Behavior Genetics
https://doi.org/10.1007/s10519-019-09982-7

ORIGINAL RESEARCH

Pathways Between a Polygenic Score for Educational Attainment and Higher Educational Attainment in an African American Sample

Jill A. Rabinowitz[1] · Sally I-Chun Kuo[2] · Benjamin Domingue[3] · Mieka Smart[4] · William Felder[1] · Kelly Benke[1] · Brion S. Maher[1] · Nicholas S. Ialongo[1] · George Uhl[5]

Received: 8 April 2019 / Accepted: 15 November 2019

Abstract

We investigated the extent to which performance on standardized achievement tests, executive function (EF), and aggression in childhood and adolescence accounted for the relationship between a polygenic score for educational attainment (EA PGS) and years of education in a community sample of African Americans. Participants ($N=402$; 49.9% female) were initially recruited for an elementary school-based prevention trial in a Mid-Atlantic city and followed into adulthood. In first and twelfth grade, participants completed math and reading standardized tests and teachers reported on participants' aggression and EF, specifically impulsivity and concentration problems. At age 20, participants reported on their years of education and post-secondary degrees attained and their genotype was assayed from blood or buccal swabs. An EA PGS was created using results from a large-scale GWAS on EA. A higher EA PGS was associated with higher education indirectly via adolescent achievement. No other mediating mechanisms were significant. Adolescent academic achievement is thus one mechanism through which polygenic propensity for EA influences post-secondary education among urban, African American youth.

Keywords Educational attainment · Polygenic score · Achievement · Executive function · Aggression · Childhood · Adolescence

Introduction

The last decade has witnessed a revolution in our ability to process and analyze molecular genetics information that has facilitated novel studies of genetic variation associated with education attainment (EA) and related phenotypes, correlates, and outcomes. Despite advances in molecular genetics technologies and NIH efforts to increase recruitment of diverse populations (H3Africa Consortium 2014), most of the genetic discoveries on EA, in addition to other phenotypes, have included predominantly individuals of European descent (Hindorff et al. 2018; Mills and Rahal 2019). The failure to include diverse populations, such as African Americans, is a significant problem given differences in environmental experiences (e.g., poverty, Albrecht et al. 2005) and allele frequencies across ethnic groups (1000 Genomes Consortium 2015) that may limit generalizability of findings derived from samples of European descent (H3Africa Consortium). Indeed, a number of studies have indicated an attenuation in the variance accounted for when the sample ancestry from genetic discoveries does not match that of the population under study (Martin et al. 2019). While there may be a number of historical, cultural, and logistical reasons why African Americans and other ethnic minority populations are underrepresented in genomics research (Bentley et al. 2017), intervention initiatives aimed

Handling Editor: Valerie S Knopik, Ph.D.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10519-019-09982-7) contains supplementary material, which is available to authorized users.

✉ Jill A. Rabinowitz
jrabino3@jhu.edu

1 Department of Mental Health, Johns Hopkins Bloomberg School of Public Heath, Baltimore, USA

2 Department of Psychology, Virginia Commonwealth University, Richmond, USA

3 Department of Sociology, Stanford University, Stanford, USA

4 College of Human Medicine, Michigan State University, East Lansing, USA

5 New Mexico VA Health Care System, Las Vegas, USA

at improving education and related outcomes for all people cannot be realized until such work is prioritized.

In terms of EA, among samples of European ancestry, genome-wide association studies (GWAS) have identified multiple genetic variants associated with higher educational attainment (Okbay et al. 2016; Lee et al. 2018) that have been linked to better scholastic performance, self-control, and interpersonal skills (Ward et al. 2014; Belsky 2016; Selzam et al. 2017). While a relationship has been established between polygenic propensity for EA and higher education, less is known about the phenotypic mechanisms that underpin this relationship. Moreover, it is also unclear whether EA polygenic influences play a greater role in the expression of certain phenotypes during some developmental periods compared to others. For example, results from twin studies suggest that genetics have a greater impact on the manifestation of behaviors as individuals age (Plomin and Deary 2015). In the current study, we examine (a) the extent to which various phenotypes (i.e., achievement, executive function (EF), aggression) explain the relationship between genetic propensity for EA and higher education (see Fig. 1 for conceptual model); and (b) whether these relationships change depending on the developmental period considered (i.e., childhood vs. adolescence). We build on previous work conducted among predominantly European ancestry individuals by investigating these relationships in a low-income urban, African American sample, a population that may experience a number of hardships and challenges that complicate the study of EA.

Phenotypic predictors of higher education

Research has shown that there are a number of important predictors of higher educational attainment. A natural candidate, given the ways in which educational experiences build on those that precede them, is academic achievement (Bierman et al. 2013). Children who perform better academically may be more likely to receive positive feedback from teachers that encourages subsequent learning; this learning may lead to selection into advanced classes that further promotes knowledge and skill acquisition. Moreover, individuals that excel academically may exhibit higher levels of academic self-efficacy and be better equipped to excel when faced with academic challenges in an educational setting (Honicke and Broadbent 2016).

In addition to academic achievement, other individual-specific factors, such as executive function (EF) and disruptive behavior problems (i.e., aggression), can influence educational aspirations and higher educational attainment. Higher levels of EF, defined as the ability to modulate cognitive and emotional states in an effort to achieve goals and adapt to environmental demands, has been positively associated with educational outcomes (Cartwright 2012). EF is often conceptualized as a multidimensional construct that includes cognitive and behavioral components such as inattention and impulsivity, respectively (Nigg 2017). For example, youth that display higher levels of inattention often have difficulty persisting on academic tasks and problem solving (Sayal et al. 2015; Colomer et al. 2017), which may confer

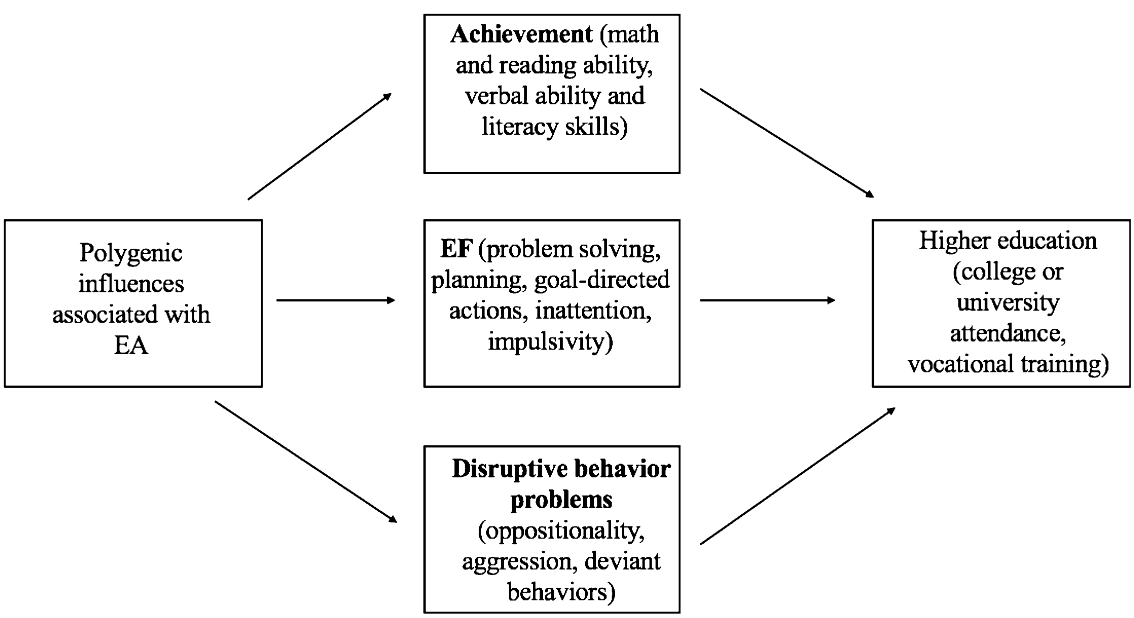

Fig. 1 Conceptual model linking polygenic influences associated with EA, achievement, EF, disruptive behavior problems, and higher education. *EA* educational attainment; *EF* executive function

risk for poorer performance on achievement tests (Merrell et al. 2017). Inattention has also been associated with less positive attitudes towards school (Colomer et al. 2017), behavior problems (Hoaken et al. 2003; Fantuzzo et al. 2005), and school failure (Rabiner et al. 2016). Children that are more impulsive are also at risk for school dropout and reduced scholastic performance, possibly because they may be more likely to act without considering the consequences of their actions, which may present challenges in a classroom setting that requires restraint (Valiente et al. 2013; Knouse et al. 2014; Colomer et al. 2017).

Moreover, a number of studies have linked aggressive behaviors to lower academic motivation (Frey et al 2009), grade retention (Bierman et al. 2013), and decreased academic performance (Hinshaw 1992; Metcalfe et al. 2013; Turney and Mclanahan 2015). It is possible that reduced scholastic achievement may increase the likelihood of youth feeling helpless and less motivated to academically excel, which may increase risk for behavior problems (Metcalfe et al. 2013). In addition, children who engage in disruptive, aggressive behaviors may be more likely to be disciplined and/or removed from classrooms (Horner et al. 2010), which may have a long-term impact on their educational success. Given the role that achievement, EF, and aggressive phenotypes play in higher education, an investigation into whether these phenotypes act as mediators in the relationship between polygenic propensity for EA and higher education is warranted.

Genetic associations with higher educational attainment

There is a significant corpus of work that has linked polygenic propensity for EA (assessed via polygenic scores) to achievement, EF, and aggression among individuals of European ancestry (Ward et al. 2014; Belsky et al. 2018). For example, a higher EA polygenic score (PGS) is associated with better academic performance on standardized tests, greater verbal ability, acquired reading skills at younger ages, lower levels of impulsivity, and a reduced likelihood of criminal offending (Ward et al. 2014; Belsky 2016; Selzam et al. 2017; Wertz et al. 2018). In addition, recent work conducted by Rabinowitz et al. (2019) using an urban, African American sample indicated that a higher EA PGS was positively associated with math, but not reading, achievement in early childhood.

While the literature has indicated direct associations between EA PGSs and achievement, EF, and aggressive phenotypes, it is unclear the extent to which these phenotypes account for the relationship between polygenic propensity for EA and higher educational attainment. There are, however, a number of reasons to expect that achievement, EF, and aggression are key pathways that influence the link between polygenic propensity for EA and higher education. As noted previously, a higher EA PGS may augment the likelihood of individuals performing well academically which may open up educational opportunities. Individuals with a greater polygenic propensity for EA may also manifest higher EF and be more equipped to suppress and shift behaviors in response to changing academic environments; these advantages may accordingly set youth on a trajectory towards higher education. In addition, greater interpersonal skills and decreased behavior problems frequently associated with polygenic propensity for EA may decrease the likelihood of academic disengagement and school disciplinary actions, increasing the likelihood of youth being able to pursue educational and/or professional opportunities. Thus, achievement, EF, and aggressive phenotypes may play an important role in the EA PGS-higher education association observed in previous studies, although there is a dearth of research in this area.

Developmental considerations

The effect of genetics on higher educational attainment may also vary depending on the developmental period during which academic achievement, EF, and aggressive behaviors are examined. For example, twin data have indicated that the heritability of cognitive ability increases by about 40% from infancy to adolescence (Plomin and Deary 2015). Other work has shown that the association of an EA PGS with academic achievement accounted for 14% of the variance in this outcome when youth were 16 years of age compared to 5% of the variance when youth were 7 years of age (von Stumm et al. 2019). Environmental experiences are thought to play a greater role in cognitive abilities in early childhood, whereas genetic influences often account for a greater amount of variance in this outcome during adolescence and adulthood (Plomin and Deary 2015). Consistent with the genetic amplification hypothesis, as youth become older and more autonomous from parents, they are able to seek out, modify, and create environments that reflect their genetic characteristics, thus increasing the impact of genetics on the manifestation of phenotypes (Plomin and Deary 2015). For example, relative to early childhood, adolescents generally have a greater choice in course enrollment and may select classes that reflect their interests and cognitive ability. Thus, when considering the mediating role of achievement, EF, and aggressive behaviors in terms of the relationship between polygenic propensity for EA and higher educational attainment, we would expect the variance accounted for by these variables to be attenuated in early childhood relative to adolescence.

The current study

The current study sought to build on our recent work demonstrating that an EA PGS was associated with higher education in a sample of urban, African Americans (Rabinowitz et al. 2019). Here, we extend this work by considering whether early childhood and adolescent academic achievement, EF, and aggression mediated the relationship between an EA PGS and higher education. We expected that a higher EA PGS would be associated with higher levels of academic achievement and EF and lower levels of aggression which in turn, would be associated with a greater likelihood of higher education. We also hypothesized that the mediators referenced above would account for a greater amount of variance in the EA PGS-higher education link in adolescence relative to early childhood.

Method

Participants

Participants ($N = 678$; 46.6% female; 86.3% African American) were originally recruited as first grade students in the fall of 1993 as part of a randomized controlled, universal preventive intervention trial in nine Mid-Atlantic urban elementary schools. Children were assigned to one of three conditions: (1) a classroom-centered intervention which focused on improving the curriculum, increasing behavior management, and helping students struggling academically; (2) a family-centered intervention which was designed to improve achievement and reduce early aggression, shy behavior, and concentration problems by enhancing parent–teacher communication and providing parents with effective teaching and child behavior management strategies; and (3) no intervention (Ialongo et al. 1999). The prevention trial and follow-up studies were approved by a University Institutional Review Board and participants provided informed consent as adults and assent prior to the age of 18. Additional information regarding the interventions can be found elsewhere (Ialongo et al. 1999).

We restricted the sample to individuals who self-identified as African American, had a successfully genotyped sample, and who reported on whether they attended a higher education institution ($n = 402$). Participant demographic information for the analytic sample is outlined in Table 1. About half the analytic sample was male and 66.4% were assigned to an intervention condition. The analytic sample generally reflects the characteristics of the whole sample (i.e., 678 participants) with respect to participant sex (whole sample, 53.4% males vs. analytic sample, 50.5% males) and percentage assigned to an intervention condition (whole sample, 67.7% vs. analytic sample, 66.4%).

Table 1 Sample characteristics

Characteristic	*n* (%)
Young adult sex	
Male	203 (50.5%)
Female	199 (49.5%)
Intervention status	
Yes	267 (66.4%)
No	135 (33.6%)
Young adult income	
< $10,000	187 (57.5%)
$10,000–$20,000	65 (20.0%)
$20,000–$35,000	60 (18.5%)
> $35,000	13 (4.0%)
Young adult education	
< High school diploma	88 (21.9%)
High school diploma only	114 (28.4%)
GED only	22 (5.5%)
> High school diploma	178 (44.3%)
Parent education[a]	
≤ High school diploma	191 (54.4%)
> High school diploma	160 (45.6%)
Parent income[a]	
< $5,000	38 (11.8%)
$5,000–$10,000	56 (17.4%)
$10,000–$20,000	73 (22.7%)
$20,000–$30,000	58 (18.0%)
$30,000–$40,000	44 (13.7%)
> $40,000	53 (16.4%)

GED general education diploma

[a]Parent education and income information were obtained when participants were in first grade

Measures

Participant demographic information

To help contextualize the sample, Table 1 presents caregiver level of education and income at baseline (fall of first grade) and participants' self-reported income and level of education at age 20.

Post-secondary education

At age 20, participants reported on the number of years of education they completed and degrees attained. About 22.0% of the sample did not report receiving a high school diploma, 28.4% of the sample reported just having a high school diploma, 5.5% reported receiving a General Education Diploma (GED), and 44.3% reported attending a higher education institution or vocational school. Given this distribution, we created a dichotomous variable to reflect individuals that had a high school education or GED or less (coded as 0) and individuals that reported attending a post-secondary education institution (e.g., vocational school, college/university) (coded as 1).

Standardized achievement test performance

In the fall of first grade, the Comprehensive Test of Basic Skills (CTBS) was administered and used to measure academic achievement (McGraw-Hill 1981). Subtests in the CTBS cover both verbal (word analysis, vocabulary, comprehension, spelling, and language mechanics and expression) and quantitative topics (computation, concepts, and applications). Standard scores for reading and mathematics achievement were provided, with higher scores reflecting higher academic achievement. The CTBS has shown convergent validity with other achievement tests (e.g., Academic Performance Rating Scale) and concurrent validity with teacher ratings of inattention and aggression (Atkins et al. 1989; Dupaul et al. 2013).

In the spring of twelfth grade, participants completed the reading and mathematics subtests from the Kaufman Test of Educational Achievement (KTEA; Kaufman and Kaufman 1985). The KTEA is an individually administered diagnostic battery that measures reading, mathematics, and spelling skills. We used the reading subtest (decoding and comprehension) from the brief form and the mathematics computation subtest from the comprehensive form. Each form provides age- and grade-based standard scores ($M = 100$, $SD = 15$), grade equivalents, percentile ranks, normal curve equivalents, and stanines. The KTEA norms are based on a nationally representative sampling of over 3000 children from grades 1 to 12. For the current study, the standard scores were used. Both the reading and mathematics subtests have shown convergent validity with other achievement measures (e.g., Peabody Individual Achievement Test, Wechsler Individual Achievement Test), excellent reliability (alpha > 0.90), and test–retest stability (Worthington 1987; Gentry et al. 1995).

Executive function and aggressive behaviors

Teachers completed (a) the Teacher Observation of Classroom Adaptation-Revised (TOCA-R; Werthamer-Larsson et al. 1991) when youth were in first grade; and (b) the Teacher Report of Classroom Behavior Checklist (TRCBC), an adaptation of the TOCA-R (Werthamer-Larsson et al. 1991), when youth were in twelfth grade. The TOCA-R is a structured interview that was administered to teachers by a trained research assistant, whereas the TRCBC is a checklist that was completed by teachers. Both measures assess teacher perceptions of children's aggressive/disruptive behaviors and EF. For the present study, the impulsivity and concentration problems subscales were used to index EF, and the aggressive/disruptive behaviors subscale was used to index aggressive behaviors. The impulsivity subscale assesses whether participants typically wait for their turn or blurt out answers to questions, whereas the concentration problems subscale measures the extent to which youth can pay attention and focus on a given task. The aggressive behavior subscale includes items such as whether youth hurt others or initiate fights with peers. Items across subscales were rated on a scale from 1 (*almost never*) to 6 (*almost always*). The subscales of the TOCA-R (Grades 1–3, Ialongo et al. 1999) and TRCBC (Grades 6–12, Petras et al. 2011; Liu et al. 2012) have shown predictive validity with a number of adult outcomes (e.g., marijuana use; Liu et al. 2013).

DNA and genotyping

In young adulthood, DNA was extracted from blood or saliva samples and was genotyped using Affymetrix 6.0 microarrays comprising 1 million single nucleotide polymorphisms (SNPs) across the genome (Affymetrix, Santa Clara, CA, US). Standard quality control steps were implemented to ensure accurate genotypes were included in subsequent analyses. Subjects with > 5% missing genotype data were removed. SNPs were also removed from further analysis when they had a minor allele frequency < 0.01, missingness > 0.05, or departures from Hardy–Weinberg equilibrium at $p < 0.0001$. These steps were performed using PLINK 2.0 (Chang et al. 2015). Genotypes were imputed to the 1000 Genomes Phase 3 reference panel (1000 Genomes Project Consortium 2010) using IMPUTE2 (Howie et al. 2009), whereas pre-phasing was performed in SHAPEIT (Delaneau et al. 2013). Resulting variants imputed with an INFO (quality) score < 0.8 were removed. Uncertainty adjusted dosage data, instead of called alleles, were used to generate the polygenic score.

When exploring genetic associations, it is important to identify and control for population stratification or genetic differences between subpopulations so that any significant associations observed are not confounded by ancestry (Cardon and Palmer 2003). We used principal components analysis in PLINK 2.0 to create the population stratification control variables (Chang et al. 2015). This process uses an orthogonal transformation to reduce the multi-dimensional genome-wide SNP data into a smaller number of genetic ancestry principal components (PCs). We used all the available measured SNPs (roughly 900,000) to generate these components. Although these were not a priori identified ancestry information markers, it has been shown that "randomly" selected SNPs perform equally as well (Pritchard and Rosenberg 1999). Our analytic sample includes individuals whose ancestry was geographically homogenous (see supplementary material for PC scatterplot). Consistent with prior work that has included 10 PCs in genetic analyses (e.g., Derks et al. 2012; Hartz et al. 2017; Vassos et al. 2017), we identified and controlled for 10 PCs which sufficiently accounted for population stratification.

PGS generation

We used the discovery sample results from a mega GWAS conducted recently by Lee et al. (2018). This GWAS included 1.1 million individuals pooled from a large number of samples, such as the Netherlands Twin Registry, Finnish Twin Cohort, Swedish Twin Registry, the Avon Longitudinal Study of Parents and Children, the UK Biobank, and 23andMe. The authors conducted a sample size weighted meta-analysis of SNPs that were associated with years of schooling completed, measured continuously.

Our genotype data contained 741,174 (26.3%) directly genotyped SNPs from the discovery list. After imputation, 2,554,305 (90.5%) SNPs from the discovery dataset were available in the current sample. Palindromic (A/T or C/G) SNPs were excluded, as methods for properly orienting multiple datasets are error-prone. LDPred, a method that includes direct modeling of linkage disequilibrium (LD) (Vilhjálmsson et al. 2015), was used to generate the PGS in our target sample using discovery results derived from the GWAS on EA referenced above. In this approach, the posterior mean effect size in a target sample is estimated based on the LD pattern in the target sample, as well as the LD pattern and effect sizes in the discovery sample. The reference sample used to estimate the LD pattern was based on a local sample that included individuals of European ancestry ($N = 336$), and was used to determine weights for each SNP that informed the PGS in our target African American sample. In this way, SNP pruning based on p-value thresholding is not necessary. The PGS was regressed on the ten principal components described above and standardized ($M = 0$, $SD = 1$). The residualized, standardized EA PGS was used in all the analyses.

Statistical analyses

Bivariate correlations and descriptive statistics were conducted to investigate the relations among the study variables using SPSS Version 25 (IBM 2017). The demographic and participant variables were coded as follows: (female = 0, male = 1; no intervention = 0, received an intervention = 1).

Multiple mediation analyses were conducted using Mplus (Muthén and Muthén 1998/2017) to examine the direct and indirect effects of the EA PGS on higher education attendance via early childhood and adolescent academic achievement, impulsivity, concentration problems, and aggression (see Fig. 2 for analytic model). We also controlled for intervention status when examining the outcomes that were assessed post-intervention (i.e., aggression, impulsivity, concentration problems, and academic achievement in twelfth grade). Parental education was also included as a covariate in our analyses given that parental education is robustly

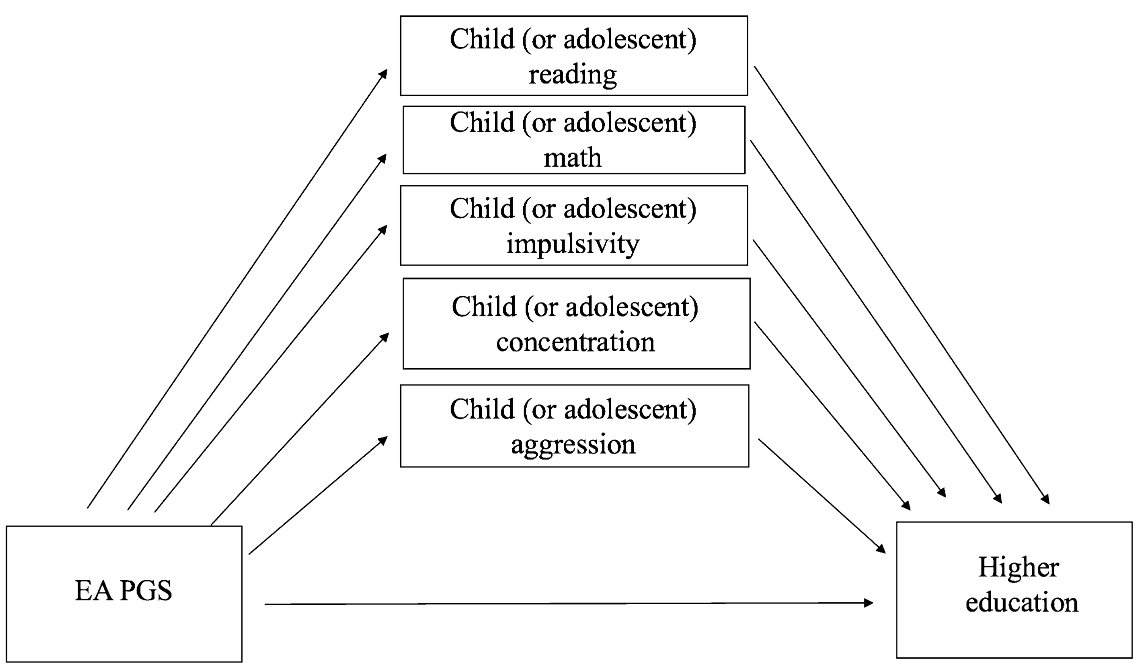

Fig. 2 Mediation models involving the EA PGS, early childhood and adolescent variables, and higher education. Separate analyses were conducted for each mediator

associated with children's achievement-related outcomes and educational attainment (Davis-Kean 2005; Grant et al. 2010; Belsky et al. 2018). Last, we controlled for early childhood variables when considering the adolescent variables given that early performance on achievement tests, EF, and aggression may persist into adolescence (Fergusson and Lynskey 1998; Rabiner et al. 2016). For example, when we examined twelfth grade aggression, first grade aggression was included in the model as a covariate. Indirect effects were assessed using bias-corrected bootstrapped (1000 times) 95% confidence intervals (CI) (Mackinnon and Luecken 2008). If the 95% CI for the indirect effect estimate did not include zero, it was concluded that the indirect effect was statistically significant (Shrout and Bolger 2002). Missing data was handled using full information maximum likelihood (FIML) (Enders 2011).

Results

Bivariate correlations between the EA PGS and study variables are presented in Table 2. Small positive correlations were observed between the EA PGS and early childhood math and reading achievement, impulsivity, and aggression; and a very small negative correlation was observed between the EA PGS and concentration problems. Moderate, positive correlations were observed between the EA PGS and adolescent math and reading achievement, and small, negative correlations were observed between the EA PGS and adolescent impulsivity, concentration problems, and aggression. Results from the primary analyses are presented below.

Participant sex ($b = -0.23$, $p < 0.005$) and parental education ($b = 0.18$, $p = 0.004$) were predictive of higher education. As shown in Table 3, first grade reading achievement ($b = 0.32$, $p < 0.005$), math achievement ($b = 0.17$, $p = 0.011$), concentration problems ($b = -0.29$, $p < 0.005$), and aggression ($b = -0.19$, $p = 0.006$) were predictive of higher education such that individuals who performed better on reading and math tests and whose teachers rated them as behaving less aggressively and as having fewer concentration problems were more likely to pursue higher education. The EA PGS was significantly positively associated with higher education ($b = 0.13$, $p = 0.036$) and first grade math achievement ($b = 0.12$, $p = 0.029$), but was not associated with reading achievement, impulsivity, concentration problems, or aggression. None of the mediation analyses associated with first grade achievement, impulsivity, concentration problems, or aggression were significant.

Twelfth grade reading ($b = 0.42$, $p < 0.005$) and math achievement ($b = 0.49$, $p < 0.005$) positively predicted higher educational attainment, whereas twelfth grade aggression ($b = -0.29$, $p = 0.001$) and concentration problems ($b = -0.48$, $p < 0.005$) negatively predicted higher education (Table 3). The EA PGS was significantly positively associated with twelfth grade reading ($b = 0.27$, $p = 0.001$) and math achievement ($b = 0.45$, $p < 0.005$).

In terms of the mediation analyses, there were significant indirect effects associated with twelfth grade reading ($b = 0.11$, 95% CI [0.01, 0.26]) and math achievement ($b = 0.22$, 95% CI [0.06, 0.54] such that a higher EA PGS was positively associated with adolescent math and reading achievement which, in turn, was associated with a greater likelihood of higher educational attainment. None of the other mediation analyses involving the adolescent phenotypes were significant.

Table 2 Bivariate correlations of the EA PGS and study variables

	1st grade	12th grade
Reading	0.06	0.20**
Math	0.12*	0.19**
Impulsivity	0.06	−0.09
Concentration problems	−0.04	−0.07
Aggression	0.05	−0.12*

[a]The EA PGS included in the analyses was regressed on the ten genetic ancestry principal components

*$p < .05$; **$p < .01$

Discussion

Upon the completion of secondary schooling, youth are faced with the decision to pursue higher education which often has a tremendous impact on their future earnings, occupation, and health across adulthood. While substantial advances in molecular genetics technologies have increased our understanding of the role that genetics play in post-secondary education, this work has failed to consider diverse populations, such as African Americans. African Americans in the United States often face barriers to educational attainment, including depressed social and economic conditions that may compromise their ability to succeed academically (Rothstein 2015). Moreover, phenotypes that yield one set of outcomes in majority groups may yield substantially different outcomes among African Americans (Conley and Conley 2009).

Thus, we argue that it is important to conduct preliminary studies in diverse populations, even though we still have relatively small samples in non-majority populations. We sought to harness recent scientific genetic discoveries on EA by examining potential pathways associated with EA polygenic influences and higher education among African Americans, which may inform developmental models of EA in this population. Using an African American, low-income sample, the present study examined whether a number of

Table 3 Standardized path coefficients, direct and indirect effects associated with the EA PGS, achievement, impulsivity, concentration problems, aggression, and higher education

Direct effects	*b* (SE)	*p*
EA PGS[a] → 1st grade reading	0.07 (0.06)	.220
EA PGS → 1st grade math	0.12 (0.06)	.029
EA PGS → 1st grade impulsivity	0.08 (0.06)	.177
EA PGS → 1st grade concentration problems	−0.03 (0.06)	.602
EA PGS → 1st grade aggression	0.07 (0.06)	.220
EA PGS → 12th grade reading	0.27 (0.08)	.001
EA PGS → 12th grade math	0.45 (0.10)	< .005
EA PGS → 12th grade impulsivity	−0.09 (0.08)	.234
EA PGS → 12th grade concentration problems	−0.07 (0.08)	.385
EA PGS → 12th grade aggression	−0.12 (0.09)	.166
EA PGS → higher education	0.13 (0.06)	.036
1st grade reading → higher education	0.32 (0.07)	< .005
1st grade math → higher education	0.17 (0.07)	.011
1st grade impulsivity → higher education	−0.11 (0.06)	.080
1st grade concentration problems → higher education	−0.29 (0.06)	< .005
1st grade aggression → higher education	−0.19 (0.07)	.006
12th grade reading → higher education	0.42 (0.07)	<005
12th grade math → higher education	0.49 (0.09)	< .005
12th grade impulsivity → higher education	−0.03 (0.07)	.664
12th grade concentration problems → higher education	−0.48 (0.06)	< .005
12th grade aggression → higher education	−0.29 (0.09)	.001
Indirect effects	***b* [95% CI]**	***p***
EA PGS → 1st grade reading → higher education	0.02 (−0.02, 0.07)	.231
EA PGS → 1st grade math → higher education	0.02 (−0.001, 0.07)	.103
EA PGS → 1st grade impulsivity → higher education	−0.10 (−0.05, 0.10)	.341
EA PGS → 1st grade concentration problems → higher education	0.01 (−0.04, 0.06)	.612
EA PGS → 1st grade aggression → higher education	−0.01 (−0.06, 0.01)	.308
EA PGS → 12th grade reading → higher education	0.11 (0.01, 0.26)	.009
EA PGS → 12th grade math → higher education	0.22 (0.06, 0.54)	.005
EA PGS → 12th grade impulsivity → higher education	0.003 (−0.01, 0.06)	.775
EA PGS → 12th grade concentration problems → higher education	0.03 (−0.08, 0.14)	.394
EA PGS → 12th grade aggression → higher education	0.04 (−0.03, 0.14)	.278

EA PGS educational attainment polygenic score. Separate analyses were conducted for each mediator

[a]The EA PGS included in the analyses was regressed on the ten genetic ancestry principal components

phenotypes (i.e., achievement, EF, and aggression) explained the association between polygenic propensity for EA and higher education.

We found that the EA PGS was positively associated with math achievement in early childhood, and reading and math achievement in adolescence. In addition, adolescent academic achievement mediated the relationship between the EA PGS and higher education such that a higher EA PGS predicted higher standardized test performance in adolescence which in turn, predicted post-secondary education, consistent with our hypotheses. Our findings are in line with twin studies indicating that the association between an EA PGS and academic performance increased in strength from childhood to adolescence (von Stumm et al. 2019). While the genes implicated in achievement and higher education are the same across the developmental course, genetics may better account for achievement-related outcomes as youth enter adolescence and young adulthood. Indeed, these developmental periods are often characterized by increased independence from caregivers and the family (Drabick and Steinberg 2011), which may allow youth to make choices and construct environments in line with their genetic predispositions. The qualities associated with greater polygenic propensity for EA, such as greater cognitive faculties and higher levels of achievement, may give youth greater confidence in their ability to excel and they may accordingly seek out more rigorous academic options, particularly during adolescence when they have greater volition to pursue them. Higher levels of adolescent academic achievement may also make youth more competitive for educational and professional opportunities, which may increase the likelihood of youth applying to and being granted entry into higher education institutions.

We found that childhood achievement, EF, and aggression did not mediate the relationship between the EA PGS and higher education. As noted previously, during early childhood, environmental factors in the family context, for example, may better explain youth's achievement, EF, and displays of aggression above that of genetic influences (Plomin and Deary 2015). Our findings also indicate that adolescent impulsivity, concentration problems, and aggression did not mediate the association between the EA PGS and higher education attainment. It is possible that EF and aggression did not explain the link between the EA PGS and post-secondary education because other factors better accounted for these behaviors, such as the quality of student–teacher relationships, the school climate, and peers. Indeed, more negative interactions with teachers has been associated with decreased student engagement (Wang and Eccles 2012; Martin et al. 2018), which may influence compliance with teachers' expectations and youth's ability and/or willingness to remain on task. In addition, it is developmentally normative for youth to engage in risk taking behaviors (e.g., rebelling from authority figures) in adolescence and their behaviors in the classroom may be heavily influenced by their peers and the classroom culture (Barth et al. 2004). Future research is needed to examine other factors in developmentally salient domains, such as the school context, which may account for why teacher reports of student behavior did not explain the association between polygenic propensity for EA and higher education.

Our results should be interpreted in the context of the following limitations. The EA PGS was based on a GWAS that included individuals of predominantly European ancestry. Prediction from GWAS is most accurate when the ancestry of the discovery sample matches the ancestry of the target sample (Scutari et al. 2016). Allelic heterogeneity and differences in allele frequencies and LD across populations may result in the most relevant SNPs for a given phenotype being missed if the ancestry of the discovery and target samples are dissimilar (Grinde et al. 2019). Indeed, SNP effect sizes among African Americans may differ from European ancestry samples due to epistatic influences or gene by environment interplay, which may result in a substantial reduction in the variance accounted for when discovery results from European ancestry GWAS are used to predict phenotypes in non-European ancestry samples (Martin et al. 2018). Recent approaches have been developed to better predict phenotypes in admixed populations, such as the multi-ethnic polygenic risk score approach, which involves estimating the optimal combination of summary statistics using discovery samples from European ancestry and smaller admixed samples (Marquez-Luna et al. 2017). Simulation and empirical work have shown that using both European-ancestry and admixed samples to estimate SNP weights improve the predictive utility for a number of traits in diverse samples (Kranzler et al. 2019; Grinde et al. 2019). Future work is warranted to replicate our findings using other methods that may reduce bias in cross-ancestry prediction models. Moreover, gene identification efforts must prioritize recruiting and maintaining diverse and underrepresented populations which may help increase prediction from polygenic scores to phenotypes and outcomes in diverse samples and has the potential to reduce health disparities (Bentley et al. 2017). Future research should also examine whether the present findings are generalizable to individuals of other ancestral backgrounds (Carlson et al. 2013).

In terms of next steps, consistent with a call for identifying nomological networks associated with GWAS discovery sample results (Belsky and Harden 2019), future research should identify whether genetic variants linked to higher education are associated with neurobiological, cognitive, and behavioral outcomes, which may help clarify the mechanisms through which the identified SNPs influence post-secondary attendance. Moreover, while most GWAS to date have identified genetic variants linked with EA among adults, the importance of examining intermediates associated with these genetic variants is warranted to identify how genetic influences unfold across development. Conducting this work in underrepresented populations is paramount to elucidate how polygenic propensity for EA, in conjunction with social contextual factors, influences student learning and educational success across the developmental course.

Compliance with Ethical Standards

Conflict of interest Jill A. Rabinowitz, Sally I-Chun Kuo, Benjamin Domingue, Mieka Smart, William Felder, Kelly Benke, Brion S. Maher, Nicholas S. Ialongo, and George Uhl declare that they have no conflict of interest.

Human and Animal Rights and Informed Consent All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee (Johns Hopkins Bloomberg School of Public Health Institutional Review Board #9223) and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

References

1000 Project Genome Consortium (2015) A global reference for human genetic variation. Nature 526:68–74. https://doi.org/10.1038/nature15393

Affymetrix Inc. Genome-Wide Human SNP Nsp/Sty 6.0 User Guide. Santa Clara, California: Affymetrix Inc. Rev 1; 2007. https://www.affymetrix.com/. Accessed 3 Nov 2018.

Albrecht DE, Albrecht CM et al (2005) Minority concentration, disadvantage, and inequality in the nonmetropolitan United States. Sociol Q 46:503–523

Atkins MS, Pelham WE, Licht M (1989) The differential validity of teacher ratings of inattention/overactivity and aggression. J Abnorm Child Psychol 17:423–435

Barth JM, Dunlap ST, Dane H et al (2004) Classroom environment influences on aggression, peer relations, and academic focus. J School Psychol 42:115–133

Belsky DW (2016) The genetics of success: How SNPs associated with educational attainment relate to life-course development. Psychol Sci 27:957–972. https://doi.org/10.1038/nprot.2015.121.Human

Belsky DW, Domingue BW, Wedow R et al (2018) Genetic analysis of social-class mobility in five longitudinal studies. Proc Natl Acad Sci USA 115:E7275–E7284. https://doi.org/10.1073/pnas.1801238115

Belsky DW, Harden KP (2019) Phenotypic annotation: Using polygenic scores to translate discoveries from genome-wide association studies from the top down. Curr Dir Psychol Sci 28:82–90. https://doi.org/10.1177/0963721418807729

Bentley AR, Callier S, Rotimi CN (2017) Diversity and inclusion in genomic research: why the uneven progress? J Commun Genet 8:255–266. https://doi.org/10.1007/s12687-017-0316-6

Bierman KL, Coie J, Dodge K et al (2013) School outcomes of aggressive-disruptive children: Prediction from kindergarten risk factors and impact of the Fast Track prevention program. Aggress Behav 39:114–130. https://doi.org/10.1002/ab.21467

Cardon LR, Palmer LJ (2003) Population stratification and spurious allelic association. Lancet 361:598–605

Carlson CS, Matise TC, North KE et al (2013) Generalization and dilution of association results from European GWAS in populations of non-European ancestry: The PAGE study. PLoS Biol 11:1–11. https://doi.org/10.1371/journal.pbio.1001661

Cartwright KB (2012) Insights from cognitive neuroscience: The importance of executive function for rarly reading development and education. Early Educ Dev 23:24–36. https://doi.org/10.1080/10409289.2011.615025

Chang CC, Chow CC, Tellier LCAM et al (2015) Second-generation PLINK: Rising to the challenge of larger and richer datasets. GigaScience 4:1–16. https://doi.org/10.1186/s13742-015-0047-8

Colomer C, Berenguer C, Roselló B et al (2017) The impact of inattention, hyperactivity/impulsivity symptoms, and executive functions on learning behaviors of children with ADHD. Front Psychol 8:1–10. https://doi.org/10.3389/fpsyg.2017.00540

Conley D, Conley D (2009) The promise and challenges of incorporating genetic data into longitudinal social science surveys and research. Biodemogr Soc Biol 55:238–251. https://doi.org/10.1080/19485560903415807

Comprehensive Test of Basic Skills (1981) Monterey, CA. CTB/McGraw-Hill, US

Davis-Kean PE (2005) The influence of parent education and family income on child achievement: The indirect role of parental expectations and the home environment. J Fam Psychol 19:294–304. https://doi.org/10.1037/0893-3200.19.2.294

Delaneau O, Zagury J, Marchini J (2013) Improved whole-chromosome phasing for disease and population genetic studies. Nat Methods 10:5–6. https://doi.org/10.1038/nmeth.2307

Derks EM, Vorstman JAS, Ripke S et al (2012) Investigation of the genetic association between quantitative measures of psychosis and schizophrenia: A polygenic risk score analysis. PLoS ONE 7:e37852. https://doi.org/10.1371/journal.pone.0037852

Drabick DAG, Steinberg L (2011) Developmental psychopathology. Encyclopedia of Adolescence, Three-Volume Set 3:136–142

Dupaul GJ, Perriello LM, Rapport MD (2013) Teacher ratings of academic skills: the development of the academic performance rating scale. School Psychol Rev 20:284–300

Enders CK (2001) A primer on maximum likelihood algorithms available for use with missing data. Struct Equ Modeling 8:128–141. https://doi.org/10.1207/s15328007sem0801_7

Fantuzzo JW, Bulotsky-Shearer R, Fusco RA et al (2005) An investigation of preschool classroom behavioral adjustment problems and social—emotional school readiness competencies. Early Child Res Q 20:259–275. https://doi.org/10.1016/j.ecresq.2005.07.001

Fergusson DM, Lynskey MT (1998) Conduct problems in childhood and psychosocial outcomes in young adulthood: a prospective study. J Emot Behav Disord 6:2–18

Frey A, Ruchkin V, Martin A et al (2009) Adolescents in transition: school and family characteristics in the development of violent behaviors entering high school. Child Psychiatry Hum Dev 40:1–13. https://doi.org/10.1007/s10578-008-0105-x

Gentry N, Sapp GL, Daw JL (1995) Scores on the Wechsler individual achievement test and the Kaufman test of educational-achievement comprehensive form for emotionally conflicted adolescents. Psychol Rep 76:607–610

Grant MD, Kremen WS, Jacobson KC et al (2010) Does parental education have a moderating effect on the genetic and environmental influences of general cognitive ability in early adulthood? Behav Genet 40:438–446. https://doi.org/10.1007/s10519-010-9351-3

Grinde KE, Qi Q, Thornton TA et al (2019) Generalizing polygenic risk scores from Europeans to Hispanics/Latinos. Genet Epidemiol 43:50–62

H3Africa Consortium (2014) Enabling the genomic revolution in Africa. Science 344:1346–1348

Hartz SM, Horton AC, Oehlert M et al (2017) Association between substance use disorder and polygenic liability to schizophrenia. Biol Psychiatry 82:709–715. https://doi.org/10.1016/j.biopsych.2017.04.020

Hindorff LA, Bonham VL, Brody LC et al (2018) Prioritizing diversity in human genomics research. Nat Rev Genet 19:175–185. https://doi.org/10.1038/nrg.2017.89

Hinshaw SP (1992) Externalizing behavior problems and academic underachievement in childhood and adolescence: causal relationships and underlying mechanisms. Psychol Bull 111:127–155

Hoaken PNS, Shaughnessy VK, Pihl RO (2003) Executive cognitive gunctioning and aggression: is it an issue ofiImpulsivity? Aggress Behav 29:15–30. https://doi.org/10.1002/ab.10023

Honicke T, Broadbent J (2016) The influence of academic self-ef fi cacy on academic performance: a systematic review. Educ Res Rev 17:63–84. https://doi.org/10.1016/j.edurev.2015.11.002

Horner SB, Fireman GD, Wang EW (2010) The relation of student behavior, peer status, race, and gender to decisions about school discipline using CHAID decision trees and regression modeling. J School Psychol 48:135–161. https://doi.org/10.1016/j.jsp.2009.12.001

Howie BN, Donnelly P, Marchini J (2009) A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet. https://doi.org/10.1371/journal.pgen.1000529

Ialongo NS, Werthamer L, Kellam SG et al (1999) Proximal impact of two first-grade preventive interventions on the early risk behaviors for later substance abuse, depression, and antisocial behavior. Am J Community Psychol 27:599–641

IBM Corp. Released, 2017. IBM SPSS Statistics for Windows, Version 25.0. IBM Corp., Armonk, NY.

Kaufman AS, Kaufman NL (1985) Kaufman test of educational achievement. American Guidance Service, Pines, MN

Kranzler HR, Zhou H, Kember RL et al (2019) Genome-wide association study of alcohol consumption and use disorder in 274, 424 individuals from multiple populations. Nat Commun 10:1499

Knouse LE, Feldman G, Blevins EJ (2014) Executive functioning difficulties as predictors of academic performance: examining the role of grade goals. Learn Individ Differ 36:19–26. https://doi.org/10.1016/j.lindif.2014.07.001

Lee JJ, Wedow R, Okbay A et al (2018) Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nat Genet 23:1112–1121. https://doi.org/10.1038/s41588-018-0147-3

Liu W, Lee GP, Goldweber A et al (2012) Impulsivity trajectories and gambling in adolescence among urban male youth. Addiction 108:780–788. https://doi.org/10.1111/add.12049

Liu W, Lynne-Landsman SD, Petras H, Masyn K, Ialongo N (2013) The evaluation of two first-grade preventive interventions on childhood aggression and adolescent marijuana use: a latent transition longitudinal mixture model. Prev Sci 14:206–217. https://doi.org/10.1007/s11121-013-0375-9

Mackinnon DP, Luecken LJ (2008) How and for whom? Mediation and moderation in health psychology. Health Psychol 27:99–100. https://doi.org/10.1037/0278-6133.27.2(Suppl.).S99

Martin AR, Kanai M, Kamatani Y, et al (2018) Hidden 'risk' in polygenic scores: Clinical use today could exacerbate health disparities. bioRxiv 1–26

Martin AR, Kanai M, Kamatani Y, et al (2019) Clinical use of current polygenic risk scores may exacerbate health disparities. Nat Genet 51:584–591. https://doi.org/10.1038/s41588-019-0379-x

Marquez-Luna C, Loh P-R, South Asian Type II Diabetes (SAT2D) Consortium, et al (2017) Multi-ethnic polygenic risk scores improve risk prediction in diverse populations. Genet Epidemiol 41:811–823

Merrell C, Sayal K, Tymms P, Kasim A (2017) A longitudinal study of the association between inattention, hyperactivity and impulsivity and children's academic attainment at age 11. Learn Individ Differ 53:156–161. https://doi.org/10.1016/j.lindif.2016.04.003

Metcalfe LA, Harvey EA, Laws HB (2013) The longitudinal relation between academic/ cognitive skills and externalizing behavior problems in preschool children. J Educ Psychol 105:881–894. https://doi.org/10.1037/a0032624

Mills MC, Rahal C (2019) A scientometric review of genome-wide association studies. Commun Biol 9:2. https://doi.org/10.1038/s42003-018-0261-x

Muthén LK, Muthén BO (1998–2017) Mplus user's guide. 8th edition. Muthén and Muthén, Los Angeles, CA.

Nigg JT (2017) Annual research review: On the relations among self-regulation, self-control, executive functioning, effortful control, cognitive control, impulsivity, risk taking, and inhibition for developmental psychopathology. J Child Psychol Psychiatry 4:361–383. https://doi.org/10.1111/jcpp.12675

Okbay A, Beauchamp JP, Fontana MA et al (2016) Genome-wide association study identifies 74 loci associated with educational attainment. Nature 533:539–542. https://doi.org/10.1038/nature17671

Petras H, Masyn K, Ialongo N (2011) The developmental impact of two first grade preventive interventions on aggressive/disruptive behavior in childhood and adolescence: an application of latent transition growth mixture modeling. Prev Sci 12:300–313. https://doi.org/10.1007/s11121-011-0216-7

Plomin R, Deary IJ (2015) Genetics and intelligence differences: five special findings. Mol Psychiatry 20:98–108. https://doi.org/10.1038/mp.2014.105

Pritchard JK, Rosenberg NA (1999) Use of unlinked genetic markers to detect population stratification in association studies. Am J Hum Genet 65:220–228. https://doi.org/10.1086/302449

Rabiner DL, Godwin J, Dodge KA (2016) Predicting academic achievement and attainment: the contribution of early academic skills, attention difficulties, and social competence. Sch Psychol Rev 45:250–267

Rabinowitz JA, Kuo SI, Felder W et al (2019) Associations between an educational attainment polygenic score with educational attainment in an African American sample Genes. Brain Behav. https://doi.org/10.1111/gbb.12558

Rothstein R (2015) The racial achievement gap, segregated schools, and segregated neighborhoods: a constitutional insult. Race Soc Probl 7:21–30. https://doi.org/10.1007/s12552-014-9134-1

Sayal K, Washbrook E, Propper C (2015) Childhood behavior problems and academic outcomes in adolescence: longitudinal population-based study. J Am Acad Child Adolesc Psychiatry 54:360–368. https://doi.org/10.1016/j.jaac.2015.02.007

Scutari M, Mackay I, Balding D (2016) Using genetic distance to infer the accuracy of genomic prediction. PLoS Genet 12:1–19. https://doi.org/10.1038/ng1840

Selzam S, Krapohl E, Von Stumm S et al (2017) Predicting educational achievement from DNA. Mol Psychiatry 22:267–272. https://doi.org/10.1038/mp.2016.107

Shrout PE, Bolger N (2002) Mediation in experimental and nonexperimental studies: new procedures and recommendations. Psychol Methods 7:422–445. https://doi.org/10.1037//1082-989X.7.4.422

The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467:1061–1073

Turney K, Mclanahan S (2015) The academic consequences of early childhood problems. Soc Sci Res 54:131–145. https://doi.org/10.1016/j.ssresearch.2015.06.022

Valiente C, Eisenberg N, Tracy L et al (2013) Effortful control and impulsivity as concurrent and longitudinal predictors of academic achievement. J Early Adolesc 33:946–972. https://doi.org/10.1177/0272431613477239

Vassos E, Di FM, Coleman J et al (2017) An examination of polygenic score risk prediction in individuals with first-episode psychosis. Biol Psychiatry 81:470–477. https://doi.org/10.1016/j.biopsych.2016.06.028

Vilhjálmsson BJ, Yang J, Finucane HK et al (2015) Modeling linkage disequilibrium increases accuracy of polygenic risk scores. Am J Hum Genet 97:576–592. https://doi.org/10.1016/j.ajhg.2015.09.001

Von Stumm S, Smith-Wooley E, Ayorech Z, et al (2019) Predicting educational achievement from genomic measures and socioeconomic status. bioRxiv. https://doi.org/10.1101/538108

Wang M, Eccles JS (2012) Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. Child Dev 83:877–895. https://doi.org/10.1111/j.1467-8624.2012.01745.x

Ward ME, McMahon G, St Pourcain B et al (2014) Genetic variation associated with differential educational attainment in adults has anticipated associations with school performance in children. PLoS ONE 9:1–7. https://doi.org/10.1371/journal.pone.0100248

Werthamer-Larsson L, Kellam S, Wheeler L (1991) Effect of first-grade classroom environment on shy behavior, aggressive behavior, and concentration problems. Am J Commun Psychol 19:585–602. https://doi.org/10.1007/BF00937993

Wertz J, Caspi A, Belsky DW et al (2018) Genetics and crime: Integrating new genomic discoveries into psychological research about antisocial behavior. Psychol Sci 29:791–803. https://doi.org/10.1177/0956797617744542

Worthington CF (1987) Kaufman test of educational achievement, comprehensive form and brief form. J Couns Dev 65(6):325–327