Intelligence 78 (2020) 101408

Contents lists available at ScienceDirect

Intelligence

journal homepage: www.elsevier.com/locate/intell

# Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis

Bryan J. Pesta[a,*], Emil O.W. Kirkegaard[b], Jan te Nijenhuis[c], Jordan Lasker[d], John G.R. Fuerst[a,b]

[a] Cleveland State University, USA
[b] Ulster Institute for Social Research, UK
[c] Free University, Amsterdam, the Netherlands
[d] Independent Researcher, Atlanta, GA 30332, USA

ARTICLE INFO

*Keywords:*
Ethnicity
Heritability
Intelligence
Race

ABSTRACT

Via meta-analysis, we examined whether the heritability of intelligence varies across racial or ethnic groups. Specifically, we tested a hypothesis predicting an interaction whereby those racial and ethnic groups living in relatively disadvantaged environments display lower heritability and higher environmentality. The reasoning behind this prediction is that people (or groups of people) raised in poor environments may not be able to realize their full genetic potentials. Our sample ($k$ = 16) comprised 84,897 Whites, 37,160 Blacks, and 17,678 Hispanics residing in the United States. We found that White, Black, and Hispanic heritabilities were consistently moderate to high, and that these heritabilities did not differ across groups. At least in the United States, Race/Ethnicity × Heritability interactions likely do not exist.

## 1. Introduction

In behavioral genetic research, individual variance in cognitive ability is commonly partitioned into three components. The first is the additive genetic component ($a^2$, also known as $h^2$), which refers to genetic effects on a trait that act additively. This component is called (narrow) "heritability." The second component is the common or shared environment ($c^2$), which denotes environmental effects that make family members more similar. The third component is the unshared environment ($e^2$), which consists of non-genetic effects (plus measurement error) that are not shared between family members, but which instead differentiate them from each other. Collectively, the last two components are known as "environmentality" (Plomin, DeFries, Knopik, & Neiderhiser, 2014).

These three components together comprise the "ACE" model of behavioral genetics. The model represents one basic, biometric framework behavioral geneticists may use when studying the heritability of human traits, including intelligence. The ACE model assumes that environmental and genetic influences are additive, but allows that interactions (e.g., A × E) may also exist between components; these can be estimated as well (Plomin et al., 2014; Vinkhuyzen, van der Sluis, Maes, & Posthuma, 2012). Moreover, the model is useful in intelligence research because the behavioral genetic architecture of the trait is "surprisingly simple" (Plomin et al., 2014, p. 200). Finally, the ACE model nicely fits IQ data, and ACE estimates do not require the use of cumbersome kinship designs.

The relative importance of genetic and environmental sources of individual differences in cognitive ability has been extensively studied. Results for the general population show that the proportion of variance in IQ explained by genes increases with age (Plomin et al., 2014). Specifically, in early childhood, genetic effects explain less than 50% of IQ variance, and the effect of the shared environment is relatively strong. As children age, though, genetic effects become increasingly prominent, and the environmental variance due to factors common to siblings decreases. In adults, the heritability of intelligence is 60–80%, while the effect of common environment is small, if not zero (Plomin et al., 2014). The unshared environment explains the rest.

The degree to which one can generalize heritability estimates to other populations has been debated (see, e.g., Sesardic, 2005). It is clear, though, that some variables (e.g., age; Plomin et al., 2014) moderate the heritability of cognitive ability. One putative moderator is the quality of one's environment. Poorer (richer) environments supposedly correspond to lower (higher) heritability, to a presumably measurable degree. Said differently, "natural potentials for adaptive functioning are more fully expressed in the context of more nourishing environmental experiences" (Tucker-Drob & Bates, 2016, p. 1). This prediction is known as the *Scarr-Rowe hypothesis* (Scarr-Salapatek, 1971; Turkheimer, Harden, D'onofrio, & Gottesman, 2011).

* Corresponding author at: Department of Management, Cleveland State University, 2121, Euclid Avenue, Cleveland, OH 44115, United States.
*E-mail address:* b.pesta@csuohio.edu (B.J. Pesta).

https://doi.org/10.1016/j.intell.2019.101408
Received 15 May 2019; Received in revised form 6 November 2019; Accepted 12 November 2019
Available online 28 November 2019
0160-2896/ Published by Elsevier Inc.

The Scarr-Rowe hypothesis predicts lower heritabilities for lower performing social classes and racial/ethnic groups (Scarr-Salapatek, 1971, p. 1286). Scarr-Salapatek's (1971) original hypothesis and related ones – examples include the "Threshold Hypothesis" (Jensen, 1968), the "Bio-ecological Model" (Bronfenbrenner & Ceci, 1994), and the "Gene–Gini Hypothesis" (Selita & Kovas, 2019) – predict that Scarr-Rowe interactions will result when there are environmental differences. Assuming that social class and racial/ethnic differences are largely environmental in origin, Scarr-Salapatek (1971) and others have predicted lower heritabilities for the lower scoring groups.

Does the heritability of human intelligence differ by either social class or race/ethnicity? The answer is complicated because variables like age and the country sampled can moderate the effects. For example, a meta-analysis by Tucker-Drob and Bates (2016) found greater heritability with higher socioeconomic status, but these effects existed only with participants from the United States. Regarding age, recent data from Germany suggest the existence of a Scarr-Rowe interaction, but one which declines with increasing age (Gottschling et al., 2019).

While Scarr-Rowe interactions for social class are relatively well-studied, interactions for race or ethnicity are less so. Hence, whether Scarr-Rowe interactions for race or ethnicity exist is unclear. Some reviews suggest that the heritability of intelligence is similar across cultures (Plomin et al., 2014) and ethnic groups (Jensen, 1998; Rushton & Jensen, 2005). Others suggest differently (Turkheimer, Harden, & Nisbett, 2017).

The issue is relevant for several reasons, including evaluating the trans-ethnic validity of polygenic scores. Recently, Lee et al. (2018) developed polygenic scores for both intelligence and educational levels. These scores were derived from European samples and they showed lower predictive accuracy in non-European groups such as African Americans. The typical explanation offered for attenuated predictive accuracy is decay of linkage disequilibrium (LD) which results in differences in the correlations between SNPs across different ancestry groups (Zanetti & Weale, 2018). Another hypothesis appeals to lower within-group heritability in non-White groups (see, e.g., Rabinowitz et al., 2019). Both explanations are plausible since the predictive accuracy of polygenic scores is a joint function of (1) the validity of the scores as predictors of the traits, and (2) the within-group heritability of the traits in question (i.e., the association between the genotype and the phenotype; Daetwyler, Villanueva, & Woolliams, 2008). While LD decay might be a theoretically adequate explanation for attenuated predictive accuracy of PGS (Zanetti & Weale, 2018), whether it is the actual explanation can only be properly evaluated when the heritabilities of the trait within the different subgroups are known.

Our aim is to shed light on these matters by conducting a systematic review and meta-analysis. The goal is to test for the presence of Scarr-Rowe interactions with respect to race/ethnicity. Our specific research question is whether the heritability of intelligence differs across racial/ethnic groups residing in the United States (we searched for studies worldwide but found only samples from this country).

## 2. Method

### 2.1. Method for study identification, screening, and selection

We first created a database of all studies reporting ACE estimates for multiple racial/ethnic groups. Next, we conducted a meta-analysis of these estimates in order to assess the existence of Scarr-Rowe interactions for race/ethnicity. We included a qualitative review of these studies in the supplementary file. We also reported the PRISMA statement requirements that were relevant for our review and meta-analysis. Outlined below are the steps for the literature review, data extraction, data preparation, and data analysis.

#### 2.1.1. Information sources and eligibility criteria

There was no formal, registered review protocol for this study. To begin, we conducted searches to identify studies for inclusion in our database. First, we reviewed the literature discussed by Loehlin, Lindzey, and Spuhler (1975), Jensen (1998), Rushton and Jensen (2005), and *Fuerst & Dalliard, 2014. This review revealed 17 articles for potential inclusion in our database. Next, we scanned the following, major Gene × SES interaction review papers: Turkheimer et al. (2011), Nisbett et al. (2012), Hanscombe et al. (2012), Tucker-Drob, Briley, and Harden (2013), Turkheimer and Horn (2014), Tucker-Drob and Bates (2016), Selita and Kovas (2019), together with the papers cited therein. This search revealed approximately 30 (more or less non-redundant) potential papers in addition to the previous 17.

Third, we conducted Google Scholar, PsycINFO, and Medline searches on the subject matter. All searches were limited to literature in the English language. We considered both published and unpublished results. As for the timeline, we searched for abstracts published between 1970 and 2018.

#### 2.1.2. Electronic search

For Google Scholar, we searched for papers that included "heritability" and "race OR ethnicity OR African OR Black OR Latino OR Hispanic OR Asian Or Pacific Islander" and "cognitive ability OR achievement OR intelligence OR IQ." This produced many abstracts, ranked by relatedness to the topic, so we limited consideration to the first 20 pages (100 results) for each year. These totaled 4,800 abstracts. For PsycINFO, we used a Boolean, all-text search with the terms: "heritability OR genetic factors OR shared environment OR unshared environment" AND "race OR ethnicity OR minority OR Asian OR Pacific Islander OR African OR Hispanic OR Black OR Latino" AND "cognitive functioning OR cognitive ability OR IQ OR intelligence OR achievement OR math OR reading OR executive function OR verbal ability OR spatial ability." This process revealed 222 abstracts for potential inclusion in the study. For Medline, we used a Boolean, all-text search with the terms: "twins OR heritability OR shared environment OR unshared environment OR genetic factors" And "ethnicity OR minority OR Asian OR Pacific Islander OR African OR Hispanic OR Black OR Latino" AND "cognitive function OR IQ OR Intelligence OR achievement OR math OR reading OR executive function OR verbal ability OR spatial ability." We identified 207 abstracts via this search.

From the pool of approximately 5000 potential papers, we scanned all abstracts and texts (when warranted) for any that discussed: Variance decomposition (e.g., heritability, genetic factors, or twins), race/ethnic groups (along with specific ethnic groups), and cognitive ability (e.g., cognitive function, intelligence, or IQ). When a study implied that estimates meeting our criteria had been computed but not reported, we tried to contact the authors to request the subgroup estimates. Specifically, we tried to contact authors for reanalysis whenever: (1) a study reported variance estimates for a population while also noting the race/ethnicity of the sample, (2) sample sizes were thought to probably include an analyzable ($n_{pairs} > 50$ by kin class) minority sample, and (3) we did not already request the data for the same sample from other research teams. As re-analysis would involve non-trivial effort on the original authors' parts, we exercised constraint when making our requests. As a result, we emailed 25 corresponding authors/research teams for additional data.

#### 2.1.3. Eligibility criteria

One of us (JGRF) reviewed the abstracts on a rolling basis over the course of several years. The inclusion criteria were as follows:

1. We had to be able to copy or calculate ACE (or at least A) estimates from the reported data.
2. We had no minimum sample size for contrasting kinship classes (the minimum sample size mentioned above applied only to requesting data from other authors). All identified studies for which we had data were included in the meta-analysis.
3. The ACE or A estimates had to be interpretable in light of theory.

Due to this criterion, we excluded a study by Zhang and Pierce (2014). Pierce (December, 18, 2018, personal communication) reported a heritability of .999 for African Americans using GCTA. The authors interpreted this to mean that the sample size was too small to accurately estimate heritability using this technique.

4. The ACE or A estimates had to be based on some measure of mental ability, as defined by Jensen (1998, pp. 52–53; see also, Rindermann, 2018, pp. 43–45).
5. The samples had to include data for more than one race or ethnic group to allow for comparisons using both the same tests and methods for computing ACE estimates. For example, Johnson et al. (2007) reported the heritability of complex reasoning in a sample of Caribbean Latinos. We did not include this study here because there was no other racial or ethnic group with which to compare scores. In contrast, we did include ACE estimates for heterogenous racial/ethnic groups such as "non-majority" or "non-White." An example of the latter case is the study by Rhemtulla and Tucker-Drob (2012), which included estimates for Whites and "non-Whites".
6. When more than one study reported data from the same sample, we included only the study with the largest harmonic *N* and the most comprehensive analysis. For example, data from Beaver et al. (2013) overlapped with that from both Rowe, Jacobson, and Van den Oord (1999), and Guo and Stearns (2002). We included only the Beaver et al. (2013) analysis, because it had larger *N*s than Guo and Stearns (2002) and because it included longitudinal data (i.e., multiple waves), unlike Rowe et al. (1999).

#### 2.1.4. Identified studies and study selection

After contacting researchers for unpublished data, we identified 22 studies for possible inclusion in the meta-analysis. Next, two of us independently rated the set of potential studies for usability. Percent agreement was then computed, resulting in an interrater reliability of .91 (i.e., 20 out of 22). Thereafter, the ratings were discussed again until complete agreement was reached.

Fig. 1 depicts the flowchart for study identification. Note that one of these studies was an unpublished analysis (Fuerst, 2014), which we then reanalyzed using the updated NLSY Kinship Links data. Of the 22 studies in Fig. 1, we excluded seven from the qualitative review (provided in the Supplementary Materials). We did this because six of the seven contained redundant samples and one of the seven had uninterpretable results (according to the authors; Zhang & Pierce, 2014; Pierce, personal communication, December 18, 2018). For the meta-analysis, we excluded another two studies because we did a re-analysis of the first (*Fuerst & Dalliard, 2014), whereas the second (Loehlin et al., 1975) reported scores that were redundant with follow-up data we already had. The remaining 13 studies yielded 16 independent sets of samples, allowing for 40 dyadic group comparisons. Table 1 provides details for all these studies.

We did not limit our search to just the United States, but all samples came from there. This outcome likely occurred for two reasons. First, only fairly well-developed countries have extensive biometric research programs. Second, developed countries, apart from the United States, were until recently largely ethnically homogenous, with demographic changes occurring only within the last 30 or so years.

### 2.2. Data items

For each study included in the qualitative review, we recorded the: (1) study or survey from which the sample came (e.g., the Georgia Twin Study), (2) race/ethnic subgroups (e.g., White), (3) ACE estimates and intraclass correlations for subgroups and tests, (4) standard errors for the ACE estimates, when available, (5) samples sizes or percent sample sizes for each kinship class by race/ethnicity, (6) specific cognitive tests used, and (7) subgroup means and standard deviations, when available. We also coded any pertinent data regarding whether the authors or others reported concerns with the samples, methods of analysis, tests employed, etc.

#### 2.2.1. Data extraction and collection process

When a study reported intraclass correlations but did not report SEM or regression-based ACE estimates, we computed the estimates with Falconer's formula. Otherwise, we used reported ACE estimates if they were standardized. If they were not, we first set any negative ACE estimates to zero and then we standardized the values, so their total variance summed to one. We evaluated the effect of constraining estimates from 0 to 1, since doing so could potentially bias estimates of the variance components. Details are provided in Supplementary Materials Table S5b. We concluded that this was not a concern as only 4 of the 16 studies could have allowed for alternative estimates and all of these had small *N*s.

We also recorded means and standard deviations for all cognitive tests with ACE estimates. When sample means and standard deviations were neither provided nor available, we found or computed means and standard deviations based on the source data from which the sample came. For example, scores were not provided for the Add Health sample. However, we were able to obtain Wave I and III PPVT means and standard deviations by race/ethnicity from the publicly available Add Health survey data.

#### 2.2.2. Computations of ACE estimates and group difference effect sizes

We either computed or recomputed ACE estimates using Falconer's formula and its derivations (see, e.g., Plomin et al., 2014, pp. 381–382). The formulae are:

$$\mathrm{Var(P)} = a^2 + c^2 + e^2 \quad (1)$$

$$a^2 = 2\,(r\mathrm{MZ} - r\mathrm{DZ}), \quad (2)$$

$$c^2 = (r\mathrm{MZ} - a^2), \quad (3)$$

$$e^2 = (1 - r\mathrm{MZ}) \quad (4)$$

where Var(P) designates the phenotypic variance, $a^2$ (or A) designates variance attributable to additive genetics, $c^2$ (or C) designates variance attributable to the shared environment, $e^2$ (or E) designates variance attributable to the non-shared environment, *r*MZ designates the intraclass correlation for monozygotic or "identical" twins, and *r*DZ designates the intraclass correlations for dizygotic or "fraternal" twins.

For data from one study (Scarr-Salapatek, 1971), we first had to compute the identical twin intraclass correlations from the same sex twin correlations. To do this, we used the formula given by Loehlin et al. (1975, p. 288). We then used the results of this formula in the equations above.

For some samples, ACE estimates were given for multiple tests/subtests. For these, we created summary estimates by averaging ACE values within samples (across tests). For a few of these samples, additional subtests were given to only partial samples (e.g., Osborne, 1980, wherein 304 White sibling pairs were given the Basic Test Battery while only 63 of these pairs were given the Cattell Culture Fair test). In these cases, we weighted individual ACE estimates by the harmonic *N* for the kinship pairs to not lose information. We used the harmonic *N* because this considers unbalanced sibships used in computing ACE estimates (e.g., if one group comprised $N = 50$, but the other group is comprised of $N = 400$). The rationale for using harmonic *N* is detailed by te Nijenhuis and van den Hoek (2016), and we employ this statistic for the same reason. The formula is:

$$Nh = (N \times N)/(1/x_1 + 1/x_2) \quad (5)$$

where *N* is the number of samples, and $x_1$ is the sample size for group 1. When there were more than two groups, we used the largest group (usually full siblings) for $x_1$ and the remainder (e.g., half-siblings, cousins, adoptees) for $x_2$. Details are provided in Tables S3 and S4 of the Supplementary Materials (SM) file.

Preferably, across studies, ACE estimates would be weighted by the

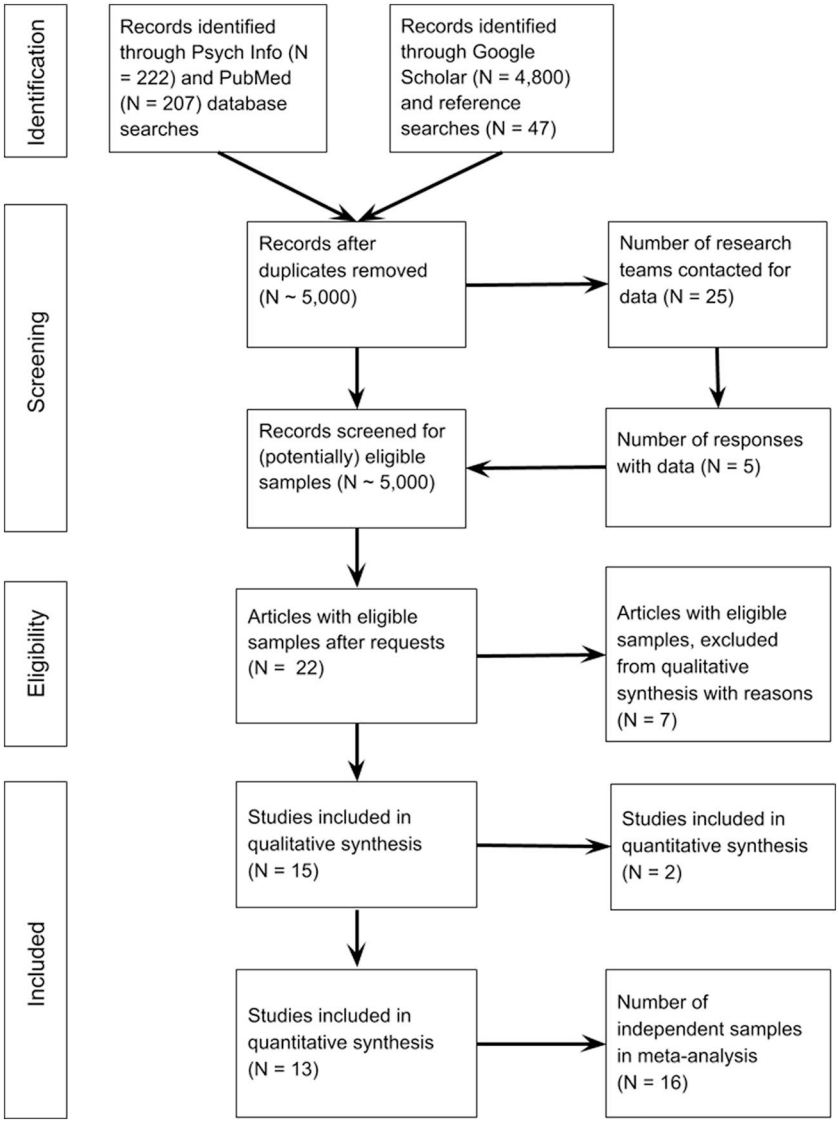


**Fig. 1.** Flowchart for study identification.

standard errors of the variance components, since kinship designs and samples differ regarding how precisely the variance components are estimated. However, except in two cases (see below) we did not have information to directly compute standard errors. That said, all things being equal, standard errors will be proportional to the theoretical, intraclass correlations (Falconer, 1960). Loehlin et al. (1975, pp. 288–289) provided a simplified formula for estimated standard errors based on the observed intraclass correlations and theoretical coefficients of relatedness. The formula is as follows:

$$\sigma h_B^2 \approx \frac{1}{rg1 - rg2}\sqrt{\frac{(1 - r_1^2)^2}{N1} + \frac{(1 - r_2^2)^2}{N2}} \quad (6)$$

where $N_1$ and $N_2$ are the number of pairs in the two groups, respectively, $r_g1$ and $r_g2$ are the theoretical coefficients of relatedness, respectively, and $r_1$ and $r_2$ are the observed intraclass correlations, respectively.

We used Loehlin's formula to estimate error variances (with two exceptions, see below), which we then used in the meta-analysis. When there were more than two groups, we used the largest group (usually full siblings) for $r_g1$ and $r_1$ and then the weighted average of the next largest group (e.g., half-siblings and cousins) for $r_g2$ and $r_2$. Details are provided in Tables S3 and S4 of the SM File. As for the two exceptions, Kevin Beaver (personal communication, October 3, September 24, 2013) provided confidence intervals for the Collaborative Perinatal Project and Add Health results, and Mollon et al. (2018; fig. 1) depicted confidence intervals for the Philadelphia Developmental Cohort. We calculated standard errors from the confidence intervals in these cases. These were computed as:

$$\text{S. E.} = \frac{\frac{1}{2} \times (Upper\ 95\%C.\ I. - h\text{^}2 + h\text{^}2 - Lower\ 95\%C.\ I.)}{1.96} \quad (7)$$

Because our estimated standard errors were imprecise, we also tried weighting by the harmonic *N*s of the samples as an alternative. Using these alternative weights did not alter our results.

Regarding effect sizes, we computed Cohen's *d* for all 40 pairs of groups, and for all test scores for which there were ACE estimates. When possible, we used data for the kinship sample specifically used in computing the ACE estimates. Otherwise, we used data from the survey where the kinship sample came from. Data sources are noted in Table S5 of the SM file. The formula for Cohen's *d* is:

$$d = (x_1 - x_2)/\text{pooled } SD \quad (8)$$

where *SD* stands for standard deviation and $x_1$ is the mean of Group 1.

**Table 1**
Description of included and excluded studies.

| Study | Race & ethnicity | Biometric type | Sample | Sample # | Excluded from review? |
|---|---|---|---|---|---|
| Scarr-Salapatek (1971) | CA & AA | SS and OS Twins | Philadelphia School Sample | 1 | No |
| Loehlin et al. (1975) | CA & AA | MZ & DZ Twins | The Collaborative Perinatal Project (CPP) 8 months | 2 | No, but from meta-analysis (redundant w/#3) |
| | CA & AA | MZ & DZ Twins | CPP Age 4 | | Yes (redundant w/#3) |
| Beaver et al. (2013) | CA & AA | Multiple Kinships | CPP Age 4 | 3 | No |
| | CA & AA | Multiple Kinships | CPP Age 7 | | (No, same sample as above) |
| Hodges (1976) | CA & AA & HA | MZ & DZ Twins | Los Angeles School Sample | 4 | No |
| Osborne and Miele (1969) | CA & AA | MZ & DZ Twins | The Georgia Twin Study | 5 | Yes (redundant w/#8) |
| Osborne and Gregor (1968) | CA & AA | MZ & DZ Twins | The Georgia Twin Study | 6 | Yes (redundant w/#8) |
| Vandenberg (1969) | | | The Louisville Twin Study | 7 | Yes (redundant w/#8) |
| Osborne (1980) | CA & AA | MZ & DZ Twins | The Twin Study | 8 | No |
| Scarr (1981) | CA & AA | MZ & DZ Twins | Philadelphia Twin Study | 9 | No |
| Scarr, Weinberg, and Waldman (1993) | CA & AA | Adoptees & Biological | The Minnesota Transracial Adoption Study (MTRAS) | 10 | No |
| Rowe et al. (1999) | CA & AA | Multiple Kinships | The National Longitudinal Study of Adolescent Health (Add Health) Wave 1 | 11 | Yes (redundant with #3) |
| Guo and Stearns (2002) | CA & AA | Multiple Kinships | Add Health Wave 1 | 12 | Yes (redundant with #3) |
| Beaver et al. (2013) | CA & AA | Multiple Kinships | Add Health Wave 1 | | (No, second sample from #3) |
| | CA & AA | Multiple Kinships | Add Health Wave 3 | | (No, second sample form #3) |
| Hart, Soden, Johnson, Schatschneider, and Taylor (2013) | CA & AA & HA & Other | MZ & DZ Twins | Florida Twin Project | 13 | No |
| Rhemtulla and Tucker-Drob (2012) | CA & non-White | MZ & DZ Twins | Early Childhood Longitudinal Study Birth Cohort (ECLS-B) | 14 | No |
| Zhang and Pierce (2014) | CA & AA | Individuals (GTCA) | Health and Retirement Study (HRS) | 15 | Yes (too small subsamples) |
| Woodley of Menie, Figueredo, Dunkel, and Madison (2015) | CA & AA & HA | MZ & DZ Twins | The Midlife in the United States (MIDUS II) | 16 | No |
| Figlio, Freese, Karbownik, and Roth (2017) | CA & AA & HA | SS & OS Twins | Florida Schools Sample (Grade 3 to 5) | 17 | No |
| Figlio et al. (2017) | CA & AA & HA | SS & OS Twins | Florida Schools Sample (Grade 6 to 8) | | (No, second sample form #17) |
| Mollon et al. (2018) | CA & AA | Individuals (SOLAR with SNP data) | Philadelphia Developmental Cohort | 18 | No |
| Engelhardt, Church, Paige Harden, and Tucker-Drob (2019) | CA & AA & HA & Asian American & Multiracial | MZ & DZ Twins | Texas Twin Study | 19 | No |
| Rowe and Cleveland (1996) | CA & AA | Full & Half Siblings | NLSY79 Children and Young Adults (CNLSY) | 20 | Yes (redundant with #21) |
| *Fuerst & Dalliard, 2014 | CA & AA & HA | Multiple Kinships | CNLSY | 21 | No, but from meta-analysis (redundant w/#22) |
| Pesta, Kirkegaard, te Nijenhuis, and Fuerst (2019) | CA & AA & HA | Multiple Kinships | CNLSY | 22 | No |
| Pesta et al. (2019) | CA & AA & HA | Multiple Kinships | National Longitudinal Survey of Youth 1979 (NLSY79) | | (No, second sample form #22) |

*Note:* CA – Caucasian American, AA – African American, HA – Hispanic American.

**Table 2**
Summary table for the meta-analysis.

| Study # | Study | Analysis type | Test types | Race/ethnicity | *Nh* | S.E. A | A | C | E |
|---|---|---|---|---|---|---|---|---|---|
| 1 | Scarr-Salapatek (1971) | Falconer's | IQ or g | White | 230 | 0.33 | 0.28 | 0.44 | 0.28 |
| 2 | Beaver et al. (2013) | SEM | IQ or g | White | 2604 | 0.11 | 0.48 | 0.28 | 0.25 |
| 3 | Hodges (1976) | Falconer's | IQ or g | White | 217 | 0.17 | 0.34 | 0.42 | 0.24 |
| 4 | Osborne (1980) | Falconer's | IQ or g | White | 299 | 0.16 | 0.54 | 0.28 | 0.18 |
| 5 | Scarr (1981) | Falconer's | IQ or g | White | 208 | 0.23 | 0.53 | 0.03 | 0.43 |
| 6 | Scarr et al. (1993) | Falconer's | IQ or g | White | 34 | 0.40 | 0.41 | 0.27 | 0.32 |
| 7 | Beaver et al. (2013) | SEM | IQ or g | White | 1544 | 0.08 | 0.52 | 0.20 | 0.28 |
| 8 | Rhemtulla and Tucker-Drob (2012) | SEM | Achievement | White | 320 | 0.11 | 0.37 | 0.63 | 0.00 |
| 9 | Hart et al. (2013) | Falconer's | Achievement | White | 465 | 0.10 | 0.80 | 0.02 | 0.18 |
| 10 | Woodley of Menie et al. (2015) | Falconer's | IQ or g | White | 586 | 0.11 | 0.14 | 0.46 | 0.40 |
| 11 | Figlio et al. (2017) | Regression | Achievement | White | 7444 | 0.05 | 0.57 | 0.16 | 0.27 |
| 12 | Figlio et al. (2017) | Regression | Achievement | White | 4747 | 0.07 | 0.59 | 0.18 | 0.24 |
| 13 | Mollon et al. (2018) | SOLAR | IQ or g | White | 2347 | 0.07 | 0.72 | | |
| 14 | Engelhardt et al. (2019) | SEM | IQ or g | White | 523 | 0.10 | 0.46 | 0.22 | 0.32 |
| 15 | Pesta et al. (2019) | SEM | IQ or g | White | 4596 | 0.11 | 0.67 | 0.17 | 0.16 |
| 16 | Pesta et al. (2019) | SEM | IQ or g | White | 229 | 0.38 | 0.61 | 0.29 | 0.10 |
| 3 | Hodges (1976) | Falconer's | IQ or g | Hispanic | 92 | 0.17 | 0.37 | 0.51 | 0.12 |
| 9 | Hart et al. (2013) | Falconer's | Achievement | Hispanic | 245 | 0.12 | 0.50 | 0.32 | 0.18 |
| 11 | Figlio et al. (2017) | Regression | Achievement | Hispanic | 2313 | 0.10 | 0.72 | 0.07 | 0.21 |
| 12 | Figlio et al. (2017) | Regression | Achievement | Hispanic | 1428 | 0.12 | 0.73 | 0.07 | 0.20 |
| 14 | Engelhardt et al. (2019) | SEM | IQ or g | Hispanic | 226 | 0.17 | 0.66 | 0.00 | 0.34 |
| 15 | Pesta et al. (2019) | SEM | IQ or g | Hispanic | 2532 | 0.10 | 0.80 | 0.12 | 0.08 |
| 16 | Pesta et al. (2019) | SEM | IQ or g | Hispanic | 122 | 0.54 | 0.24 | 0.46 | 0.30 |
| 1 | Scarr-Salapatek (1971) | Falconer's | IQ or g | Black | 448 | 0.29 | 0.31 | 0.32 | 0.37 |
| 2 | Beaver et al. (2013) | SEM | IQ or g | Black | 2399 | 0.10 | 0.51 | 0.18 | 0.31 |
| 3 | Hodges (1976) | Falconer's | IQ or g | Black | 172 | 0.24 | 0.20 | 0.38 | 0.42 |
| 4 | Osborne (1980) | Falconer's | IQ or g | Black | 116 | 0.29 | 0.59 | 0.13 | 0.28 |
| 5 | Scarr (1981) | Falconer's | IQ or g | Black | 154 | 0.24 | 0.48 | 0.11 | 0.41 |
| 6 | Scarr et al. (1993) | Falconer's | IQ or g | Black | 42 | 0.44 | 0.52 | 0.20 | 0.28 |
| 7 | Beaver et al. (2013) | SEM | IQ or g | Black | 542 | 0.20 | 0.45 | 0.20 | 0.34 |
| 9 | Hart et al. (2013) | Falconer's | Achievement | Black | 167 | 0.16 | 0.89 | 0.00 | 0.11 |
| 10 | Woodley of Menie et al. (2015) | Falconer's | IQ or g | Black | 10 | 0.85 | 0.94 | 0.00 | 0.06 |
| 11 | Figlio et al. (2017) | Regression | Achievement | Black | 2904 | 0.09 | 0.56 | 0.12 | 0.32 |
| 12 | Figlio et al. (2017) | Regression | Achievement | Black | 1789 | 0.11 | 0.48 | 0.20 | 0.32 |
| 13 | Mollon et al. (2018) | SOLAR | IQ or g | Black | 970 | 0.14 | 0.61 | | |
| 14 | Engelhardt et al. (2019) | SEM | IQ or g | Black | 72 | 0.30 | 0.13 | 0.45 | 0.42 |
| 15 | Pesta et al. (2019) | SEM | IQ or g | Black | 3544 | 0.09 | 0.81 | 0.12 | 0.07 |
| 16 | Pesta et al. (2019) | SEM | IQ or g | Black | 648 | 0.29 | 0.62 | 0.07 | 0.31 |
| 9 | Hart et al. (2013) | Falconer's | Achievement | Asian/Other | 68 | 0.40 | 0.24 | 0.22 | 0.54 |
| 14 | Engelhardt et al. (2019) | SEM | IQ or g | Asian/Other | 48 | 0.42 | 0.38 | 0.27 | 0.36 |
| 14 | Engelhardt et al. (2019) | SEM | IQ or g | Multi- racial | 62 | 0.48 | 0.43 | 0.00 | 0.57 |
| 8 | Rhemtulla and Tucker-Drob (2012) | SEM | Achievement | Non- White | 205 | 0.15 | 0.35 | 0.65 | 0.00 |

## 3. Meta-analytic methods

### 3.1. General procedure

Recently, various meta-analyses of heritability have been published (Polderman et al., 2015; van Houtem et al., 2013; Vukasović & Bratko, 2015; Willems, Boesen, Li, Finkenauer, & Bartels, 2019) and we followed their general procedures. The summary characteristics for the meta-analysis were ACE estimates for the 16 samples' summary cognitive measures together with the Harmonic *N*s and the standard errors of the heritability estimates. For the summary data, we (1) retained Mollon et al.'s (2018)*g*, (2) used our own summary estimate for Scarr's (1981) data, and (3) used our own CNLSY estimates for *g* (see the review in the SM). Note that for one of the samples (i.e., Mollon et al., 2018), we only had $h^2$ estimates. As such, the *k*s differ for some of the ACE components. Table 2 summarizes the data (see the SM files for additional data and computation).

### 3.2. Main analyses

We based the meta-analyses on a random-effects model which we conducted with the Hunter and Schmidt psychometric meta-analytic package (Schmidt & Le, 2004). Our measure of consistency was the percent of variance explained by sampling error between the data points. We conducted two meta-analyses, each using alternate weights. In one we used harmonic *N*s and in the other we used the standard errors of the heritability estimates. The latter adjusts for the precision of the estimates.

We meta-analyzed the ACE estimates for each racial/ethnic group, and then tested for the existence of moderator variables. The potential moderator variables were age (less than 12 years or not), biometric method (Falconer's formula or other), and measure of cognitive ability (academic achievement test scores or *g*/IQ). We selected these for several reasons. First, age is known to be a moderator of heritability estimates (Plomin et al., 2014). Second, Falconer's formula may give less reliable estimates relative to SEM or regression-based approaches. Third, tests which better measure *g* may more strongly predict racial and ethnic group differences, while also being more heritable (e.g., te Nijenhuis, Kura, & Hur, 2014). Following Schmidt and Hunter's (2015) recommendations regarding moderation, we split our database in two and applied meta-analytic techniques to each subset. Signs of moderation include finding substantially different mean correlations, together with an increase in the amount of variance explained by sample size.

We next meta-analyzed ACE estimate differences for the pairs of races/ethnicities in each sample. The comparison groups were Whites and each minority group. We did not include minority/minority group comparisons since the number of such comparisons was too small. For these analyses there were 26 dyadic comparisons. This difference-score analysis effectively controlled for several possible moderators, as the

ethnic subsamples in each main sample had (1) comparable ages, (2) ACE estimated the same way, and (3) taken the same cognitive tests.

### 3.3. Publication bias analysis

Publication bias is of paramount concern when conducting a meta-analysis. Bias exists when the meta-analyzed set of articles in some research domain fails to adequately represent the entire set of studies (published or not) in that domain. It may arise, for example, when studies finding some particular outcome are rejected more so than are other studies without the finding. Consequently, when publication bias exists, effect sizes derived via meta-analysis will over- or underestimate the true effect size. In testing for publication bias, we selected the random effects model and included funnel plots, Begg and Mazumdar's test (Begg & Mazumdar, 1994), Egger's test, and Duval and Tweedie's trim-and-fill method. Our analytic choices are detailed in the SM file.

### 3.4. Additional analysis

The Scarr-Rowe hypothesis predicts that *d* values will be positively correlated with heritability differences ($\Delta h^2$). Specifically, when *d* values are large, lower (higher) scoring populations should also show lower (higher) test heritabilities. We examined this prediction by looking at the correlation between *d* and $\Delta h^2$ across samples. We did so via four separate analyses. First, we used all pairs of *d* and $\Delta h^2$ estimates, with the first group being the higher-scoring group. Next, we looked at the associations between *d* and $\Delta h^2$ for White-Black, White-Hispanic, and Hispanic-Black comparisons. Other group pairings (e.g., White-multiracial) had too few samples ($k \leq 2$), so analyses were not possible. For these group comparisons, we instead weighted the data points by the pooled standard error of $h^2$, computed using the Satterthwaite Approximation. However, these results may not be robust, owing to possible confounding factors between samples (e.g., age differences). Therefore, as an alternative approach, we looked at the relation between *d* and $\Delta h^2$ within samples. Only three samples had a reasonable number of subtests (i.e., five or more), and so this analysis was limited to these three samples.

## 4. Review of studies and samples

The Supplement Material file contains a short review of each sample.

## 5. Meta-analytic results

### 5.1. Meta-analysis of ACE estimates

Meta-analytic results for the summary estimates appear in Table 3. Columns one through five, respectively, show the (1) various groups, (2) number of data points (*K*), (3) mean meta-analytic values of A, C, and E ($rho^2$), respectively, and (4) percentage of variance in the meta-analytic data points explained by sampling error (% var). The table shows that all heritabilities are moderate to high, except for those in the "Asian/Other" and "Non-White" categories. Estimates for these categories, however, were based on only one or two samples of data, with correspondingly high standard errors. The table also shows that almost all values of C and E are substantially smaller than are the values of A.

**Table 3**
Meta-analytic ACE outcomes for combinations of matched and unmatched groups by race/ethnicity.

| SIRE | A | | | C | | | E | | |
|---|---|---|---|---|---|---|---|---|---|
| | *K* | $rho^2$ | % var. | *K* | $rho^2$ | % var. | *K* | $rho^2$ | % var. |
| All groups | 42 | .60 | 2.02 | 40 | .17 | 9.87 | 40 | .23 | 9.74 |
| White | 16 | .58 | 2.31 | 15 | .20 | 7.61 | 15 | .24 | 14.73 |
| Black | 15 | .60 | 2.00 | 14 | .15 | 26.69 | 14 | .25 | 7.14 |
| Hispanic | 7 | .73 | 2.28 | 7 | .11 | 13.91 | 7 | .17 | 13.36 |
| Asian/Other | 2 | .30 | 306.6 | 2 | .24 | 2,569 | 2 | .47 | 137.04 |
| Multi-racial | 1 | .43 | n.a. | 1 | .00 | n.a. | 1 | .57 | n.a. |
| Non-White | 1 | .35 | n.a. | 1 | .65 | n.a. | 1 | .00 | n.a. |

*Note.* *K* = number of data points; $rho^2$ = mean meta-analytic value of, respectively, A, C, and E; % var = percentage of variance in the meta-analytic data points explained by sampling error.

### 5.2. Meta-analysis of matched groups

We analyzed all samples in Table 3. The estimates, however, are not directly comparable because the groups differed in terms of the samples in which they participated. For example, "Asian/Other" appeared in only two, relatively small samples; whereas, Whites appeared in all 16 samples. Table 4 shows analyses on matched groups (e.g., Whites and Blacks from the same study). These analyses comprised five comparisons: Whites versus Blacks, Whites versus Hispanics, Whites versus Asians/others, Whites versus multiracials, and Whites versus non-Whites. The design here is strong, as the groups are matched on several background variables. In each comparison we first meta-analyzed all data points (e.g., all 30 heritabilities of Blacks and Whites), and then we meta-analyzed the data points by the racial/ethnic group in the comparison (e.g., all 16 data points for Whites, plus all 16 data points for Blacks). This allowed us to compute differences between the meta-analytic estimates of the A, C, and E (denoted as $\Delta rho^2$) for each racial/ethnic group.

Almost all the $\Delta rho^2$ values are trivial or small, especially for those in Table 4 with a reasonable number of comparisons between samples (i.e., five or more): White-Black ($\Delta rho^2 = -.02$, 15 comparisons), White-Hispanic ($\Delta rho^2 = -.13$, seven comparisons), White-Asian/other ($\Delta rho^2 = .32$, two comparisons), White-multiracial ($\Delta rho^2 = .03$, one comparison), and White-non-White ($\Delta rho^{2=}$ .02, one comparison). Note that Scarr hypothesized a lower value of A for Hispanics relative to Whites. Here, however, the value of A is higher for Hispanics. Moreover, given the Scarr-Rowe hypothesis, one would expect a higher value of A for Asians relative to Whites, since Asians score better than Whites on IQ tests in the USA and since they also scored higher in the two samples we had data on. In earlier generations when Whites had higher social status (but still had lower test scores), Asians should have had lower heritabilities. We find a lower value of A for Asians relative to Whites. In sum, virtually no support exists for the Scarr-Rowe hypothesis as applied to race in the present data.

One could consider these differences in light of the effect originally reported by Turkheimer, Haley, Waldron, D'Onofrio, and Gottesman (2003). When treating SES as a dichotomous variable, Turkheimer et al. (2003) reported that the low SES group had a $h^2$ of .10 while the high SES group had a $h^2$ of .72. This represents a large effect by conventional standards. We find nothing like this in the present meta-analysis. Rather, our results are more in line with a recent meta-analysis of the Scarr-Rowe effect on SES, which showed little to no overall effect (Tucker-Drob & Bates, 2016), at least outside of the USA. Alternatively, one could compare the effects here to those that exist between age groups. Plomin et al. (2014, p. 202) reported that heritability increases significantly from approximately 40% in childhood to 80% in late adulthood. This ΔA represents a medium-sized effect, which we do not see here in the context of differences between self-identified racial/ethnic groups.

In almost all analyses, a very small amount of variance between data points was explained by sampling error. However, only the comparisons between Whites and Blacks have enough data points to allow for exploratory moderator analyses. We thus tested age, method, and measure as moderators for these groups. In the Schmidt and Hunter (2015) approach to testing for moderators in a meta-analysis, two things are taken into consideration. First, do the meta-analytical outcomes (meta-analytical correlations or meta-analytical means or other outcomes)

**Table 4**
Meta-analytic ACE outcomes for matched groups by race/ethnicity.

| SIRE | A | | | | C | | | | E | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | K | $rho^2$ | $\Delta rho^2$ | % var. | K | $rho^2$ | $\Delta rho^2$ | % var. | K | $rho^2$ | $\Delta rho^2$ | % var. |
| Whites vs Blacks | 30 | .59 | -0.02 | 2.13 | 28 | .18 | 0.05 | 14.12 | 28 | .25 | -0.01 | 10.08 |
| Whites | 15 | .58 | | 2.25 | 14 | .20 | | 10.62 | 14 | .24 | | 17.21 |
| Age ≤ 11 | 6 | .58 | | 2.45 | 6 | .18 | | 10.31 | 6 | .22 | | 20.60 |
| Age ≥ 12 | 9 | .57 | | 2.17 | 8 | .22 | | 11.66 | 8 | .26 | | 24.99 |
| SEM | 5 | .58 | | 2.88 | 5 | .21 | | 21.00 | 5 | .21 | | 14.15 |
| Not SEM | 10 | .57 | | 2.09 | 9 | .18 | | 8.75 | 9 | .26 | | 31.10 |
| IQ or g | 12 | .58 | | 2.00 | 11 | .23 | | 11.88 | 11 | .23 | | 15.96 |
| Achievement | 3 | .59 | | 5.56 | 3 | .16 | | 26.01 | 3 | .26 | | 48.98 |
| Blacks | 15 | .60 | | 2.00 | 14 | .15 | | 26.69 | 14 | .25 | | 7.14 |
| Age ≤ 11 | 6 | .64 | | 0.94 | 6 | .14 | | 21.84 | 6 | .22 | | 3.90 |
| Age ≥ 12 | 9 | .51 | | 12.00 | 8 | .22 | | 11.66 | 8 | .33 | | 196.14 |
| SEM | 5 | .66 | | 0.90 | 5 | .21 | | 21.00 | 5 | .20 | | 4.19 |
| Not SEM | 10 | .53 | | 6.81 | 9 | .17 | | 26.89 | 9 | .32 | | 59.58 |
| IQ or g | 12 | .62 | | 1.70 | 11 | .16 | | 26.96 | 11 | .22 | | 7.35 |
| Achievement | 3 | .54 | | 5.35 | 3 | .15 | | 26.83 | 3 | .31 | | 34.41 |
| Whites vs Hispanics | 14 | .64 | -0.13 | 2.23 | 14 | .15 | 0.06 | 13.48 | 14 | .22 | 0.06 | 11.53 |
| Whites | 7 | .60 | | 3.91 | 7 | .17 | | 22.15 | 7 | .23 | | 14.23 |
| Hispanics | 7 | .73 | | 2.28 | 7 | .11 | | 13.91 | 7 | .17 | | 13.36 |
| Whites vs Asians/others | 4 | .59 | 0.32 | 4.38 | 4 | .14 | -0.11 | 34.45 | 4 | .28 | -0.22 | 33.22 |
| Whites | 2 | .62 | | 2.67 | 2 | .13 | | 19.71 | 2 | .25 | | 36.35 |
| Asians/others | 2 | .30 | | 306.6 | 2 | .24 | | 2,569 | 2 | .47 | | 137.04 |
| Whites vs Multi-Racial | 2 | .46 | 0.03 | 2,533 | 2 | .20 | 0.22 | 69.52 | 2 | .35 | -0.25 | 45.11 |
| Whites | 1 | .46 | | n.a. | 1 | .22 | | n.a. | 1 | .32 | | n.a. |
| Multi-racial | 1 | .43 | | n.a. | 1 | .00 | | n.a. | 1 | .57 | | n.a. |
| Whites vs non-Whites | 2 | .36 | 0.02 | 3,033 | 2 | .64 | -0.02 | 1,414 | 2 | .00 | 0 | n.a. |
| Whites | 1 | .37 | | n.a. | 1 | .63 | | n.a. | 1 | .00 | | n.a. |
| Non-Whites | 1 | .35 | | n.a. | 1 | .65 | | n.a. | 1 | .00 | | n.a. |

*Note.* $K$ = number of data points; $rho^2$ = mean meta-analytic value of, respectively, A, C, and E; $\Delta\ rho^2$ = White $rho^2$ minus non-White $rho^2$; % var = percentage of variance in the meta-analytic data points explained by sampling error.

differ substantially when comparing subsets of the data? Second, does sampling error explain substantially more variance between the data points in the subsets than in all the data points in the meta-analysis? When both these outcomes are found, it is concluded there is a clear moderator in the meta-analysis.

Regarding Blacks, the effect sizes for A are slightly larger for studies with SEM in comparison to studies not using SEM, and also for younger children compared to older children. In contrast, these effects are not found for Whites. Also, the amount of variance between data points explained by sampling error does not increase systematically in comparison to the amount of variance in the analyses of all 30 data points. So, for A, we found no evidence of moderator effects. Very little support exists for the existence of moderators for C. This is also true for the analyses on the White samples for E. However, the Black samples for E show quite small differences in effect sizes between the subgroups for age and method used (SEM versus not SEM), but the amount of variance between the data points explained by sampling error does not increase systematically. Taken together, the data do not show clear evidence of moderator effects. An obvious limitation here is that the number of data points in several of the comparisons was quite modest.

### 5.3. Publication bias

The analyses for publication and experimenter bias were carried out via a random effects model. We conducted the analyses using both $N$ and the standard error. Moreover, we tested for experimenter bias using the trim-and-fill method which lets us choose the direction of the biasing effect (i.e., left or right of the inverted funnel). Fig. 2 shows the funnel plots for A, C, and E, respectively. The plot for A shows a hole on the left of the distribution, for the smaller studies, which indicates some experimenter bias. The plot for C shows a symmetrical distribution, and so there is no indication of publication bias. In the plot for E, there is a hole in the right of the distribution for the smaller studies, which indicates some experimenter bias.

Table 5 shows the outcomes of three additional tests of publication bias. The analyses show almost no sign of bias for either the Egger or the Begg tests, and at best only negligible effects using the trim-and-fill method in the case of E. Thus, all these tests combined show no indication of even a small amount of publication bias.

There is no conflict between these three outcomes and those derived when visually inspecting the funnel plots. This is because the missing values at the left or the right of the funnel plot for A and E occur almost only at the level of the smaller samples, so there are quite a few large studies in the meta-analytic database which dwarf the influence of the smaller studies. We conclude that overall there is no clear proof of publication bias, including an absence of clear proof for experimenter bias. Repeating the publication-bias analyses using *SE* instead of *N* leads to the same general conclusion.

### 5.4. Heritability × Group differences

The Scarr-Rowe hypothesis predicts that $d$ values will be positively correlated with heritability differences. The specific prediction is that when $d$ values are large, lower scoring populations should also show lower test heritabilities. We examined this prediction by looking at all samples together (e.g., White-Black, White-Hispanic, and Hispanic-Black). Surprisingly, given only small differences between groups in heritability, $\Delta h^2$ was consistently positively associated with $d$. The correlations for all pairs (higher-scoring groups minus the lower-scoring) were: $r$ = .59 ($N$ = 40; regression equation: $d$ = .64 + .52 $\Delta h^2$); White-Black, $r$ = .98 ($N$ = 15; regression equation: $d$ = 0.85 + 0.85 $\Delta h^2$); White-Hispanic, $r$ = .76 ($N$ = 7; regression equation: $d$ = .66 + .31$\Delta h^2$); Hispanic-Black, $r$ = .64 ($N$ = 7; regression equation: $d$ = .31 + .18 $\Delta h^2$). Note, for these analyses, we weighted values by the Satterthwaite approximation of the pooled error for heritability.

This pattern of correlations may represent a Scarr-Rowe effect of sorts. However, the pattern does not support the typical Scarr-Rowe

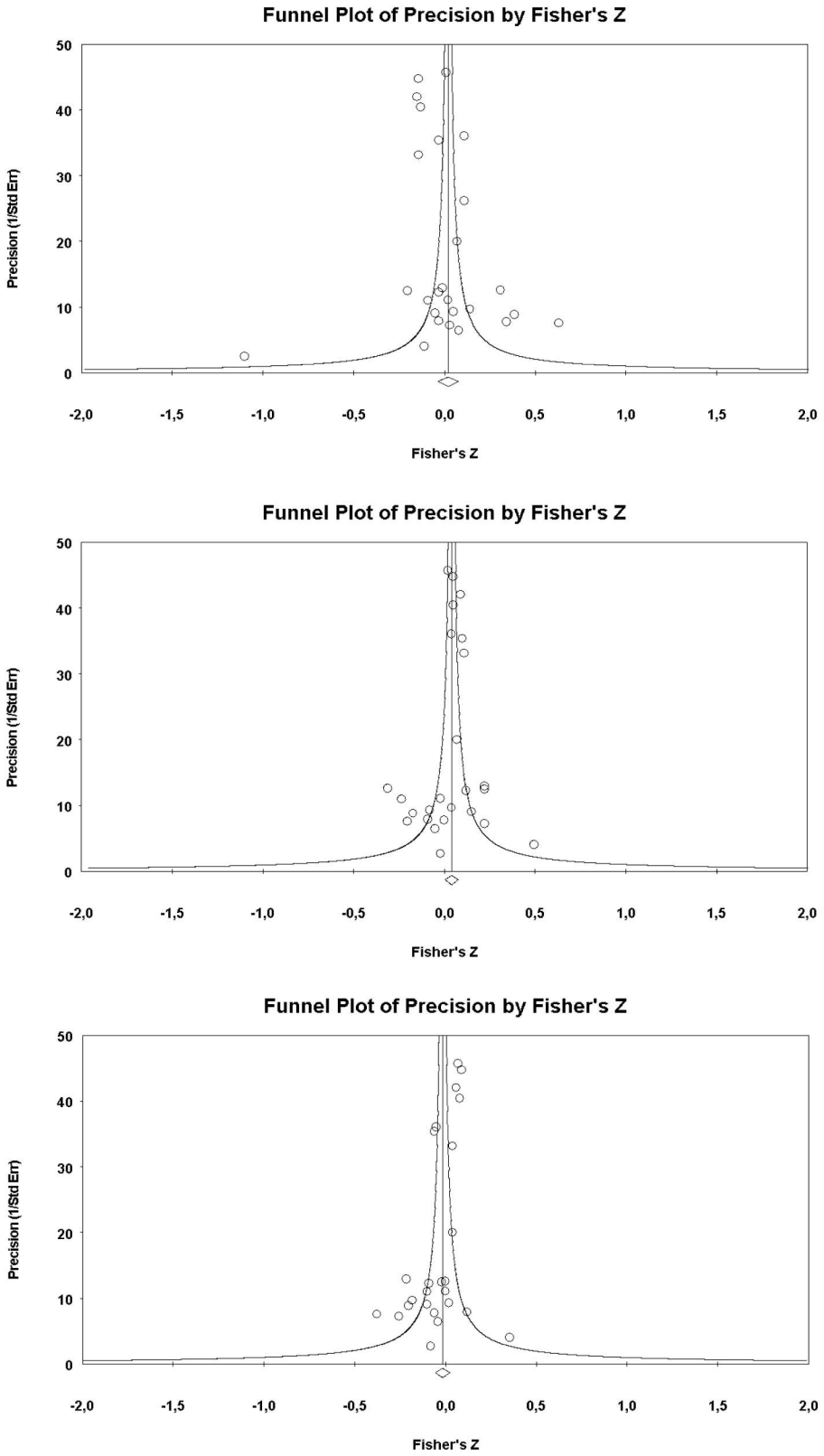


**Fig. 2.** Funnel plots of precision by Fisher's Z for A, C, and E, respectively. The x-axis shows Fisher's Z and the y-axis shows precision, measured as the inverse of the standard error.

interpretation that differences are due to a lack of "nourishing environmental experiences" (Tucker-Drob & Bates, 2016, p. 1). This can be seen by looking at the *d*s when delta $h^2 = 0$. This value is the point where "environmental equality" and "equal opportunity for mobility" (Selita & Kovas, 2019) are indicated. Based on the regression equations, setting $\Delta h^2$ to zero, the *d* values were: All pairs, $d = .64$, White-Black, $d = .85$, White-Hispanic, $d = .66$, and Hispanic-Black, $d = .31$. This pattern emerged perhaps because lower-scoring groups (about as often

**Table 5**
Outcomes of publication bias analyses and experimenter bias analyses for A, C, and E.

| Datasets analyzed | Egger | | Begg | | Trim-and-fill | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | Significance | Bias | Significance | Bias | Direction[a] | Number | Observed | Adjusted | Bias |
| A | .09 | No | .55 | No | **Left** | 0 | | | No |
| | | | | | Right | 2 | −.04 | −.04 | No |
| C | .38 | No | .74 | No | Left | 0 | | | No |
| | | | | | **Right** | 3 | .06 | .06 | No |
| E | .01 | Yes | .21 | No | Left | 0 | | | No |
| | | | | | **Right** | 7 | .03 | .04 | Yes |

[a] Direction indicates whether missing studies are expected at the right or the left side of the distribution. Directions in bold indicate where missing studies are expected in case of experimenter bias. The analyses are based on the sample size.

**Table 6**
Correlations between $\Delta h^2$ and d for samples with multiple ($\geq$ 5) subtests given at the same time.

| Study | *Nh* | Subtest *N* | *r* ($\Delta h^2 \times d$); (Regression Eq.) | *r* ($h^2 \times d$) |
|---|---|---|---|---|
| Osborne (1980) | 334.71 | 12 | −.26; ($d = 0.64 - 0.35\,\Delta h^2$) | −.11 |
| Scarr (1981) | 354.26 | 5 | .87; ($d = 0.70 + 0.96\,\Delta h^2$) | .18 |
| Mollon et al. (2018) | 2745.36 | 15 | −.23; ($d = 0.40 - 0.50\,\Delta h^2$) | .50 |
| All subtests | Unweighted | 32 | −.33; ($d = 0.53 - 0.35\,\Delta h^2$) | .44 |

*Note*: *Nh* is the harmonic *N* of the two samples' harmonic *N* for kinship pairs. Subtest *N* is the number of subtests.

as not) had higher heritabilities, relative to higher-scoring groups.

We next looked at the relationship between $\Delta h^2$ and *d* values within samples. First, we identified three samples (see Table 6) that had a reasonable number of subtests (i.e., five or more). We then ran the alternative analysis on these samples. Table 6 shows correlations between $\Delta h^2$ and *d* for each of these three samples (along with the regression equations). The results are inconsistent. The *r* ($\Delta h^2 \times d$) was positive only with the Scarr (1981) data. On the other hand, results for *r* ($h^2 \times d$), a prediction of a hereditarian model (Jensen, 1998), were also inconsistent, as also shown in Table 6. This is despite high correlations between *d* and *g* (*rs* = .65, .61, and .92) and modest ones between *g* and $h^2_{average}$(*rs* = .35, .79, and .43) for Osborne (1980), Scarr (1981), and Mollon et al. (2018). respectively.

## 6. Conclusion

We examined whether Scarr-Rowe interactions for race/ethnicity exist using a meta-analytic sample composed of Blacks, Whites, Hispanics, Asian/others, multiracials, and (non-specified) non-Whites from the United States. We tested the Scarr-Rowe hypothesis, which predicts that the heritability of intelligence will differ across these groups because of relative differences in the quality of each's environment. Our meta-analysis revealed that heritabilities were moderate to high for the major groups (Blacks, Whites, and Hispanics). We also found that genes accounted for about half of the IQ variance across groups, while shared and nonshared environmental effects explained the other half. Since the average age across samples was 15 (range of sample averages: 4 to 61; mean: 15; median: 12), our findings are consistent with those reported by Plomin et al. (2014). Additionally, there is no clear proof of publication bias or experimenter bias. It is possible that sometimes researchers simply do not submit manuscripts that do not support their hypothesis, but there is no proof for that in these studies.

Regarding the Scarr-Rowe hypothesis, we found that ethnic groups did not substantially differ in the heritability of intelligence (White-Black: $\Delta h^2 = -.02$, 15 comparisons; White-Hispanic: $\Delta h^2 = -.13$, 7 comparisons; White-Asians/others: $\Delta h^2 = .32$, 2 comparisons; Whites-multiracial: $\Delta h^2 = .03$, 1 comparison; and Whites-non-Whites: $\Delta h^2 = .02$, 1 comparison). This was despite moderate to large cognitive differences between some of the groups (e.g., White-Black mean $d = .83$ and White-Hispanic mean $d = .60$. Moreover, when we did find a non-trivial difference it was in the "opposite" direction from that predicted by Scarr-Rowe. For example, Hispanics had higher A values than Whites despite lower cognitive scores; Asians/others had lower A values than Whites, despite higher cognitive scores. In sum, the claim that the "heritability of intelligence, although never zero, is markedly lower among American children raised in poverty, at least with regard to less affluent "socially defined racial groups" (Turkheimer et al., 2017), is not supported.

A caveat is in order regarding whether the Scarr-Rowe interaction actually exists. When we looked across samples, we found evidence consistent with the interaction. Specifically, differences in within-group heritability covaried with the magnitude of the observed cognitive differences. However, evidence was lacking when we looked at the association in samples for which there were multiple subtests. What explains the differences we found when looking across versus within samples? The samples with multiple subtests generally did not result in precise $h^2$ estimates, at least judging from the reported standard errors. Across samples, however, several confounds existed, such as differences in age, methods of estimated $h^2$, and differences in cognitive measures. Regardless, even when looking across samples, the effect sizes for IQ differences were medium-to-large at the point in the regression plots where heritability was equal between higher- and lower-scoring races/ethnicities.

While Turkheimer et al. (2003) reported a Heritability × SES interaction for intelligence, they did not find a significant Heritability × Race/Ethnicity interaction (Turkheimer, personal communication, October 4, 2013). The same phenomena of an ACE × SES, but not an ACE × Race/Ethnicity interaction was found both in the Early Childhood Longitudinal Study (ECLS) analyzed by Rhemtulla and Tucker-Drob (2012) and in the Add Health sample analyzed by Beaver et al. (2013), Schwartz (2015), and others (Cf. Guo & Stearns, 2002). These results are consistent with the $\Delta h^2$ for SES contrasted with race, as found by Scarr-Salapatek (1971; see also Hart et al., 2013). Why this is the case is not clear. Regardless, we conclude that ACE × SES interactions, when found, are not being driven by ACE × Race/Ethnicity interactions and, as well, do not elicit ACE × Race/Ethnicity interactions.

Many researchers argue that heritability is an index of fairness in some sense. Various authors have labeled it as an indicator of "social justice" (Scarr, 1995), "equality of educational opportunity" (Krapohl et al., 2014), "equity" (Colodro-Conde, Rijsdijk, Tornero-Gómez, Sánchez-Romera, & Ordoñana, 2015), "equality of opportunity and

meritocracy," (Rimfeld et al., 2018), "environmental equality" (Selita & Kovas, 2019) or "social mobility" (Selita & Kovas, 2019). The reasoning is that when heritability is high, little room may exist for "environmental differences that convey privilege or privation" (Rimfeld et al., 2018, p. 270). Likewise, "where access to quality education is the privilege of the rich, differences in academic outcomes are largely due to socioeconomic disparities" (Selita & Kovas, 2019, p. 20).

Moreover, according to Scarr-Salapatek's (1971) environmental disadvantage hypothesis, groups with lower intelligence (due to being exposed to cognitively depressing environments) should exhibit attenuated heritabilities. This follows because "both [the environmental disadvantage and genetic difference hypotheses] make differential predictions about the proportions of genetic and environmental variance in IQ within lower and higher social class groups" (Scarr-Salapatek, 1971, p. 1286).

By the above logic, if racial/ethnic group differences in cognitive ability are due to poorer-quality environments, these groups should show lower heritabilities relative to people raised in more stimulating environments (i.e., the Scarr-Rowe interaction for race/ethnicity). While it is not always explicitly stated in the literature, by the same logic, the finding of similar heritabilities across advantaged/disadvantaged groups supports the genetic difference hypothesis (see, e.g., Jensen, 1968; Scarr-Salapatek, 1971; Selita & Kovas, 2019)

Our general findings are at odds with the predictions of Scarr-Salapatek's (1971) environmental disadvantage hypothesis. Scarr-Salapatek (1971) predicted that lower-scoring racial/ethnic groups would have substantially weaker genotype-phenotype correlations (heritabilities) than higher-scoring ones. It was assumed that the environmental factors causing the cognitive disadvantages would attenuate the genotype-phenotype correlations in the disadvantaged groups. The finding of similar genotype-phenotype correlations across groups could be because the alternative genetic hypothesis is correct. Alternatively, the results may imply that the general model's key prediction is incorrect. Perhaps "environmental disadvantage" between groups does not substantially lower heritability within groups, even when those groups themselves are disadvantaged in a cognitively impactful way. Given the present data, we suggest a re-evaluation of the Scarr-Rowe hypothesis. Proponents of the Scarr-Rowe hypothesis should try to model their predicted effects regarding group differences more explicitly.

Our results are relevant to the interpretation of the trans-ethnic predictive accuracy of education and intelligence polygenic scores. The predictive accuracy of polygenic scores is a function of both the validity of the scores and the within-group heritability of the traits in question. Rabinowitz et al. (2019) has shown that European polygenic scores weakly (and often insignificantly) predict African-American test scores (see, also, Lasker, Pesta, Fuerst, & Kirkegaard, 2019). They suggest that the attenuation relative to European samples could be due to the depressed heritability of cognitive ability among African Americans. The authors also cite Turkheimer et al.'s (2003) study which did not find heritability differences by race despite there being differences in SES by race. While this conjecture is reasonable (and could be explored with respect to other traits like education or social status attainment), our results indicate that the within-group heritability of cognitive ability is not depressed among African Americans. This suggests that the attenuated association for European-derived PGS is not likely to be due to a Scarr-Rowe interaction.

Our study has several limitations. These include: a limited number of studies, and only studies from the United States. In the future, new studies should be added to an updated meta-analysis. The research program begun by Scarr-Salapatek (1971) and popularized by Turkheimer et al. (2003) is a worthwhile endeavor. To advance the aims of this, in addition to SES, researchers are encouraged to publish heritability data by ethnicity, race, and culture.

In conclusion, our meta-analysis reveals that the heritability of cognitive ability is generally moderate to high for Whites, Blacks, and Hispanics in the United States. The other groups featured here (e.g., Asians) had sample sizes that were too small to allow making strong conclusions. We also found that differences in heritability across these three groups were mostly trivial. Nonetheless, we cannot rule out the existence of modest differences in population parameters in our analyses. We can, however, conclude that the correlations between phenotype and genotype are essentially the same for Whites, Blacks, and Hispanics residing in the USA.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.intell.2019.101408.

## References[1]


*Beaver, K. M., Schwartz, J. A., Connolly, E. J., Nedelec, J. L., Al-Ghamdi, M. S., & Kobeisy, A. N. (2013). The genetic and environmental architecture to the stability of IQ: Results from two independent samples of kinship pairs. *Intelligence, 41*, 428–438.

Begg, C. B., & Mazumdar, M. (1994). Operating characteristics of a rank correlation test for publication bias. *Biometrics, 50*, 1088–1101.

Bronfenbrenner, U., & Ceci, S. J. (1994). Nature-nurture reconceptualized in developmental perspective: A bioecological model. *Psychological Review, 101*, 568–586.

Colodro-Conde, L., Rijsdijk, F., Tornero-Gómez, M. J., Sánchez-Romera, J. F., & Ordoñana, J. R. (2015). Equality in educational policy and the heritability of educational attainment. *PLoS One, 10*, 111–117.

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

*Engelhardt, L. E., Church, J. A., Paige Harden, K., & Tucker-Drob, E. M. (2019). Accounting for the shared environment in cognitive abilities and academic achievement with measured socioecological contexts. *Developmental Science, 22*, 126–139.

Falconer, D. S. (1960). *Introduction to quantitative genetics.* Edinburgh: London: Oliver and Boyd.

*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, 84*, 291–299.

Fuerst, G. R. (2014). *ACE Analysis of the NLSY79 AFQT by Race/Ethnicity.* Unpublished manuscript.

*Fuerst, J., & Dalliard, M. (2014). Genetic and environmental determinants of IQ in black, white and hispanic americans: A meta-analysis and new analysis. *Open Behavioral Genetics*. https://doi.org/10.26775/OBG.2014.09.15.

Gottschling, J., Hahn, E., Beam, C. R., 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*(3), 881–910.

Hanscombe, K. B., Trzaskowski, M., Haworth, C. M., Davis, O. S., 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.

*Hart, S. A., Soden, B., Johnson, W., Schatschneider, C., & Taylor, J. (2013). Expanding the environment: Gene x School-level SES interaction on reading comprehension. *Journal of Child Psychology and Psychiatry, 54*(10), 1047–1055.

*Hodges, P. M. (1976). *Estimates of heritability in different populations (a preliminary report)*.

van Houtem, C. M. H. H., Laine, M. L., Boomsma, D. I., Ligthart, L., van Wijk, A. J., & de Jongh, A. (2013). A review and meta-analysis of the heritability of specific phobia subtypes and corresponding fears. *Journal of Anxiety Disorders, 27*, 379–388.

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

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

Johnson, B., Santana, V., Schupf, N., Tang, M. X., Stern, Y., Mayeux, R., & Lee, J. H. (2007). The heritability of abstract reasoning in caribbean latinos with familial Alzheimer's disease. *Dementia and Geriatric Cognitive Disorders, 24*, 411–417.

Krapohl, E., Rimfeld, K., Shakeshaft, N. G., Trzaskowski, M., McMillan, A., Pingault, J. B., ... Plomin, R. (2014). The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proceedings of the National Academy of Sciences, 111*(42), 15273–15278.

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

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

*Loehlin, J. C., Lindzey, G., & Spuhler, J. N. (1975). *Race differences in intelligence.* San Francisco: Freeman.

*Mollon, J., Knowles, E. E., Mathias, S. R., Gur, R., Peralta, J. M., Weiner, D. J., ... Glahn, D. C. (2018). Genetic influence on cognitive development between childhood and adulthood. *Molecular Psychiatry, 1*.


[1] Articles preceded by an asterisk appear in the meta-analysis.

te Nijenhuis, J., Kura, K., & Hur, Y. M. (2014). The correlation between g loadings and heritability in Japan: A meta-analysis. *Intelligence, 46*, 275–282.
te Nijenhuis, J., & van den Hoek, M. (2016). Spearman's hypothesis tested on black adults: A meta-analysis. *Journal of Intelligence, 4*(2), 6.
Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist, 67*(2), 130.
*Osborne, R. T. (1980). *Twins, black and white.* Athens: Foundation for Human Understanding.
Osborne, R. T., & Gregor, A. J. (1968). Racial differences in heritability estimates for tests of spatial ability. *Perceptual and Motor Skills, 27*, 735–739.
Osborne, R. T., & Miele, F. (1969). Racial differences in environmental influence on numerical ability as determined by heritability estimates. *Perceptual and Motor Skills, 28*, 535–538.
*Pesta, B. J., Kirkegaard, E. O. W., te Nijenhuis, J., & Fuerst, J. G. R. (2019). Racial and ethnic group differences in the heritability of intelligence: A systematic review and meta-analysis. *Intelligence*.
Plomin, R., DeFries, J. C., Knopik, V. S., & Neiderhiser, J. M. (2014). *Behavioral genetics* (6th ed.). New York: Worth Publishers.
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*, 702–709.
Rabinowitz, J. A., Kuo, S. I., Felder, W., Musci, R. J., Bettencourt, A., Benke, K., ... Ialongo, N. S. (2019). Associations between an educational attainment polygenic score with educational attainment in an African American sample. *Genes, Brain, and Behavior*. https://doi.org/10.1111/gbb.12558.
*Rhemtulla, M., & Tucker-Drob, E. M. (2012). Gene-by-socioeconomic status interaction on school readiness. *Behavior Genetics, 42*, 549–558.
Rimfeld, K., Krapohl, E., Trzaskowski, M., Coleman, J. R., Selzam, S., Dale, P. S., ... Plomin, R. (2018). Genetic influence on social outcomes during and after the Soviet era in Estonia. *Nature Human Behaviour, 1*, 212–222.
Rindermann, H. (2018). *Cognitive capitalism: Human capital and the wellbeing of nations.* Cambridge University Press.
Rowe, D. C., & Cleveland, H. H. (1996). Academic achievement in Blacks and Whites: Are the developmental processes similar? *Intelligence, 23*(3), 205–228.
Rowe, D. C., Jacobson, K. C., & Van den Oord, E. J. (1999). Genetic and environmental influences on vocabulary IQ: Parental education level as moderator. *Child Development, 70*, 1151–1162.
Rushton, J. P., & Jensen, A. R. (2005). Thirty years of research on race differences in cognitive ability. *Psychology, Public Policy, and Law, 11*, 235–294.
*Scarr, S. (1981). *Race, social class, and individual differences in IQ.* Hillsdale: Erlbaum.
Scarr, S. (1995). What is equality? Review of the bell curve. *Issues in Science and Technology, 11*(2), 82–85.
*Scarr, S., Weinberg, R. A., & Waldman, I. D. (1993). IQ correlations in transracial adoptive families. *Intelligence, 17*, 541–555.
*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.). Thousand Oaks, CA: Sage.
Schmidt, F. L., & Le, H. (2004). *Software for the Hunter-Schmidt meta-analysis methods.* Iowa City, IQ 42242: University of Iowa, Department of Management and Organization.
Schwartz, J. A. (2015). Socioeconomic status as a moderator of the genetic and shared environmental influence on verbal IQ: A multilevel behavioral genetic approach. *Intelligence, 52*, 80–89.
Selita, F., & Kovas, Y. (2019). Genes and Gini: What inequality means for heritability. *Journal of Biosocial Science, 51*(1), 18–47.
Sesardic, N. (2005). *Making sense of heritability.* Cambridge University Press.
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.
Tucker-Drob, E. M., Briley, D. A., & Harden, K. P. (2013). Genetic and environmental influences on cognition across development and context. *Current Directions in Psychological Science, 22*, 349–355.
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. (2011). The Scarr–Rowe interaction between measured socioeconomic status and the heritability of cognitive ability. *Experience and development* (pp. 89–106). Psychology Press.
Turkheimer, E., Harden, K. P., & Nisbett, R. E. (2017). Charles Murray is once again peddling junk science about race and IQ. *Vox*. Available at: https://www.vox.com/the-big-idea/2017/5/18/15655638/charles-murray-race-iqsam-harris-science-free-speech [Accessed July 29, 2017a] .
Turkheimer, E., & Horn, E. E. (2014). Interactions between socioeconomic status and components of variation in cognitive ability. In Behavior genetics of cognition across the Lifespan. *Advances in Behavior Genetics, 1*, 41–68.
Vandenberg, S. G. (1969). A comparison of heritability estimates of US Negro and White high school students. *Acta Geneticae Medicae et Gemellologiae, 19*, 280–284.
Vinkhuyzen, A. A., van der Sluis, S., Maes, H. H., & Posthuma, D. (2012). Reconsidering the heritability of intelligence in adulthood: Taking assortative mating and cultural transmission into account. *Behavior Genetics, 42*, 187–198.
Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin, 141*, 769–785.
Willems, Y. E., Boesen, N., Li, J., Finkenauer, C., & Bartels, M. (2019). The heritability of self-control: A meta-analysis. *Neuroscience and Biobehavioral Reviews, 100*, 324–334.
*Woodley of Menie, M. A. W., Figueredo, A. J., Dunkel, C. S., & Madison, G. (2015). Estimating the strength of genetic selection against heritable *g* in a sample of 3520 Americans, sourced from MIDUS II. *Personality and Individual Differences, 86*, 266–270.
Zanetti, D., & Weale, M. E. (2018). Transethnic differences in GWAS signals: A simulation study. *Annals of Human Genetics, 82*(5), 280–286.
Zhang, C., & Pierce, B. L. (2014). Genetic susceptibility to accelerated cognitive decline in the US Health and Retirement Study. *Neurobiology of Aging, 35*(6), 1512–1535.