenotype with a SNP-based heritability of 0.35 using GCTA, assuming 10000 causal loci drawn from the HapMap SNP set and the subset of individuals of British ancestry in the UK Biobank. To mimic varying degrees of measurement error, we progressively introduced random noise into this latent phenotype, reducing the effective reliability such that observed heritabilities decreased stepwise to approximately 0.30, 0.25, 0.20, 0.15, and 0.10. We then performed a GWAS using PLINK on a randomly selected training subset comprising 80% of individuals (N=273763) and estimated SNP-heritability using Linkage Disequilibrium Score Regression (LDSC). As expected, decreasing reliability proportionally lowered the observed SNP-heritability estimates; however, after applying our reliability correction, these estimates consistently and accurately recovered the true latent heritability across all simulated scenarios (Figure 2A).

The correlation between a polygenic score and an observed measure similarly underestimates the true correlation with the latent trait due to measurement error. The standard correction for correlations (Spearman's correction) gives:

$$r(\mathrm{PGS}, g) = \frac{r\,(\mathrm{PGS}, g_{\mathrm{obs}})}{\sqrt{\rho}}$$

where PGS is the standardized PGS and $r(\mathrm{PGS}, g)$ the correlation.

We confirmed the effectiveness of this correction through additional simulations (Figure 2B). Using the GWAS summary statistics generated at different reliability levels, we constructed polygenic scores by clumping and thresholding SNP associations (threshold p=1). We then validated these scores within the aforementioned 20% test set of 68514 UK Biobank participants of British ancestry. At perfect reliability, the true correlation given the unattenuated SNP-heritability of 0.35 was $r(\mathrm{PGS}, g)$=0.419, while lower reliability measures systematically underestimated the observed PGS-phenotype correlations. After applying the reliability correction, the estimates consistently recovered the true latent correlation across the range of reliabilities simulated (Figure 2B).

Thus, we adjust both the population-level (β) and within-family (δ) standardized associations to account for reliability, yielding corrected latent-trait estimates ($\beta_g = \beta/\sqrt{\rho}$ and $\delta_g = \delta/\sqrt{\rho}$) (Table 1 and Figure 2), using test-retest reliability ($\rho = r_{\text{test-retest}}$=0.607) for fluid intelligence/VNR in UKB from Fawns-Ritchie & Deary[6], and McDonald's Omega ($\rho = \omega$=0.803) that we calculated for our measure of general cognitive ability in ABCD[7]. Based on the UKB results, we estimate β_g=0.521 (SE=0.0113) without family-fixed effects, and δ_g=0.456 (SE=0.0276) with family fixed-effects. The latter reflects the standardized direct effect of the PGS on latent g. Our results based on ABCD indicated slightly lower correlation with latent g than in the UK Biobank (β_g=0.425, SE 0.0259), possibly due to lower heritability at ages 9–10 than in adulthood (Bouchard, 2013), as well as the fact that, in contrast to UKB, the ABCD PGS was constructed from imputed chip data, which is known to attenuate PGS-outcome associations (Li et al., 2021). The fixed-effects inverse-variance weighted meta-analysis estimate of the within-family correlation with latent g is δ_g=0.448, SE=0.025.

The validity of these adjustments is independently confirmed by a separate approach using Common-Path Slope estimation, an errors-in-variables correlated-vectors regression under a one-factor MIMIC model (see Supplement I): because both the PGS and the observed g_{obs} are imperfect measures of the same latent trait, g, their associations with external outcomes should differ only by a constant multiplicative factor if the PGS acts primarily via g. Concretely, across outcomes Y_j we expect $\beta_j^{(\mathrm{PGS})} \approx k\beta_j^{(g_{\mathrm{obs}})}$ with a line through the origin; the slope k captures the relative attenuation from measurement error in g_{obs} and limited predictive performance of the PGS. Fitting an errors-in-variables (Deming) regression to $\left(\beta_j^{(\mathrm{PGS})}, \beta_j^{(g_{\mathrm{obs}})}\right)$ and combining the slope with the population effect PGS-g association recovers the latent correlation $r\,(\mathrm{PGS}, g)$ and the implied reliability of the observed g_{obs}.

Within ABCD, across 20 explicitly non-cognitive outcomes selected for an exploratory pleiotropic scan (see below), the estimated Deming regression slope was k=0.539 (SE=0.028). Under a simple latent-factor model where PGS and g_{obs} measure the same latent trait, we can combine the association between PGS and observed GCA for ABCD given in Table 1, and use the Deming regression to estimate the latent correlation $r(PGS, g)$=0.453 (95% CI 0.417–0.488). This is in line with our attenuation-corrected population estimates in both cohorts. The same identity implies for our GCA measure in ABCD a reliability of 0.705 (95% CI 0.595–0.815), which matches our psychometric reliability estimate using McDonald's ω of 0.803 (for details, see Supplement 1).

To address psychometric skepticism about classical-test-theory deattenuation on potentially multidimensional batteries, we further re-estimated the association between the cognitive PGS and latent general ability using structural equation models with family-clustered (robust) standard errors and tested measurement invariance (see Supplement II). Because the latent models define a technically slightly different phenotype (multi-test latent g estimated by SEM and fit with FIML on individuals), we report them in a separate table. Table 2 shows the latent-scale within-family slope. Whereas a simple one-factor model fit the UKB data well, the ABCD data were better captured by a hierarchical model with three first-order domains loading on g (Supplementary Table S9). The within-family standardized latent coefficients are $\delta_{g_{SEM}} \approx 0.44$ in both cohorts, closely matching the reliability-corrected within-family estimates. In contrast to the classical test theory-derived results presented above, latent variable-based population estimates $\beta_{g_{SEM}}$ align well between both cohorts, closing the gap in performance between UK Biobank and ABCD and demonstrating that our headline results do not hinge on a single reliability parameter. We further verified metric measurement invariance across PGS strata and found no material item-level differential prediction (DIF) by within-family PGS (see Supplementary Tables S10–S12).

6 This published estimate aligns well with the Cronbach's alpha as provided elsewhere (Hagenaars et al., 2016) as 0.62. We also estimated test-retest reliability in the sample of UKB participants that took the test twice over the years (N = 29,879) as 0.647.

7 We chose test–retest reliability for the UK Biobank fluid intelligence/VNR test because it is a single, brief cognitive assessment administered at two time points. Test–retest reliability directly captures the proportion of variance that remains stable across repeated administrations, isolating genetically relevant variation from transient noise. In contrast, our ABCD cognitive measure is a composite factor derived from multiple cognitive tests. For such multi-item composites, McDonald's ω is the appropriate reliability index: it estimates the proportion of variance in the composite measure that is attributable to the common latent factor (here, general cognitive ability, or g), rather than idiosyncratic measurement error associated with individual tests or items.

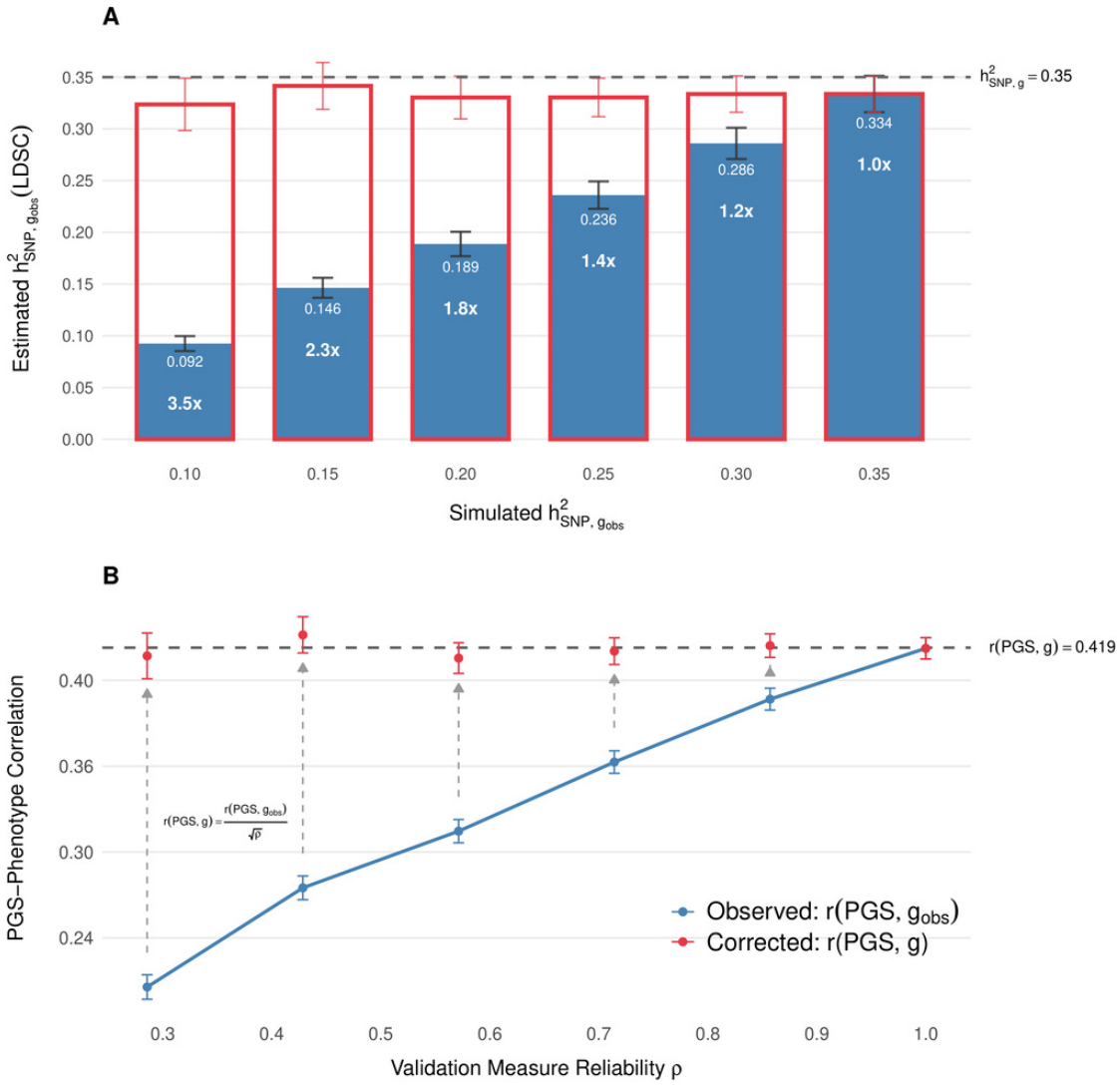


Figure 2. The effect of reliability on SNP heritability and PGS association.

In UK Biobank, we simulated a latent phenotype with true SNP heritability equaling 35% (an illustrative value consistent with observed $h^2 \approx 0.22$ after accounting for test reliability $\rho \approx 0.6$). We then examined the impact of the reliability of different measures of this latent phenotype on the estimated SNP heritability ***(A)*** and PGS-phenotype correlation ***(B)***. As expected, lower levels of reliability reduced apparent SNP heritability and PGS-phenotype correlation. However, reliability corrected estimates of both SNP heritability and PGS-phenotype correlation were never statistically significantly different from the true value for the latent phenotype. Bars in ***(A)*** show observed estimates with x-fold underestimation; red outlines denote reliability-corrected estimates

Table 2. Latent-variable within- and population associations of the cognitive PGS with general ability (*g*).

Cohort	Measurement model	N (individuals)	δ_{gSEM}	β_{gSEM}	CFI	TLI	RMSEA
UKB	One-factor *g*	36,126	0.439	0.525	0.969	0.927	0.022
			(0.018)	(0.008)			
ABCD	Second-order *g*	1,560	0.435	0.509	0.955	0.947	0.031
			(0.053)	(0.031)			

This table reports standardized latent-scale coefficients from structural equation models: in UK Biobank a one-factor g (8 indicators), and in ABCD a second-order g with first-order domains (crystallized, speed, memory). Coefficients are standardized on the latent scale (std.lv), estimated by MLR with FIML and family-clustered SEs in lavaan. Standard errors are shown in parentheses on the line below each estimate. $\delta_{g_{SEM}}$ is the within-family (direct) slope on PGS_w. $\beta_{g_{SEM}}$ is the latent population slope from a model regressing latent g on individual PGS without family fixed effects. Sample sizes reflect individuals contributing via FIML (partially missing indicators and incomplete pairs). Reported CFI/TLI/RMSEA summarize measurement fit. Unlike the reliability-adjusted coefficients in Table 1, these SEM estimates are already on the latent scale and nevertheless reproduce the same pattern of results.

Within-family PGS effects on life outcomes

Phenotypically, cognitive ability is associated with diverse life outcomes (Calvin et al., 2017; Strenze, 2007). Figure 1B demonstrates which associations persist within families, providing evidence for causal relationships between the polygenic score and several important life outcomes. Within UK Biobank sibling pairs, a one standard deviation increase in the polygenic score predicted 0.154 SD higher educational attainment ($p<0.001$). We also observed significant associations with occupational status measured by CAMSIS scores ($\beta=0.157$, $p<0.001$) and family income ($\beta=0.104$, $p<0.001$), aligning with known associations on cognitive ability and socioeconomic attainment (Marks, 2022).

The polygenic score was positively associated with better self-reported health ($\beta=0.069$, $p<0.0056$), higher satisfaction with friendships ($\beta=0.102$, $p<0.0056$) and lower neuroticism (higher emotional stability) ($\beta=0.049$, $p<0.0056$). However, we found no significant within-family associations with general happiness ($p=0.88$) or work satisfaction ($p=0.94$). These null findings for subjective wellbeing (albeit potentially a result of limited power), align with behavioral genetic studies showing that while GCA correlates with life outcomes, it has limited direct effects on subjective well-being (Bartels, 2015).

As an additional confirmation of validity, Figure 1C illustrates how mean polygenic scores vary across occupations in the UK Biobank. The systematic gradient aligns with cognitive demands of different professions (Schmidt & Hunter, 2004), with highest scores among medical practitioners, academics, and IT professionals, and lowest scores in manual occupations. (Wolfram, 2023)

Predicting disease risk

Motivated by the within-family association of the GCA PGS with better self-reported health, we estimated its effect on 19 common diseases using snipar in up to 17661 UKB sibling sets (Figure 3). At the population level, the PGS showed broadly protective associations with cardiometabolic diseases, which persisted within families. Direct genetic effects remained negative for all five cardiometabolic outcomes, though due to limited power, Bonferroni-significance was only achieved for hypertension. Direct genetic effect estimates for immune/inflammatory and cancer outcomes were generally small or imprecise. Overall, 14 of 19 outcomes showed the same direction for population and direct estimates, 12 of which indicate protective effects, lending further evidence to the hypothesis that part of the favorable health association of higher cognitive ability reflects direct effects rather than only family-level or population confounding.

We observed no significant association between the GCA polygenic score and Alzheimer's disease risk in our within-family analyses, though AD cases were rare in the sibling sample (N = 461) and most participants remain below typical dementia onset age.

Cross-Ancestry Performance Declines in Line with Expectations

Polygenic scores typically show substantially reduced performance in ancestry groups not represented in the training data due to differences in linkage disequilibrium patterns and allele frequencies among other factors (Mostafavi et al., 2020). Figure 4A shows our score's relative population-level performance in the ABCD Study for individuals of non-European ancestries. The score retained 89% of its standardized effect size in Hispanic/Latino Americans ($N=1985$, SE=6.5%), 88% in South Asian Americans ($N=42$, SE=40.8%), 66% in African Americans ($N=1491$, SE=7.2%), and 59% in East Asian Americans ($N=136$, SE=23.3%) relative to European Americans. Moore et al. (Moore et al., 2025) estimated the expected effect reduction for scores built using methods similar to the one employed here, which our results (despite noise due to small sample sizes) closely track (Figure 4A).

No Evidence for Gene-Environment Interactions

Gene-environment interactions for cognitive ability have often been proposed as an important factor in individual differences in cognitive ability. To assess whether there is any evidence for gene-environment interactions with our GCA PGS, we regressed the PGS, given environmental exposure, and examined their interaction on the cognitive ability measure we constructed in the European-American subset of ABCD ($N=5670$). Figure 4B shows no significant interactions between our polygenic score and environmental factors: parental education ($N=5032$, $p=0.87$), family income ($N=5395$, $p=0.25$), and family conflict ($N=5670$, $p=0.33$). This pattern is consistent with findings from well-powered twin and administrative studies in non-U.S. and population-scale U.S. samples (Bates et al., 2016; Figlio et al., 2017; Hanscombe et al., 2012), and with contemporary ABCD analyses, which likewise find SES and PGS associations to be largely additive (Paul et al., 2024).[8] While we cannot rule out the existence of any gene-environment interaction, the current results indicate that the magnitude of such interactions is likely small relative to the main effect of the PGS, indicating that predictions that do not model such interactions will remain well calibrated.

8 By contrast, the frequently cited evidence in favor of GxE by Tucker-Drob & Bates (2016) rests on nation-stratified subgroup estimates within the meta-analysis and its U.S.–non-U.S. heterogeneity does not comport with subsequent replication (Figlio et al., 2017). Given that early U.S. reports rested on modest samples, e.g., Turkheimer et al. (2003) inferred Scarr–Rowe moderation from 319 twin pairs at age 7 (114 MZ, 205 DZ), and that detecting small moderation effects typically requires thousands of twin pairs (Hanscombe et al., 2012), the accumulating record suggests Scarr–Rowe-style interactions are not a general feature of cognitive development.

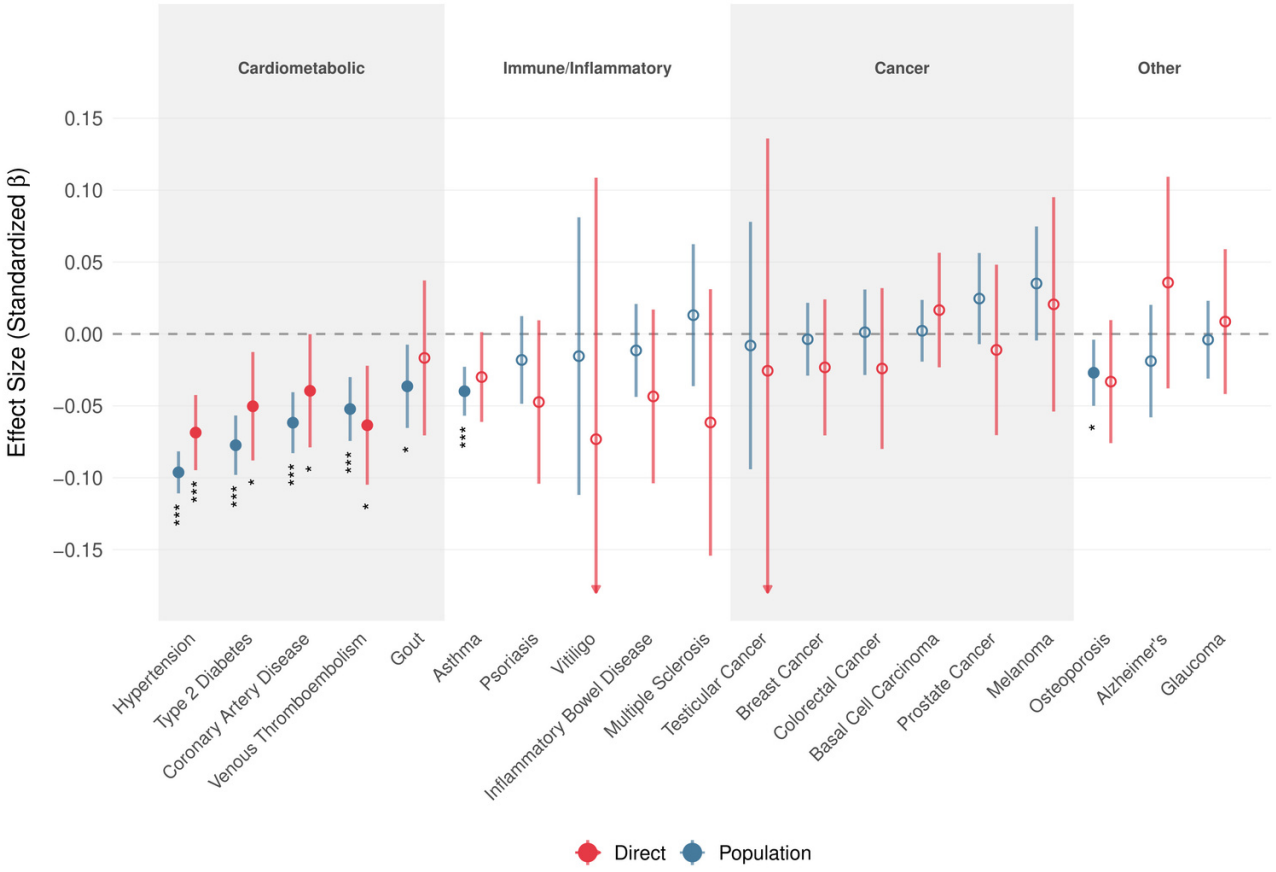


Figure 3. Association of GCA PGS with disease outcomes in UKB.

Estimates are given for the GCA PGS standardized to have variance 1, controlling for age, sex, and 10 principal components. 'Population' refers to the score's association without controlling for parental PGSs and 'direct' refers to the score's within-family effect estimated by controlling for observed or imputed parental PGSs. Bands indicate 95% confidence intervals, *** $p < .05/19$, * $p < .05$

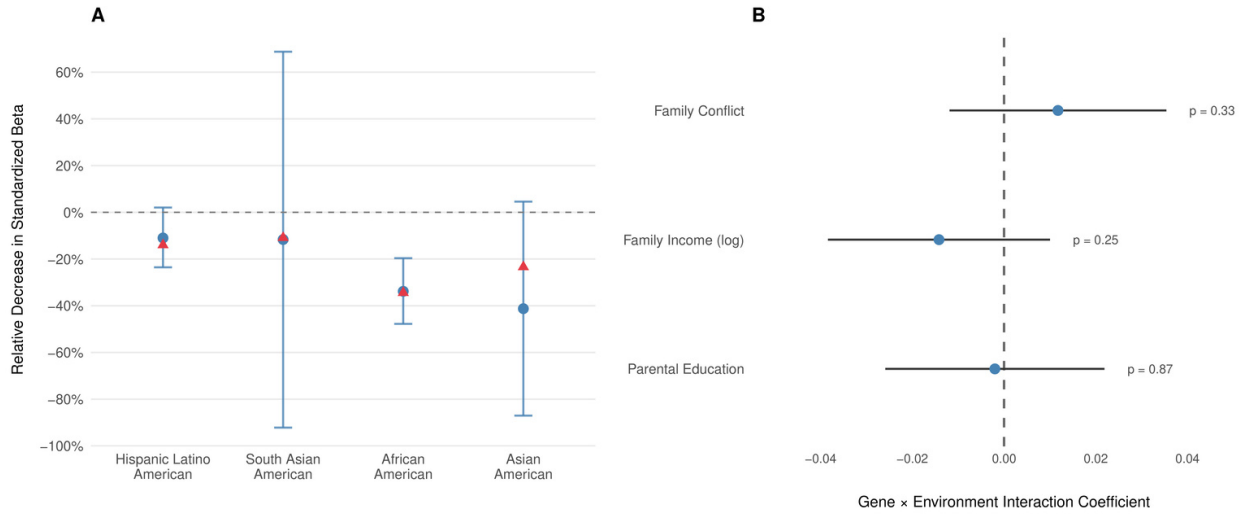


Figure 4. Robustness of Cognitive Ability Polygenic Score Effects.

(A) Relative performance across ancestry groups in the ABCD Study. The decrease is relative to the European American reference group. Error bars are 95% CIs. The observed decrease is shown in blue, and the expected decrease is shown as red triangles. ***(B)*** No significant gene-environment interactions with family conflict, family income (log), and average parental education (see materials and methods). All variables have been z-standardized. Bands indicate 95% confidence intervals.

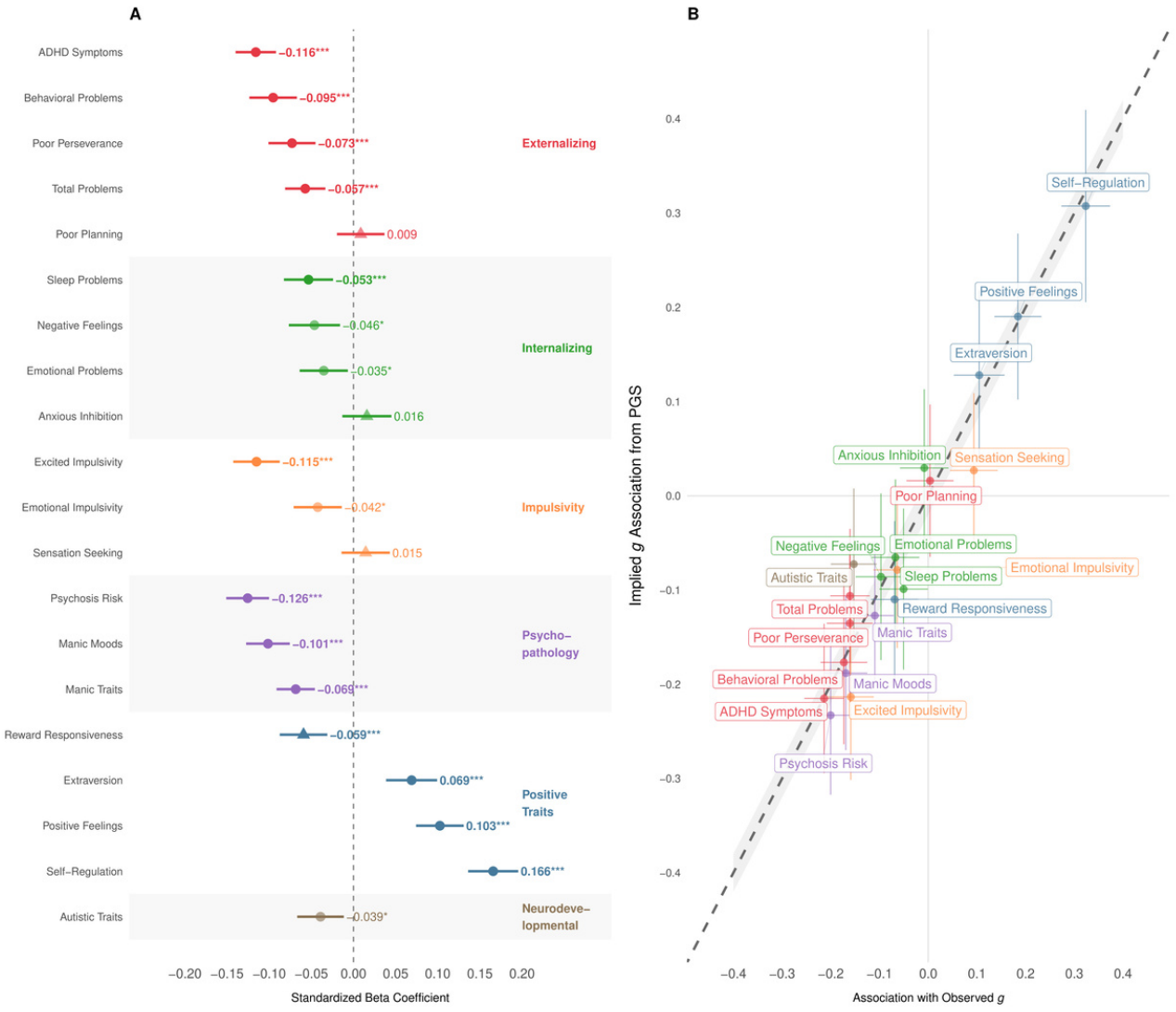


Figure 5. Pleiotropic Associations of GCA PGS.

(A) Standardized associations ($\beta, \pm$CI) between the GCA PGS and 20 psychological/neurodevelopmental outcomes in ABCD (European-ancestry youth, $N \approx 5600$–5940), controlling for age, sex, age × sex, and 10 genetic PCs; PGS and outcomes were z-standardized. Triangles indicate associations opposite to the expected beneficial direction. *** p < 0.0025 (Bonferroni); * p < 0.05 (nominal). ***(B)*** Cross-outcome scaling of betas for PGS vs. phenotypic g_{obs}. Each point is an outcome Y_j; vertical and horizontal bars show 95CIs (Bonferroni-adjusted). The dashed line indicates $y{=}x$, with points clustering around $y{=}x$ indicating (45°) that score-outcome associations are consistent with mediation by realized GCA. Bands indicate 95% confidence intervals, *** $p < .05/20$, * $p < .05$.

Assessing Potential Pleiotropic Off-Target Effects

Concerns about the interpretation of polygenic scores for cognitive ability often center on unintended associations with genetically correlated traits, particularly autism spectrum disorder (ASD) (Grove et al., 2019) and anorexia nervosa (Watson et al., 2019), a phenomenon also called "antagonistic pleiotropy". To directly address this, we screened for associations of our GCA PGS across 20 psychological traits in ABCD, including psychopathology, temperament, and autistic traits (see Materials & Methods).[9]

The results in Figure 5A show overwhelmingly beneficial associations between higher GCA PGS and outcomes indicative of improved psychological functioning. Specifically, a higher GCA PGS correlated with fewer psychotic-like symptoms (β=-0.126, $p< \times 10^{-21}$), lower ADHD symptoms (β=-0.116, $p< \times 10^{-20}$), reduced externalizing behaviors (β=-0.095, $p< \times 10^{-10}$), fewer sleep problems (β=-0.053, $p< \times 10^{-3}$), and increased positive affect (β=0.103, $p< \times 10^{-12}$). Notably, we observed no significant positive association between our cognitive PGS and autistic symptoms; indeed, the association was weakly negative

9 For statistical power reasons, analyses were conducted at the population level (between-family) rather than within-family, meaning these associations may be inflated by factors such as indirect genetic effects (Tan et al., 2024) and cross-trait assortative mating (Border et al., 2022). Therefore, effect sizes should be interpreted with caution and primarily as directional indicators.

(β=-0.039, nominally significant, Bonferroni non-significant).

The Common-Path Slope analysis (see Supplement I) shown in Figure 5B, establishes that PGS associations scale almost perfectly with the corresponding associations of observed g_{obs}, consistent with the PGS acting primarily via general ability rather than through idiosyncratic pathways specific to the PGS (pleiotropy). This supports the view that what we detect in the pleiotropy scan reflects the downstream correlates of g, not independent "off-target" mechanisms.

Overall, our comprehensive pleiotropy analysis provides little support for substantial negative off-target effects. Instead, higher cognitive ability genetic predispositions appear beneficial.

Conclusions

The interpretation of cognitive polygenic scores has been complicated by concerns about environmental confounding, limited cross-context generalizability, and the gap between PGS prediction and twin-based heritability estimates. Our results address each of these concerns through converging lines of evidence.

We achieved an R^2 of 16.4% for fluid intelligence/VNR in UKB, which represents a substantial advance over previous studies where performance has not exceeded 10% variance explained (Becker et al., 2021; Van Den Berg et al., 2025) and found this association to be robust within families. Reliability-corrected estimates indicate only slight attenuation within families ($\delta/\beta \approx 0.88$), a finding replicated using latent-variable structural equation modeling (Table 2) and confirmed using a reproducible benchmark predictor (Supplement III). These results are in line with recent findings showing that modeling cognitive ability at the latent level instead of relying solely on individual observed test scores improves polygenic prediction (Lin & Plomin, 2025). Although within-sibship GWAS can yield less biased estimates of direct genetic effects at individual loci (Howe et al., 2022), such analyses remain underpowered for cognitive phenotypes given current sibling sample sizes. Our results underscore that, across the phenotypes and predictors examined here, polygenic scores derived from conventional population-based GWAS capture substantial direct genetic effects, supporting their utility in contexts where only within-family variation is relevant (Ahangari et al., 2025; Craig et al., 2025; Moore et al., 2025).[10]

Beyond the magnitude of the effect, our analyses clarify the nature of the genetic influence on cognition. The significant within-family associations with educational attainment, occupational status, and family income provide evidence that these genetic effects translate into tangible life outcomes. Furthermore, the absence of gene-environment interactions with key socioeconomic variables strengthens the argument that, within the range of environments typical of European ancestry families in the USA, polygenic effects on cognition operate in a largely additive manner.

The public health relevance of this framework is reinforced by our disease prediction analyses. The consistently protective direct effects on cardiometabolic outcomes align with Mendelian randomization evidence regarding coronary artery disease and hypertension (Wang et al., 2023; Yang et al., 2022), suggesting these effects plausibly operate through behavioral pathways—such as diet and treatment adherence—that are influenced by cognitive ability. While our exploratory pleiotropy analysis indicated predominantly beneficial associations with psychological outcomes, future studies with larger sibling cohorts are required to confirm these population-level indications. Conversely, the null within-family effect for Alzheimer's disease must be interpreted with caution given the imperfect sensitivity of registry-based diagnoses (Wilkinson et al., 2019) and the potential confounding role of cognitive reserve.

Beyond validating a specific predictor, our reliability-aware approach has implications for how the field interprets SNP-heritability estimates, as it suggests that a substantial component of the apparent gap between SNP-h² and twin-h² in cognitive measures may reflect measurement unreliability. For instance, the relatively low SNP-based heritability ($h^2 \sim 0.22$) observed for the fluid intelligence/VNR measure in the UK Biobank, given its low reliability (~ 0.6), likely underestimates the true common variant SNP heritability of g, which, applying the attenuation correction outlined above, may exceed 0.35. Drawing parallels from height, where a significant portion of heritability resides in rare variants (Wainschtein et al., 2022), it is plausible that whole-genome sequencing analyses combined with proper modeling of measurement reliability will substantially bridge the heritability gap observed between SNP- and twin-based estimates, where the latter might suffer substantially less from measurement error, given measurement on the same day and age, often by the same rater and that aspects of the error themselves might be heritable (Nivard, 2023). Indeed, recent whole-genome analyses indicate that common and rare variants (MAF $\geq$ 0.001) can explain plausibly 33–40% of variation in cognitive ability (Wainschtein et al., 2025).

A key prediction of such a reliability-aware approach is that better psychometric measurement should yield higher SNP-based heritabilities. This highlights the potential

10 The likelihood that our within-family validations in UKB and ABCD are influenced by sample overlap with discovery datasets is minimal. The entire UKB sibling cohort used for validation was explicitly excluded from all discovery analyses, with UKB-based polygenic scores derived solely from non-sibling participants. ABCD served exclusively as an independent validation cohort, contributing no data to any discovery GWAS. Furthermore, any residual overlap from external summary statistics is numerically insignificant and incapable of generating the robust within-family effects observed, which depend solely on Mendelian segregation. The replication of within-family slopes in two independent cohorts further reduces any plausible concern about sample overlap influencing our results.

value of professionally constructed assessments, such as military IQ tests or scholastic achievement exams, whose SNP-h^2 should more closely approximate the true heritability of latent *g*. Practically, these results underscore the importance of extensive pretesting of cognitive measures before their implementation in large biobank cohorts. When assessment time is limited, adaptive, computer-assisted testing could further enhance measurement precision.

Collectively, these results highlight the diverse research settings in which a powerful, rigorously validated GCA PGS can benefit future health-related research. By establishing a validation protocol that explicitly models measurement error and family-level confounding, we provide a template for the safe and effective use of polygenic prediction in causal inference and epidemiological applications.

Materials & Methods

Scope of Disclosure

The analyses in this manuscript are designed to evaluate polygenic prediction in population and within-family designs and to characterize the role of measurement reliability using three convergent estimators. However, the score construction pipeline (including the exact engineered-phenotype search procedure, imputation system implementation details, and the final integrated scoring weights) underlies an actively maintained commercial product. We therefore describe the score construction process at a conceptual level and provide additional sensitivity analyses using an alternative GWAS-based predictor in the Supplement to demonstrate that our main validation conclusions are not specific to the proprietary predictor.

Score Development

We developed our polygenic score through a novel and comprehensive phenotype engineering pipeline designed explicitly to enhance the genetic signal of cognitive ability within the UK Biobank, restricting the analyses to individuals within the European ancestry cluster (participants of Ashkenazi, Polish, Italian, and British descent, following Privé et al., 2022 (Privé et al., 2022)). Siblings in UKB (reserved for validation purposes) were excluded from any phenotype engineering and downstream analysis to avoid data leakage.

Rather than relying on the raw cognitive measures provided by UKB, we explored a high-dimensional space of alternative score definitions for each of the 11 core cognitive tasks. Concretely, for each task we generated various candidate metrics spanning a small number of families of transformations (e.g., alternative scaling/transformations of the raw score distribution; alternative ways of combining paradata where available; and alternative item-weighting strategies for multi-item tasks). We also considered nuisance-adjusted variants that differ in which pre-specified covariate sets are removed prior to genetic screening (e.g., basic demographics and assessment context variables). Each engineered phenotype was then subjected to a highly optimized, custom implementation of a GWAS + LDSC pipeline, which enabled a rapid estimation of approximate SNP-heritabilities, allowing us to select only the single phenotype with the highest genetic signal-to-noise ratio for each cognitive domain. Candidate selection was then followed by full discovery GWAS runs using the primary GWAS pipeline described below.

To address missingness in cognitive measures within UKB, we used an autoencoder-based imputation strategy to predict missing cognitive scores from a broad panel of non-genetic variables (demographic, socioeconomic, and occupational/assessment-related variables) and other available UKB measures. The imputation model was trained and evaluated using held-out individuals within the non-sibling discovery set to avoid information leakage into the sibling validation cohort. Imputed values were used to increase completeness for phenotype engineering and auxiliary GWAS inputs. The key within-family validation analyses are conducted on observed outcomes in the respective holdout cohorts.

This pipeline yielded a suite of optimized measures which formed the basis of internal discovery Genome-wide association studies (GWAS), which were subsequently conducted on these phenotypes using REGENIE (Mbatchou et al., 2021). Individuals failing standard genomic QC were excluded. All analyses controlled for age, sex, their interaction, and the first ten principal components of genetic ancestry.

The UK Biobank's fluid intelligence/VNR test—comprising 13 offline-administered verbal-numerical reasoning items—was selected as the focal phenotype for our polygenic score training due to its substantial SNP-based heritability (LDSC h^2=0.2295) and substantial non-imputed sample size (N=182276). Where multiple fluid intelligence/VNR assessments were available per participant, scores were averaged.

We further boosted statistical power by integrating our internal cognition GWAS results with publicly available summary statistics from cognition-related and genetically correlated traits, including neurological and academic phenotypes, as well as additional GWAS conducted in the UK Biobank on heritability-maximizing transformations of such traits. We did not use educational attainment in our analyses due to its reduced within-family predictive validity (Okbay et al., 2022). Given the substantial difference in the SNP set used in our internal GWAS and in external summary statistics, we systematically imputed variants absent from individual GWAS summary statistics to comprehensively capture genetic variation. Using a hierarchical, multi-stage multi-trait meta-analysis framework, we sub-

stantially increased the effective sample size available for our focal fluid intelligence/VNR measure, significantly enhancing statistical power.[11]

Finally, we derived the polygenic score from these integrated summary statistics using SBayesRC (Zheng et al., 2024), applying a hierarchical, annotation-informed prior across an expanded genome-wide variant set. Functional annotations used were curated to augment the standard BaselineLD (Gazal et al., 2017) annotation set, including among others, evolutionary conservation metrics, cell-type-specific chromatin accessibility, and neurodevelopmental gene expression profiles.

Reliability

Using GCTA, we simulated a latent phenotype with $h^2_{\mathrm{SNP},g}$=0.35 (an illustrative value, consistent with observed estimates after reliability correction) on the subset of individuals of British ancestry in the UK Biobank (N=342277) using HapMap SNPs, assuming 10000 causal loci; we then added mean-zero noise to reduce reliability levels to the level of observed $h^2_{\mathrm{SNP},g_{\mathrm{obs}}}$ as shown in Figure 2. We then performed a simple GWAS using PLINK on a randomly selected training subset comprising 80% of individuals (N_{train}=273763) and estimated SNP-heritability using LDSC. Validation used the 20% hold-out (N_{test}=68514), where simple C+T scores (p-threshold=1, constructed using PRSice-2) were evaluated on the phenotypes of varying reliability.

Validation

Statistical models & inference

Let y denote the residualized cognitive outcome and PGS the standardized polygenic score (both variance-scaled to 1 after residualization). In our main analysis, we estimate

$$y_i = \alpha + \beta \mathrm{PGS}_i + \varepsilon_i \quad \text{(Population model)}$$
$$y_{if} = \alpha_f + \delta \mathrm{PGS}_i + \varepsilon_i \quad \text{(Family fixed-effects model)}$$

where α_f are family dummies. With this scaling, coefficients correspond to semi-partial correlations. We report the attenuation ratio R=δ/β with clustered (by family) bootstrap SEs.

Using the snipar (Young et al., 2022) within-family modeling framework, we instead estimate

$$y_{if} = \beta_{\mathrm{SNIPAR}} \mathrm{PGS}_i + u_f + \varepsilon_i \quad \text{(Population model)}$$
$$y_{if} = \delta_{\mathrm{SNIPAR}} \mathrm{PGS}_i + \alpha \hat{g}_{\mathrm{par}(f)} + u_f + \varepsilon_i \quad \text{(Within-family model)}$$

where α is instead the average non-transmitted coefficient, $\hat{g}_{\mathrm{par}(f)}$ the (imperfectly) imputed parental PGS for family f, and u_f a family-level random effect. We likewise report the attenuation ratio $\delta_{\mathrm{SNIPAR}}/\beta_{\mathrm{SNIPAR}}$ with clustered (by family) bootstrap SEs.

UK Biobank

Sample

The UKB sibling cohort reserved for validation was held out from all discovery stages. Sibling pairs were identified from UKB relationship files. We restricted to the Privé et al. (2022) white-ancestry cluster (Ashkenazi, Polish, Italian, UK) and excluded genotyping outliers, sex-chromosome aneuploidies, and heterozygosity/missingness outliers. Phenotypically, 4642 families with two genotyped siblings with non-missing outcome were retained for the validation[12].

Measure

We validated against each participant's first measure of the UKB offline fluid intelligence/VNR test, residualized for age, sex, their interaction and the first 10 PCs, Test–retest reliabilities (Pearson's r) were taken from Fawns-Ritchie & Deary (2020) as specified in Table 1.

ABCD

Cognitive factor construction

At baseline (ages 9–10), we extracted raw scores from NIH Toolbox tasks (Picture Vocabulary, Flanker, List Sorting, Card Sorting, Pattern Comparison, Picture Sequence Memory, Oral Reading Recognition), WISC-V Matrix Reasoning (raw sum), Little Man Task (percent correct), and a composite of Rey Auditory Verbal Learning Test trials (mean across trials). Each score was first residualized for age, sex, and their interaction, then IRNT-transformed. We fit a one-factor model using the `psych` package – the first factor explained $\sim 30\%$ of variance and showed good reliability ($\omega \approx 0.80$, see Table 1). The resulting factor scores (residualized for age, sex, their interaction and the first 10 PCs) served as the ABCD validation outcome.

---

11 Here, "hierarchical" refers to a staged integration strategy in which trait inputs are grouped and combined according to their relationship to the focal phenotype (e.g., cognition-related and genetically correlated auxiliary traits), rather than being treated as a single undifferentiated pool.

12 UK Biobank participation is non-random; participation probabilities increase with education/cognition and related traits, and participation itself has genetic correlates. Conditioning on inclusion (especially when restricting to complete sibling pairs) creates a collider: within included families, for a given within-family PGS difference, the lower-PGS sibling must on average have a more favorable residual in cognition to be observed. In the differenced fixed-effects equation, this induces $corr(\Delta\mathrm{PGS}, \Delta\varepsilon)<0$ (an index-event/volunteer-selection mechanism), biasing the within-family slope toward downward. Under monotone selection on cognition/residuals this attenuation is the default expectation in UKB-type samples with a "healthy volunteer bias". Under these conditions, the reported within-family coefficient can be interpreted as conservative with respect to the direct genetic association.

Sample

We identified European participants by reported ancestry (variable `demo_race`) being white and excluded individuals for whom multiple ancestry groups were given. From this set, we identified monozygotic and dizygotic twins as well as full siblings using the `genetic paired_subjectid` and `rel_relationship` fields. Following ABCD guidance, we excluded a small number of participants with inadequate visual acuity during testing (Snellen$<$4).

Twin-Heritability

Using independent MZ pairs and DZ/sibling pairs, we computed correlations within zygosity, then Falconer ACE components. Standard errors for correlations used Fisher's z transformation; ACE SEs were obtained by the delta method.

Comparative Baseline

To demonstrate that our validation framework generalizes to standard, reproducible cognitive polygenic predictors, we additionally constructed a genome-wide benchmark score using the results from Savage et al. (2018). Since publicly released summary statistics include the UK Biobank, we reconstructed their results while excluding our sibling validation sample. To achieve this, we sampled 142077 white British respondents who participated in the offline fluid intelligence/VNR test, as well as 53576 for whom only online test data was available. These numbers mirror those given in their paper, but explicitly exclude any siblings. This approach was feasible due to the increased availability of cognitive ability data since the original study was conducted. In cases where multiple data points were available for a participant, we used the first one. Using this data, we ran separate GWAS for the offline and online data and meta-analyzed them with summary statistics from Savage et al. (2018) excluding the UK Biobank. A widely used Bayesian shrinkage approach (LDpred2-auto) was used to construct a polygenic score which was subsequently used to rerun all validation analyses presented in the main text and beyond. Results are reported in Supplement III.

Associations with Other Outcomes

Within UKB sibling pairs, we regressed standardized outcomes (see Table S2 for a full list) on the standardized PGS with family fixed effects. Outcomes were pre-residualized (age, sex, age $\times$ sex, PCs) and z-scored as before. Multiple testing was controlled via Bonferroni across K=9.

Occupational Stratification

We merged the sibling holdout with their PGS values, then joined occupation and CAMSIS from the `ukbjobs` package (Akimova et al., 2025). For each occupation we computed mean standardized PGS, mean CAMSIS and sample size per occupation. Occupations with n<40 validation individuals were excluded to reduce instability. We visualized mean PGS vs CAMSIS, displaying point size by n, and added an OLS trend line to summarize the occupational gradient.

Disease prediction

To examine the cognitive ability PGS's capacity to predict disease outcomes within-families, we regressed binary outcomes on standardized sibling PGS values including or excluding imputed parental PGS values following the above *snipar* within-family modeling framework. Outcomes were analyzed using a generalized linear mixed model, fitted in **R** with the `lme4` package's `glmer()` function using the binomial probit link and including age, sex (except for sex-specific diseases) and the first 10 principal components as covariates. To achieve convergence for all 19 models, we set $nAGQ$=0 and used the *bobyqa* optimizer. The set of 19 disease outcomes were selected so as to include non-overlapping conditions collectively affecting multiple biological systems. Cases were identified within up to 17661 sibling sets from the UKB using ICD-10 hospital codes, OPCS-4 operation codes, medication usage, self-reports and responses to disease-specific surveys (Table S3).

Ethnic Portability

Effect reduction. In ABCD, we estimated population-level standardized effects separately by genetic ancestry group, using the same QC, residualization and standardization as for the sibling sample. Groups were defined using parent-reported ancestry: EUR = "white", AFR = "black", AMR = "hispanic", EAS = "chinese", "korean", "japanese", "filipino", SAS = "asian indian". Children with more than one ancestry category (with the exception of AMR) were removed. For each group g_i we fit $Y=\beta_{\mathrm{POP},g}\mathrm{PGS}$ and estimate relative performance $\mathrm{RelEff}_g=\beta_{\mathrm{POP},g}/\beta_{\mathrm{POP,EUR}}$ (N_{EUR}=5673; N_{AFR}=1491; N_{AMR}=1985; N_{EAS}=136; N_{SAS}=42). For presentation, we summarized the relative decrease and its 95% confidence interval via parametric bootstrap using the estimated β and SE for each group in Figure 4A.

Expected effect reduction. Moore et al. provided estimates for East Asian (58.7% relative R^2) and South Asian (79.4% relative R^2) populations. For African Americans, we used linear interpolation assuming 80% African/20% European ancestry based on their Nigerian validation (28.3% relative performance at 100% African ancestry). For Hispanic/Latino Americans, we applied Ding et al.'s (Ding et al., 2023) genetic distance framework: Hispanic/Latino populations cluster at intermediate genetic distances ($\sim$ 0.30–0.35) between Europeans and Africans in their analysis, reflecting their admixed ancestry. Using an intermediate distance of 0.32, a linear model based on Moore's results predicts 74% relative R^2, corresponding to 86% retention of standardized effect size, which aligns well with our observed 89%.

Gene-Environment Interaction

We exploratorily tested moderation of the PGS effect in European-ancestry ABCD at baseline using three contexts,

all using the same QC, residualization and standardization as for the sibling sample, in a standard regression framework:

$$Y = \alpha + \beta_1 \mathrm{PGS} + \beta_2 E + \beta_3 (\mathrm{PGS} \times E) + \varepsilon.$$

Moderators were **Parental education** (N=5032): mean of primary and partner education, measured as number of years in education based on highest school grade or degree attained (`demo_prnt_ed_v2`, `demo_prtnr_ed_v2`). **Family income** (N=5395): natural log of combined family income (`demo_comb_income_v2`). **Family conflict** (N=5670): Sum of parent report on the Family Environment Scale conflict subscale (`fes_p_ss_fc_pr`).

Pleiotropy Analysis

To assess potential "off-target" pleiotropy associated with our cognitive ability PGS, we conducted a systematic pleiotropy screen across 20 psychological and neurodevelopmental outcomes in ABCD (ages 9–10, $N \approx 5600$–5940 European-ancestry participants; see Table S4). Outcomes covered child psychopathology, temperament, affective traits, sleep, and autistic traits (SSRS). Each outcome was residualized for age, sex, age × sex, and 10 genetic PCs, inverse-rank-normal transformed when needed, and standardized. Analyses were run at the population-level (between-family), maximizing power, and did not include sibling fixed-effects.

Associations were tested using standardized regressions, controlling the family-wise error rate with Bonferroni correction ($\alpha = 0.05/20$). To clarify the clinical interpretation, each outcome was assigned a valence: associations were labeled "positive pleiotropy" if higher cognitive PGS correlated positively with beneficial traits (e.g., effortful control) or negatively with undesirable traits (e.g., ADHD symptoms), and "negative pleiotropy" otherwise. Given these analyses are between-family, effect sizes likely reflect upper bounds due to inflation from assortative mating and genetic nurture, and should thus primarily be interpreted for directionality.

Additionally, we applied the Common-Path Slope (CPS) method (Supplement I) to verify whether observed associations were mediated predominantly through general cognitive ability rather than independent, outcome-specific genetic pathways.

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number #103244. This work uses data provided by patients and collected by the NHS as part of their care and support.

Furthermore, data used in the preparation of this article were obtained from the Adolescent Brain Cognitive Development (ABCD) Study (https://abcdstudy.org), held in the NIMH Data Archive (NDA) under application #21063. This is a multisite, longitudinal study designed to recruit more than 10000 children aged 9–10 and follow them over 10 years into early adulthood. The ABCD Study® is supported by the National Institutes of Health and additional federal partners under award numbers U01DA041048, U01DA050989, U01DA051016, U01DA041022, U01DA051018, U01DA051037, U01DA050987, U01DA041174, U01DA041106, U01DA041117, U01DA041028, U01DA041134, U01DA050988, U01DA051039, U01DA041156, U01DA041025, U01DA041120, U01DA051038, U01DA041148, U01DA041093, U01DA041089, U24DA041123, U24DA041147. A full list of supporters is available at https://abcdstudy.org/federal-partners.html. A listing of participating sites and a complete listing of the study investigators can be found at https://abcdstudy.org/consortium_members/. ABCD consortium investigators designed and implemented the study and/or provided data but did not necessarily participate in the analysis or writing of this report. This manuscript reflects the views of the authors and may not reflect the opinions or views of the NIH or ABCD consortium investigators.

Competing interests

All authors are employees of Herasight, Inc. and hold equity in the company.

Funding

This work was funded by Herasight, Inc.

Submitted: November 21, 2025 CDT. Accepted: February 26, 2026 CDT. Published: April 06, 2026 CDT.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CCBY-4.0). View this license's legal deed at http://creativecommons.org/licenses/by/4.0 and legal code at http://creativecommons.org/licenses/by/4.0/legalcode for more information.

References

Abdellaoui, A., Martin, H. C., Kolk, M., Rutherford, A., Muthukrishna, M., Tropf, F. C., Mills, M. C., Zietsch, B. P., Verweij, K. J. H., & Visscher, P. M. (2025). Socio-Economic Status Is a Social Construct with Heritable Components and Genetic Consequences. *Nature Human Behaviour*, *9*(5), 864–876. https://doi.org/10.1038/s41562-025-02150-4

Ahangari, M., Moore, S., Davidson, I., Anomaly, J., Maier, R., Li, J. H., Christensen, M., Stern, D., & Wolfram, T. (2025). Comparative Evaluation of Cross-Ancestry Polygenic Risk Scoring of Type 1 Diabetes in the All of Us Cohort. *medRxiv*. https://doi.org/10.1101/2025.10.02.25337098

Akimova, E. T., Wolfram, T., Ding, X., Tropf, F. C., & Mills, M. C. (2025). Polygenic prediction of occupational status GWAS elucidates genetic and environmental interplay in intergenerational transmission, careers and health in UK biobank. *Nature Human Behaviour*, *9*(2), 391–405. https://doi.org/10.1038/s41562-024-02076-3

Bartels, M. (2015). Genetics of Wellbeing and Its Components Satisfaction with Life, Happiness, and Quality of Life: A Review and Meta-analysis of Heritability Studies. *Behavior Genetics*, *45*(2), 137–156. https://doi.org/10.1007/s10519-015-9713-y

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 × SES Interaction for IQ in a Representative Sample of 1176 Australian Adolescent Twin Pairs. *Intelligence*, *56*, 10–15. https://doi.org/10.1016/j.intell.2016.02.003

Becker, J., Burik, C. A. P., Goldman, G., Wang, N., Jayashankar, H., Bennett, M., Belsky, D. W., Karlsson Linnér, R., Ahlskog, R., Kleinman, A., Hinds, D. A., 23andMe Research Group, Agee, M., Alipanahi, B., Auton, A., Bell, R. K., Bryc, K., Elson, S. L., Fontanillas, P., … Okbay, A. (2021). Resource Profile and User Guide of the Polygenic Index Repository. *Nature Human Behaviour*, *5*(12), 1744–1758. https://doi.org/10.1038/s41562-021-01119-3

Border, R., Athanasiadis, G., Buil, A., Schork, A. J., Cai, N., Young, A. I., Werge, T., Flint, J., Kendler, K. S., Sankararaman, S., Dahl, A. W., & Zaitlen, N. A. (2022). Cross-Trait Assortative Mating Is Widespread and Inflates Genetic Correlation Estimates. *Science*, *378*(6621), 754–761. https://doi.org/10.1126/science.abo2059

Bouchard, T. J. (2013). The Wilson Effect: The Increase in Heritability of IQ With Age. *Twin Research and Human Genetics*, *16*(5), 923–930. https://doi.org/10.1017/thg.2013.54

Butcher, L. M., Davis, O. S. P., Craig, I. W., & Plomin, R. (2008). Genome-wide Quantitative Trait Locus Association Scan of General Cognitive Ability Using Pooled DNA and 500K Single Nucleotide Polymorphism Microarrays. *Genes, Brain and Behavior*, *7*(4), 435–446. https://doi.org/10.1111/j.1601-183X.2007.00368.x

Calvin, C. M., Batty, G. D., Der, G., Brett, C. E., Taylor, A., Pattie, A., Čukić, I., & Deary, I. J. (2017). Childhood Intelligence in Relation to Major Causes of Death in 68 Year Follow-up: Prospective Population Study. *BMJ*, *j2708*, j2708. https://doi.org/10.1136/bmj.j2708

Chen, C.-Y., Tian, R., Ge, T., Lam, M., Sanchez-Andrade, G., Singh, T., Urpa, L., Liu, J. Z., Sanderson, M., Rowley, C., Ironfield, H., Fang, T., Biogen Biobank Team, The SUPER-Finland study, Kyttälä, A., Elliott, A., Kämpe, A., Sourander, A., Tuulio-Henriksson, A., … Runz, H. (2023). The Impact of Rare Protein Coding Genetic Variation on Adult Cognitive Function. *Nature Genetics*, *55*(6), 927–938. https://doi.org/10.1038/s41588-023-01398-8

Craig, A., Wartschinski, L., Eyre, M., Davidson, I., Christensen, M., & Wolfram, T. (2025). Stage-Structured, Distributional Prediction of IVF Outcomes with Conditional Updating. *medRxiv*. https://doi.org/10.1101/2025.09.27.25336680

Davies, G., Lam, M., Harris, S. E., Trampush, J. W., Luciano, M., Hill, W. D., Hagenaars, S. P., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Liewald, D. C. M., Okely, J. A., Ahola-Olli, A. V., Barnes, C. L. K., Bertram, L., Bis, J. C., Burdick, K. E., Christoforou, A., DeRosse, P., … Deary, I. J. (2018). Study of 300,486 Individuals Identifies 148 Independent Genetic Loci Influencing General Cognitive Function. *Nature Communications*, *9*(1), 2098. https://doi.org/10.1038/s41467-018-04362-x

Ding, Y., Hou, K., Xu, Z., Pimplaskar, A., Petter, E., Boulier, K., Privé, F., Vilhjálmsson, B. J., Olde Loohuis, L. M., & Pasaniuc, B. (2023). Polygenic Scoring Accuracy Varies across the Genetic Ancestry Continuum. *Nature*, *618*(7966), 774–781. https://doi.org/10.1038/s41380-019-0394-4

Fawns-Ritchie, C., & Deary, I. J. (2020). Reliability and Validity of the UK Biobank Cognitive Tests. *PLOS ONE*, *15*(4), e0231627. https://doi.org/10.1371/journal.pone.0231627

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*, *114*(51), 13441–13446. https://doi.org/10.1073/pnas.1708491114

Gazal, S., Finucane, H. K., Furlotte, N. A., Loh, P.-R., Palamara, P. F., Liu, X., Schoech, A., Bulik-Sullivan, B., Neale, B. M., Gusev, A., & Price, A. L. (2017). Linkage Disequilibrium–Dependent Architecture of Human Complex Traits Shows Action of Negative Selection. *Nature Genetics*, *49*(10), 1421–1427. https://doi.org/10.1038/ng.3954

Grove, J., Autism Spectrum Disorder Working Group of the Psychiatric Genomics Consortium, BUPGEN, Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium, 23andMe Research Team, Ripke, S., Als, T. D., Mattheisen, M., Walters, R. K., Won, H., Pallesen, J., Agerbo, E., Andreassen, O. A., Anney, R., Awashti, S., Belliveau, R., Bettella, F., Buxbaum, J. D., Bybjerg-Grauholm, J., … Børglum, A. D. (2019). Identification of Common Genetic Risk Variants for Autism Spectrum Disorder. *Nature Genetics*, *51*(3), 431–444. https://doi.org/10.1038/s41588-019-0344-8

Hagenaars, S. P., Harris, S. E., Davies, G., Hill, W. D., Liewald, D. C., Ritchie, S. J., Marioni, R. E., Fawns-Ritchie, C., Cullen, B., … Malik, R. (2016). Shared genetic aetiology between cognitive functions and physical and mental health in UK biobank (n= 112 151) and 24 GWAS consortia. *Molecular Psychiatry*, *21*(11), 1624–1632. https://doi.org/10.1038/mp.2015.225

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M. A., Davis, O. S. P., Dale, P. S., & Plomin, R. (2012). Socioeconomic Status (SES) and Children's Intelligence (IQ): In a UK-Representative Sample SES Moderates the Environmental, Not Genetic, Effect on IQ. *PLoS ONE*, *7*(2), e30320. https://doi.org/10.1371/journal.pone.0030320

Hill, W. D., Marioni, R. E., Maghzian, O., Ritchie, S. J., Hagenaars, S. P., McIntosh, A. M., Gale, C. R., Davies, G., & Deary, I. J. (2019). A Combined Analysis of Genetically Correlated Traits Identifies 187 Loci and a Role for Neurogenesis and Myelination in Intelligence. *Molecular Psychiatry*, *24*(2), 169–181. https://doi.org/10.1038/s41380-017-0001-5

Howe, L. J., Nivard, M. G., Morris, T. T., Hansen, A. F., Rasheed, H., Cho, Y., Chittoor, G., Ahlskog, R., Lind, P. A., … Palviainen, T. (2022). Within-sibship genome-wide association analyses decrease bias in estimates of direct genetic effects. *Nature Genetics*, *54*(5), 581–592. https://doi.org/10.1038/s41588-022-01062-7

Koenen, K. C., Moffitt, T. E., Roberts, A. L., Martin, L. T., Kubzansky, L., Harrington, H., Poulton, R., & Caspi, A. (2009). Childhood IQ and Adult Mental Disorders: A Test of the Cognitive Reserve Hypothesis. *American Journal of Psychiatry*, *166*(1), 50–57. https://doi.org/10.1176/appi.ajp.2008.08030343

Kong, A., Thorleifsson, G., Frigge, M. L., Vilhjalmsson, B. J., Young, A. I., Thorgeirsson, T. E., Benonisdottir, S., Oddsson, A., Halldorsson, B. V., Masson, G., Gudbjartsson, D. F., Helgason, A., Bjornsdottir, G., Thorsteinsdottir, U., & Stefansson, K. (2018). The Nature of Nurture: Effects of Parental Genotypes. *Science*, *359*(6374), 424–428. https://doi.org/10.1126/science.aan6877

Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghzian, O., Zacher, 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 Genome-Wide Association Study of Educational Attainment in 1.1 Million Individuals. *Nature Genetics*, *50*(8), 1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Li, J. H., Mazur, C. A., Berisa, T., & Pickrell, J. K. (2021). Low-Pass Sequencing Increases the Power of GWAS and Decreases Measurement Error of Polygenic Risk Scores Compared to Genotyping Arrays. *Genome Research*, *31*(4), 529–537. https://doi.org/10.1101/gr.266486.120

Lin, Y., & Plomin, R. (2025). Charting the cognitive development of children using adult "polygenic g scores." *bioRxiv*. https://doi.org/10.64898/2025.12.19.695378

Lin, Y., Procopio, F., Keser, E., Kawakami, K., Rimfeld, K., Malanchini, M., & Plomin, R. (2025). Polygenic Score Prediction Within and Between Sibling Pairs for Intelligence, Cognitive Abilities, and Educational Traits From Childhood to Early Adulthood. *Intelligence & Cognitive Abilities*, *1*(1). https://doi.org/10.65550/001c.140654

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 Cognitive Development (ABCD) Baseline Neurocognition Battery. *Developmental Cognitive Neuroscience*, *32*, 67–79. https://doi.org/10.1016/j.dcn.2018.02.006

Lyall, D. M., Cullen, B., Allerhand, M., Smith, D. J., Mackay, D., Evans, J., Anderson, J., Fawns-Ritchie, C., McIntosh, A. M., … Deary, I. J. (2016). Cognitive test scores in UK biobank: Data reduction in 480,416 participants and longitudinal stability in 20,346 participants. *PloS One*, *11*(4), e0154222. https://doi.org/10.1371/journal.pone.0154222

Marks, G. N. (2022). Cognitive Ability Has Powerful, Widespread and Robust Effects on Social Stratification: Evidence from the 1979 and 1997 US National Longitudinal Surveys of Youth. *Intelligence*, *94*, 101686. https://doi.org/10.1016/j.intell.2022.101686

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., & Daly, M. J. (2019). Clinical Use of Current Polygenic Risk Scores May Exacerbate Health Disparities. *Nature Genetics*, *51*(4), 584–591. https://doi.org/10.1038/s41588-019-0379-x

Mbatchou, J., Barnard, L., Backman, J., Marcketta, A., Kosmicki, J. A., Ziyatdinov, A., Benner, C., O'Dushlaine, C., Barber, M., Boutkov, B., Habegger, L., Ferreira, M., Baras, A., Reid, J., Abecasis, G., Maxwell, E., & Marchini, J. (2021). Computationally Efficient Whole-Genome Regression for Quantitative and Binary Traits. *Nature Genetics*, *53*(7), 1097–1103. https://doi.org/10.1038/s41588-021-00870-7

Moore, S., Davidson, I., Anomaly, J., Li, J. H., Ahangari, M., Moissiy, L., Christensen, M., Young, A. S., Stern, D., & Wolfram, T. (2025). Development and Validation of Polygenic Scores for Within-Family Prediction of Disease Risks. *medRxiv*. https://doi.org/10.1101/2025.08.06.25333145

Mostafavi, H., Harpak, A., Agarwal, I., Conley, D., Pritchard, J. K., & Przeworski, M. (2020). Variable Prediction Accuracy of Polygenic Scores within an Ancestry Group. *eLife*, *9*, e48376. https://doi.org/10.7554/eLife.48376

Nivard, M. (2023). *Michel Nivard Tweets on Twin-Study Measurement Error (19 June 2023)* [Screenshots]. Zenodo. https://doi.org/10.5281/zenodo.17282230

Okbay, A., Wu, Y., Wang, N., Jayashankar, H., Bennett, M., Nehzati, S. M., Sidorenko, J., Kweon, H., Goldman, G., Gjorgjieva, T., Jiang, Y., Hicks, B., Tian, C., Hinds, D. A., Ahlskog, R., Magnusson, P. K. E., Oskarsson, S., Hayward, C., Campbell, A., … Young, A. I. (2022). Polygenic Prediction of Educational Attainment within and between Families from Genome-Wide Association Analyses in 3 Million Individuals. *Nature Genetics*, *54*(4), 437–449. https://doi.org/10.1038/s41588-022-01016-z

Panizzon, M. S., Vuoksimaa, E., Spoon, K. M., Jacobson, K. C., Lyons, M. J., Franz, C. E., Xian, H., Vasilopoulos, T., & Kremen, W. S. (2014). Genetic and Environmental Influences of General Cognitive Ability: Is g a Valid Latent Construct? *Intelligence*, *43*, 65–76. https://doi.org/10.1016/j.intell.2014.01.008

Paul, S. E., Elsayed, N. M., Colbert, S. M. C., Bogdan, R., Hatoum, A. S., & Barch, D. M. (2024). Family Income and Polygenic Scores Are Independently but Not Interactively Associated with Cognitive Performance among Youth Genetically Similar to European Reference Populations. *Development and Psychopathology*, 1–15. https://doi.org/10.1017/S0954579424001573

Polderman, T. J. C., 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*(7), 702–709. https://doi.org/10.1038/ng.3285

Privé, F., Arbel, J., & Vilhjálmsson, B. J. (2020). LDpred2: Better, faster, stronger. *Bioinformatics*, *36*(22–23), 5424–5431. https://doi.org/10.1101/2020.04.28.066720

Privé, F., Aschard, H., Carmi, S., Folkersen, L., Hoggart, C., O'Reilly, P. F., & Vilhjálmsson, B. J. (2022). Portability of 245 Polygenic Scores When Derived from the UK Biobank and Applied to 9 Ancestry Groups from the Same Cohort. *The American Journal of Human Genetics*, *109*(1), 12–23. https://doi.org/10.1016/j.ajhg.2021.11.008

Rey, A. (1958). *L'examen Clinique En Psychologie [The Clinical Examination in Psychology]*. Presses Universitaires De France.

Savage, J. E., Jansen, P. R., Stringer, S., Watanabe, K., Bryois, J., De Leeuw, C. A., Nagel, M., Awasthi, S., Barr, P. B., Coleman, J. R. I., Grasby, K. L., Hammerschlag, A. R., Kaminski, J. A., Karlsson, R., Krapohl, E., Lam, M., Nygaard, M., Reynolds, C. A., Trampush, J. W., … Posthuma, D. (2018). Genome-Wide Association Meta-Analysis in 269,867 Individuals Identifies New Genetic and Functional Links to Intelligence. *Nature Genetics*, *50*(7), 912–919. https://doi.org/10.1038/s41588-018-0152-6

Schmidt, F. L., & Hunter, J. (2004). General Mental Ability in the World of Work: Occupational Attainment and Job Performance. *Journal of Personality and Social Psychology*, *86*(1), 162–173. https://doi.org/10.1037/0022-3514.86.1.162

Stern, Y. (2012). Cognitive Reserve in Ageing and Alzheimer's Disease. *The Lancet Neurology*, *11*(11), 1006–1012. https://doi.org/10.1016/S1474-4422(12)70191-6

Strenze, T. (2007). Intelligence and Socioeconomic Success: A Meta-Analytic Review of Longitudinal Research. *Intelligence*, *35*(5), 401–426. https://doi.org/10.1016/j.intell.2006.09.004

Tan, T., Jayashankar, H., Guan, J., Nehzati, S. M., Mir, M., Bennett, M., Agerbo, E., Ahlskog, R., Anapaz, V. P. de A., Åsvold, B. O., Benonisdottir, S., Bhatta, L., Boomsma, D. I., Brumpton, B., Campbell, A., Chabris, C. F., Cheesman, R., Chen, Z., China Kadoorie Biobank Collaborative Group, … Young, A. S. (2024). Family-GWAS Reveals Effects of Environment and Mating on Genetic Associations. *medRxiv*. https://doi.org/10.1101/2024.10.01.24314703

Tucker-Drob, E. M., & Bates, T. C. (2016). Large Cross-National Differences in Gene × Socioeconomic Status Interaction on Intelligence. *Psychological Science*, *27*(2), 138–149. https://doi.org/10.1177/0956797615612727

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*(6), 623–628. https://doi.org/10.1046/j.0956-7976.2003.psci_1475.x

Van Den Berg, D. M., Huang, W., Malawsky, D. S., Danecek, P., Walter, K., Birney, E., Verweij, K. J. H., Smit, D. J., Lindsay, S. J., Hurles, M. E., Abdellaoui, A., & Martin, H. C. (2025). Imputation of Fluid Intelligence Scores Reduces Ascertainment Bias and Increases Power for Analyses of Common and Rare Variants. *medRxiv*. https://doi.org/10.1101/2025.06.18.25329418

Volkow, N. D., Koob, G. F., Croyle, R. T., Bianchi, D. W., Gordon, J. A., Koroshetz, W. J., Pérez-Stable, E. J., Riley, W. T., Bloch, M. H., Conway, K., Deeds, B. G., Dowling, G. J., Grant, S., Howlett, K. D., Matochik, J. A., Morgan, G. D., Murray, M. M., Noronha, A., Spong, C. Y., … Weiss, S. R. B. (2018). The Conception of the ABCD Study: From Substance Use to a Broad NIH Collaboration. *Developmental Cognitive Neuroscience*, *32*, 4–7. https://doi.org/10.1016/j.dcn.2017.10.002

Wainschtein, P., Jain, D., Zheng, Z., TOPMed Anthropometry Working Group, Aslibekyan, S., Becker, D., Bi, W., Brody, J., Carlson, J. C., Correa, A., Du, M. M., Fernandez-Rhodes, L., Ferrier, K. R., Graff, M., Guo, X., He, J., Heard-Costa, N. L., Highland, H. M., Hirschhorn, J. N., … Visscher, P. M. (2022). Assessing the Contribution of Rare Variants to Complex Trait Heritability from Whole-Genome Sequence Data. *Nature Genetics*, *54*(3), 263–273. https://doi.org/10.1038/s41588-021-00997-7

Wainschtein, P., Zhang, Y., Schwartzentruber, J., Kassam, I., Sidorenko, J., Fiziev, P. P., Wang, H., McRae, J., Border, R., … Zaitlen, N. (2025). Estimation and mapping of the missing heritability of human phenotypes. *Nature*, 1–9.

Wang, Y., Ye, C., Kong, L., Zheng, J., Xu, M., Xu, Y., Li, M., Zhao, Z., Lu, J., … Chen, Y. (2023). Independent associations of education, intelligence, and cognition with hypertension and the mediating effects of cardiometabolic risk factors: A mendelian randomization study. *Hypertension*, *80*(1), 192–203. https://doi.org/10.1161/HYPERTENSIONAHA.122.20286

Watson, H. J., Anorexia Nervosa Genetics Initiative, Eating Disorders Working Group of the Psychiatric Genomics Consortium, Yilmaz, Z., Thornton, L. M., Hübel, C., Coleman, J. R. I., Gaspar, H. A., Bryois, J., Hinney, A., Leppä, V. M., Mattheisen, M., Medland, S. E., Ripke, S., Yao, S., Giusti-Rodríguez, P., Hanscombe, K. B., Purves, K. L., Adan, R. A. H., … Bulik, C. M. (2019). Genome-Wide Association Study Identifies Eight Risk Loci and Implicates Metabo-Psychiatric Origins for Anorexia Nervosa. *Nature Genetics*, *51*(8), 1207–1214. https://doi.org/10.1038/s41588-019-0439-2

Wechsler, D. (2014). *Wechsler Intelligence Scale for Children* (5th ed.). Pearson.

Weintraub, S., Dikmen, S. S., Heaton, R. K., Tulsky, D. S., Zelazo, P. D., Bauer, P. J., Carlozzi, N. E., Slotkin, J., Blitz, D., Wallner-Allen, K., Fox, N. A., Beaumont, J. L., Mungas, D., Nowinski, C. J., Richler, J., Deocampo, J. A., Anderson, J. E., Manly, J. J., Borosh, B., … Gershon, R. C. (2013). Cognition Assessment Using the NIH Toolbox. *Neurology*, *80*. https://doi.org/10.1212/WNL.0b013e3182872ded

Whalley, L. J. (2001). Longitudinal Cohort Study of Childhood IQ and Survival up to Age 76. *BMJ*, *322*(7290), 819–819. https://doi.org/10.1136/bmj.322.7290.819

Wilkinson, T., Schnier, C., Bush, K., Rannikmäe, K., Henshall, D. E., Lerpiniere, C., Allen, N. E., Flaig, R., Russ, T. C., … Bathgate, D. (2019). Identifying dementia outcomes in UK biobank: A validation study of primary care, hospital admissions and mortality data. *European Journal of Epidemiology*, *34*(6), 557. https://doi.org/10.1007/s10654-019-00499-1

Wolfram, T. (2023). (Not Just) Intelligence Stratifies the Occupational Hierarchy: Ranking 360 Professions by IQ and Non-Cognitive Traits. *Intelligence*, *98*, 101755. https://doi.org/10.1016/j.intell.2023.101755

Yang, F., Hu, T., Chen, S., Wang, K., Qu, Z., & Cui, H. (2022). Low intelligence predicts higher risks of coronary artery disease and myocardial infarction: Evidence from mendelian randomization study. *Frontiers in Genetics*, *13*, 756901. https://doi.org/10.3389/fgene.2022.756901

Young, A. I. (2019). Solving the Missing Heritability Problem. *PLOS Genetics*, *15*(6), e1008222. https://doi.org/10.1371/journal.pgen.1008222

Young, A. I., Nehzati, S. M., Benonisdottir, S., Okbay, A., Jayashankar, H., Lee, C., Cesarini, D., Benjamin, D. J., Turley, P., & Kong, A. (2022). Mendelian Imputation of Parental Genotypes Improves Estimates of Direct Genetic Effects. *Nature Genetics*, *54*(6), 897–905. https://doi.org/10.1038/s41588-022-01085-0

Zammit, S., Allebeck, P., David, A. S., Dalman, C., Hemmingsson, T., Lundberg, I., & Lewis, G. (2004). A Longitudinal Study of Premorbid IQ Score and Risk of Developing Schizophrenia,Bipolar Disorder, Severe Depression, and Other Nonaffective Psychoses. *Archives of General Psychiatry*, *61*(4), 354. https://doi.org/10.1001/archpsyc.61.4.354

Zheng, Z., Liu, S., Sidorenko, J., Wang, Y., Lin, T., Yengo, L., Turley, P., Ani, A., Wang, R., Nolte, I., Snieder, H., Study, L. C., Yang, J., Wray, N., Goddard, M., Visscher, P., & Zeng, J. (2024). Leveraging Functional Genomic Annotations and Genome Coverage to Improve Polygenic Prediction of Complex Traits within and between Ancestries. *Nature Genetics*, *56*(5), 767–777. https://doi.org/10.1038/s41588-024-01704-y

Supplementary Materials

Supplemental Information I

Download: https://icajournal.scholasticahq.com/article/158459-interpreting-polygenic-prediction-of-cognitive-ability-evidence-for-direct-reliable-and-portable-genetic-effects/attachment/333424.docx

Supplemental Information II

Download: https://icajournal.scholasticahq.com/article/158459-interpreting-polygenic-prediction-of-cognitive-ability-evidence-for-direct-reliable-and-portable-genetic-effects/attachment/333425.docx

Supplemental Information III

Download: https://icajournal.scholasticahq.com/article/158459-interpreting-polygenic-prediction-of-cognitive-ability-evidence-for-direct-reliable-and-portable-genetic-effects/attachment/333426.docx

Supplemental Tables

Download: https://icajournal.scholasticahq.com/article/158459-interpreting-polygenic-prediction-of-cognitive-ability-evidence-for-direct-reliable-and-portable-genetic-effects/attachment/333427.pdf