

**Ziada Ayorech, Robert Plomin and [Sophie von Stumm](#)**  
**Using DNA to predict educational trajectories in early adulthood**

**Article (Accepted version)  
(Refereed)**

**Original citation:**

Ayorech, Ziada and Plomin, Robert and von Stumm, Sophie (2018) Using DNA to predict educational trajectories in early adulthood. [Developmental Psychology](#). ISSN 0012-1649 (In Press)

© 2018 [American Psychological Association](#)

This version available at: <http://eprints.lse.ac.uk/id/eprint/90958>

Available in LSE Research Online: December 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (<http://eprints.lse.ac.uk>) of the LSE Research Online website.

This document is the author's final accepted version of the journal article. There may be differences between this version and the published version. You are advised to consult the publisher's version if you wish to cite from it.

## **Using DNA to predict educational trajectories in early adulthood**

Ziada Ayorech<sup>1</sup>, Robert Plomin<sup>1</sup>, Sophie von Stumm<sup>2</sup>

<sup>1</sup> King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, London, SE5 8AF, UK

<sup>2</sup> Department of Education, University of York, University of York, Heslington, York, YO10 5DD, UK.

Correspondence: Ziada Ayorech, King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology & Neuroscience, London, SE5 8AF, UK. Email: [ziada.ayorech@kcl.ac.uk](mailto:ziada.ayorech@kcl.ac.uk)

**Abstract**

At the end of compulsory schooling, young adults decide on educational and occupational trajectories that impact their subsequent employability, health and even life expectancy. To understand the antecedents to these decisions, we follow a new approach that considers genetic contributions, which have largely been ignored before. Using genome-wide polygenic scores (EA3) from a genome-wide association (GWA) study of years of education in 1.1 million individuals, we tested for genetic influence on early adult decisions in a UK-representative sample of 5,839 18 year olds. EA3 significantly predicted educational trajectories in early adulthood (Nagelkerke  $r^2 = 10\%$ ;  $\chi^2(4) = 571.77$ ,  $p < .001$ ), indicating that young adults partly adapt their aspirations to their genetic propensities—a concept known as *gene-environment correlation*. Compared to attending university, a one standard deviation increase in EA3 was associated on average with a 51% reduction in the odds of pursuing full-time employment (OR=0.47; 95% CI [0.43-0.51]), an apprenticeship (OR=0.49; 95% CI [0.45-0.54]), or becoming NEET (Not in Education, Employment or Training; OR=0.50; 95% CI [0.41-0.60]). EA3 associations were attenuated when controlling for previous academic achievement and family socioeconomic status. Overall this research illustrates how DNA-based predictions offer novel opportunities for studying the socio-developmental structures of life outcomes.

**Keywords:** emerging adulthood, education, NEET, polygenic score, behavioral genetics

### **Introduction**

Emerging adulthood is a unique developmental period in the late teens and early twenties when young people gain autonomy over their life decisions (Arnett, 2014). In most countries, the onset of emerging adulthood coincides with the end of compulsory schooling, when young adults can choose for the first time whether or not to continue formal education. Young adults select diverse paths, with some opting to pursue a university degree, others seeking full-time employment or starting a family, and others delaying or completely disengaging in economic activities. Those who neither enroll in further education nor engage with the labor market – that is, individuals who are Not in Education, Employment or Training (NEET) -- are at greater risk of later unemployment, marginalization, criminality, poor mental and physical health, and lower life expectancy (Bradley & Corwyn, 2002; Crowley, Jones, Cominetti, & Gulliford, 2013; Feng, Ralston, Everington, & Dibben, 2017; Goldman - Mellor et al., 2016; Kelly, Dave, Sindelar, & Gallo, 2014). Each of these outcomes bears substantial costs for the individual and also constitutes an economic burden to society. By comparison, the economic benefits of increased educational and occupational trajectories within society translate to a higher standard of living overall.

For these reasons, understanding the antecedents to decisions on educational trajectories during emerging adulthood is of major societal importance.

### **Genetics of educational trajectories**

Research on predicting individual differences in educational and occupational trajectories has largely focused on differences in the socioeconomic status (SES) of the family in which children are reared, or in their educational achievement (Bradley & Corwyn, 2002; Hart & Risley, 1995; Van de Werfhorst, 2001; Werfhorst, Sullivan, & Cheung, 2003; I. R. White, Blane, Morris, & Mourouga, 1999). Results indicate children of high-SES families attain higher levels of education (Parker et al., 2012), select more prestigious degrees (Leppel, Williams, & Waldauer, 2001) and secure the highest earning professions in later life (Macmillan, Tyler, & Vignoles, 2015) when compared to their low-SES peers.

Although SES and educational achievement are typically viewed as environmental variables, quantitative genetic findings indicate that they show substantial genetic influence, like other complex human traits (Knopik, Neiderheiser, DeFries, & Plomin, 2017; Robert Plomin, DeFries, Knopik, & Neiderheiser, 2016). For example, both family SES and children's educational achievement show significant genetic influence (Calvin et al., 2012; Krapohl & Plomin, 2015; Trzaskowski et al., 2014). Moreover, correlations between SES and educational achievement are in part genetically mediated (Belsky et al., 2016; Marioni et al., 2014). It is reasonable to assume then that these genetically influenced differences impact decisions on educational trajectories, so that individuals make educational choices that are in line with their educational aptitudes and their genetic propensities. Finding genetic influence on broad life choices at the beginning of emerging adulthood would indicate that young people are not passive recipients of their environment but instead actively

modify and create their experiences in part due to their genetic propensities—a concept known as *gene-environment (GE) correlation*.

GE correlation has traditionally been investigated using family-based methods including the twin method, which compares identical and fraternal twin pairs. However only two studies have applied the twin design to study educational choices (Ayorech, Krapohl, Plomin, & von Stumm, 2017; Rimfeld, Ayorech, Dale, Kovas, & Plomin, 2016). Rimfeld and colleagues (2016) found nearly 50% of differences in whether children take an advanced school leaving exam (A-levels) versus leaving school two years earlier with a lower certificate are due to genetic factors, while the choice of what subjects to take during A-levels showed even more genetic influence (42-80%).

Using the same twin sample, we demonstrated that parents' educational choices influence their offspring's educational choices—for genetic reasons (Ayorech et al., 2017). For example, genetics contributed to 50% of the likelihood that children of parents without a university education would surpass the constraints typically associated with low SES and continue past compulsory education to sit A-level examinations.

Beyond twin studies, DNA-based methods including genome-wide polygenic scores (GPS), which aggregate genetic variants identified in genome-wide association (GWA) studies, can be used to investigate GE correlation in educational choice. GPS are a game changer for studies of GE correlation. Unlike the twin design, which estimates genetic influence by comparing phenotypic resemblance between siblings growing up in the same home, GPS estimates are based on unrelated individuals. For this reason, GPS can test for the influence of GE correlation on an environmental measure that is the same for all family members, such as the family's socioeconomic

status, which is not possible using classical twin methods. Previous studies have demonstrated the utility of GPS for understanding environmental selection developmentally (Belsky et al., 2016), as well as interactions between individuals' family SES and their educational attainment (Papageorge & Thom, 2017).

We have previously used a GPS for years of education (*EA2*; Okbay et al., 2016) to study genetic influence on intergenerational educational mobility (Ayorech et al., 2017). We found large mean *EA2* GPS differences between offspring of university-educated parents, who themselves took A-level examinations, and offspring of parents without a university education, who did not pursue A-levels ( $d= 0.64$ ).

Taken together, research suggests that genetic differences contribute to differences in decisions on educational trajectories after compulsory education, and that parents and their offspring are similar in their educational paths for genetic reasons. The present study extends this literature by using summary statistics from the largest years of education GWA (*EA3*; Lee et al., 2018), based on a sample of 1.1 million individuals, to calculate GPS to test for genetic influence on a range of educational choices. Specifically, we use GPS to predict going into full-time employment, apprenticeships or becoming NEET (Not in Education, Employment or Training), as compared to opting for continuing formal education. In addition, we test the extent to which GPS predict early adult education decisions beyond the predictive power of family SES and past academic achievement.

### **The present study**

Using 5,839 unrelated genotyped participants from the UK-representative Twins Early Development Study (TEDS), we assessed the extent to which the GPS from the latest GWA based on educational attainment (*EA3*; Lee et al., 2018) in an adult sample could predict educational trajectories at the beginning of emerging adulthood

at age 18. The *EA3* GPS was used to test for genetic differences between those who continue to university or take a gap-year, those who pursue full-time employment or an apprenticeship, and those who are NEET. In a second step, we controlled these analyses for an individual's family SES and past academic achievement. Finding that family background and past achievement attenuate the association between *EA3* and educational decision in emerging adulthood would suggest that the effect of *EA3* on educational trajectories is not direct, but that *EA3* acts on family- and individual-specific factors that in turn influence young adult's occupational and educational aspirations. Although our mediation analyses unpack one possible mechanism by which polygenic scores influence environmental selection, namely through an individual's cognitive ability and family background, there are many other unmeasured mechanisms that deserve future research effort.

The unique contribution of the present study includes the increased power to detect effects due to the considerably larger sample size compared to previous analyses (Belsky et al., 2016) and the age of the current sample, which is at the beginning of emerging adulthood, a unique period of identity exploration in the late teens and early twenties (Arnett, 2014). In addition, we used summary statistics based on the largest ( $N = 1.1$  million individuals) educational achievement GWA (Lee et al., 2018) that explains three times as much variance in years of education as the GWA (Rietveld et al., 2013) used in previous studies similar to ours. Finally, our mediation analyses capitalize on the longitudinal data available from our sample to unpack the mechanisms that drive genetic influence on educational opportunities.

## **Materials and Methods**

### **Sample**

The sample was drawn from the Twins Early Development Study (TEDS) and consisted of 5,839 genotyped unrelated individuals (one member of a twin pair) with



data on family SES, past academic achievement and trajectories after compulsory education. TEDS is a birth cohort of over 10,000 twin pairs born between 1994 and 1996. The TEDS sample is representative of the British population on ethnicity, family SES and parental occupation (Haworth, Davis, & Plomin, 2013) as is the genotyped subsample (Selzam et al., 2016). Ethical approval (PNM/09/10-104) was granted to the Twins Early Development Study by the Institute of Psychiatry, Psychology and Neuroscience ethics committee at King's College London and consent was obtained prior to data collection.

### **Measures**

The measures included data about the family's socioeconomic status, and about the twins' previous academic achievement and academic and non-academic choices at the end of compulsory education.

#### **Family socioeconomic status (SES).**

Converging evidence suggests that several factors together represent a family's socioeconomic status better than any single factor (White, 1982). As a result, we based family SES on maternal age at birth of eldest child, the mean score of maternal and paternal highest education level, as well as mothers' and fathers' occupation level according to the Standard Occupational Classification (Office for National Statistics, 2000) at first contact, when the twins were age 2.

#### **Previous academic achievement.**

Previous academic achievement was based on self-reported exam grades from the General Certificate of Secondary Education (GCSE), a standardized examination undertaken by all school students in the UK at the end of compulsory education at the age of 16. We have previously demonstrated the accuracy of self-reported exam results in our sample (Rimfeld, Kovas, Dale, & Plomin, 2015).

From the available achievement data we created a mean GCSE grade score representing the average grade achieved for the three compulsory core subjects, mathematics, English and science (mean of any science subject taken). Grades were coded to range from 4 (G; the minimum pass grade) to 11 (A\*; the highest possible grade).

### **Trajectories after compulsory education.**

A questionnaire, designed to assess destinations at the end of compulsory education, was sent to all TEDS families at the end of the academic school year when twins reached age 18. The full questionnaire was completed either by twins themselves or by their parents. Individuals indicated the main activity they had been doing for the past twelve months and the main activity they expect to do in the next twelve months; options included studying at university, working full-time, apprenticeship or training, taking a year gap in schooling before entering university (gap year), unemployment or full-time parent.

Accuracy of self-report measures of educational and occupational choices was maintained by requiring agreement across multiple questions. For example, those in the university group indicated that they intend to go to university within the next 12 months, provided the name of the university for which they held a place, the name of the course they are registered for and provided results for their A-level (General Certificate of Education Advanced Level) examination, which is a two-year course required in the UK for university entry. Those participants in the NEET group indicated they have been NEET for the previous 12 months, intend to be NEET for the next 12 months, and have not completed their A-levels.

In the UK a gap year is considered a constructive time out to travel and often work abroad, taken in the transition between life stages, before continuing to formal education at the end of the year. Although the gap year students in the present study

did not hold a place at a university when they indicated their intent, it is expected that most will continue studying.

## **Statistical Analyses**

### **Construction of EA3**

The EA3 GPS was calculated for each of the 5,839 unrelated genotyped individuals in the TEDS sample based on summary statistics from the largest years of education GWA (Lee et al., 2018). Following quality control (see Supplementary Methods S1 for details), genotypic data for 515,100 genotyped or imputed SNPs ( $info=1$ ) was available.

EA3 was constructed using a Bayesian approach, *LDpred* (Vilhjálmsson et al., 2015), which estimates causal effect sizes from GWA by assuming a prior for the genetic architecture and LD information from a reference panel. The constructed EA3 GPS is the weighted sum of trait-increasing alleles for each unrelated genotyped individual in the TEDS sample. We applied a causal fraction of 1, which assumes that all SNPs contribute to the development of the trait. The EAR GPS was adjusted for the first ten principal components of the genotype data, chip, batch and plate effects using the regression method. The standardized residuals were used for all subsequent analyses. Further details on creating polygenic scores using *LDpred* are available in Supplementary Methods S2.

### **Predicting education trajectories from DNA**

A multinomial logistic regression was conducted to test whether the *EA3* polygenic score could significantly predict educational and occupational trajectories following compulsory education. Multinomial logistic regression is an extension of binary logistic regression that allows for more than two unordered categories of the outcome variable. Using maximum likelihood estimation, multinomial logistic regression evaluates the odds of categorical membership depending on the value of an

exposure or predictor variable. For a nominal dependent variable with  $k$  categories, the model estimates  $k-1$  logit equations. Although all combinations of  $k$  groups are compared, a single group is selected as a reference category and prediction of membership in this group is compared to each of the other groups. Each model within the regression has its own intercept and coefficients as the predictor variable can affect each category differently (Aldrich & Nelson, 1984).

Here, we used the decision to go to university as the reference category and then tested whether we could predict, using the EA3 alone, who would pursue a university degree compared to taking a gap year, pursuing full-time employment or apprenticeship, or becoming NEET. The chi-square statistic  $\chi^2$  (Satorra & Bentler, 2001) and *Nagelkerke*  $r^2$  (Nagelkerke, 1991) were used to indicate, respectively, the relative fit of the model and the variance explained in trajectories by EA3. Odds ratios were calculated to test the likelihood of membership in each group as compared to going to university with a one standard deviation increase in EA3.

### **Comparative model**

Finally, to test the extent to which EA3 can predict educational trajectories above and beyond the predictors of family SES and past academic achievement, we performed an additional multinomial logistic regression analysis. Here family SES and GCSE scores were added to the previous model, and model coefficients and fit statistics were compared between the original model (EA3 predicting educational trajectories) and this comparative model.

### **Results**

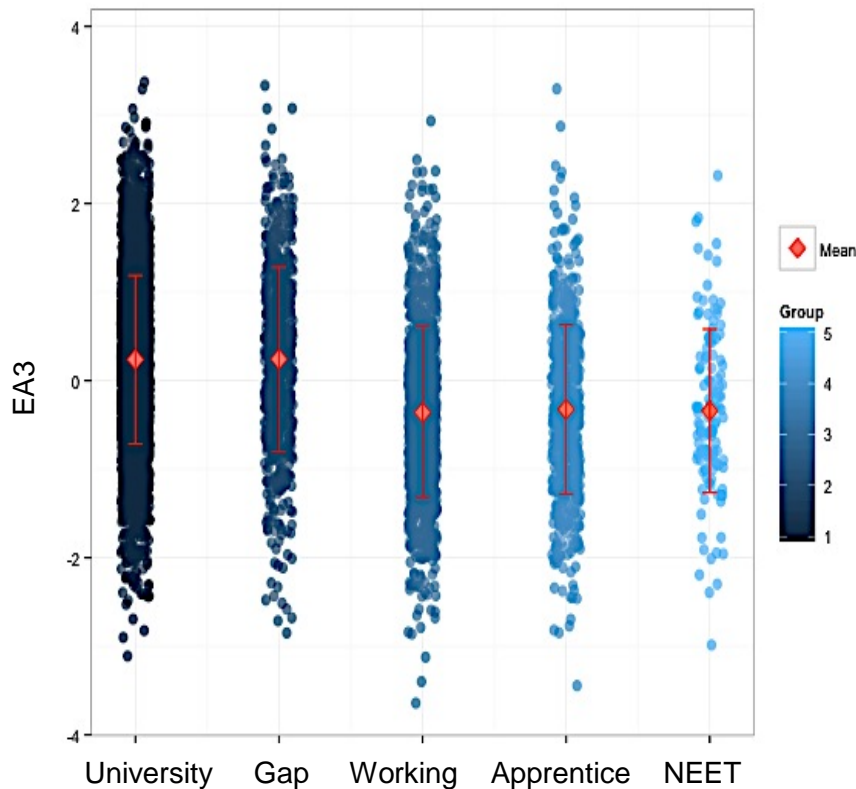
Data on decisions made at the end of compulsory education were available for 5,839 genotyped individuals of which 3,390 (58%) pursued a university education, 651 (11%) chose to take a gap year, 1,071 (18%) sought full-time employment, 598 (10%) began an apprenticeship, and 129 (2%) were NEET. Comparisons are

available with UK national averages at the time our TEDS data were collected ([https://www.gov.uk/government/uploads/system/uploads/attachment\\_data/file/532793/Main\\_text\\_SFR22\\_2016.pdf](https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/532793/Main_text_SFR22_2016.pdf)). TEDS participants were similar to UK national averages for pursuing university (58% vs 55%), for gap-year students (11% vs 7%), for full-time employment (18% vs 16%) and for apprenticeships (10% vs 9%). However, TEDS participants include notably fewer young adults considered NEET than the UK national average (2% vs 13%).

### **EA3 differences across groups.**

Mean EA3 differed across groups [ $F(4,5838) = 149.77, p < .001$ ]. Figure 1 shows the average standardized EA3 GPS for each of the educational and occupational choice groups (diamond) and the spread of EA3 scores within each group. The density of the columns refers to the size of the groups while the length of the columns indicates how far apart EA3 scores are for members within that group. Although some data points appear slightly off the central line for their respective group, the relative distance of a single dot from the central line should not be interpreted.

Young adults who pursued a university degree ( $M=0.33; SD=0.95$ ) or who took a gap year prior to going to university ( $M=0.31; SD=1.03$ ) had EA3 scores more than half a standard deviation greater than those who ended formal education and sought full-time employment ( $M=-0.35; SD=0.94$ ), an apprenticeship ( $M=-0.31; SD=0.89$ ) or were NEET ( $M=-0.29; SD=0.90$ ).



**Figure 1.** Distribution of the educational attainment (EA3) genome-wide polygenic score according to education status at the end of compulsory schooling. The diamond indicates the mean and the line represents the standard error.

### Predicting educational trajectories from DNA.

A multinomial logistic regression indicated that the EA3 GPS significantly predicted educational and occupational trajectories at the end of compulsory education [ $\chi^2(4, N=5,389) = 571.77, p < .001$ ] and explained 10% of the variance (*Nagelkerke*  $r^2 = 0.103$ ).

Odds ratios were calculated from the multinomial logistic regression (Table 1) and represent the likelihood of going to university compared to each of the other post-compulsory education options, given an individual's EA3. An odds ratio of one indicates there is no difference in the likelihood of being in a given post-compulsory choice group relative to being in the university group. An odds ratio greater than one

indicates an increased likelihood of membership in a given group relative to the university group. An odds ratio below one indicates a decreased likelihood of membership in a given group relative to the university group.

Compared to adolescents who went to university, a one-SD increase in EA3 was associated on average with a 51% decrease in the odds of being in full-time employment (OR=0.47;  $p < .01$ ), doing an apprenticeship (OR=0.49;  $p < .01$ ) or becoming NEET (OR=0.50;  $p < .01$ ). The odds of taking a gap year did not differ from the odds of going to university with a one-SD increase in EA3 (OR=0.98;  $p = 0.65$ ).

**Table 1.** Odds ratios for educational trajectories compared to entering university based on the genome-wide polygenic score (EA3)

	Intercept (SE)	B (SE)	Odds Ratio (CI 95%)
Gap year	-1.64 (0.05)**	-0.020 (0.05)	0.98 (0.90-1.07)
Working	-1.16 (0.04)**	-0.76 (0.04)**	0.47 (0.43-0.51)
Apprenticeship	-1.73 (0.05)**	-0.71 (0.05)**	0.49 (0.45-0.60)
NEET	-3.26 (0.09)**	-0.69 (0.10)**	0.50 (0.41-0.60)

Note: NEET= not in education, employment or training. All comparisons are made using university as the reference group;  $R^2 = 0.103$  (Nagelkerke). Model  $\chi^2(4) = 571.77$ ,  $p < .001$ ; \*\*= $p < .001$

### Comparative model

To test if EA3 added significantly to the prediction of educational decisions beyond previous academic achievement (GCSE) and family SES, we added family SES and children's GCSE achievement to the multinomial logistic regression (Table 2). Family SES and GCSE scores explained 39% of the variance (Nagelkerke  $r^2 = 0.39$ ) in decisions on educational trajectories at the end of compulsory education. When EA3 was added to the model, it contributed significantly to the prediction ( $p < .01$ ). All predictors together explained 40% of the variance (Nagelkerke  $r^2 = 0.40$ ), suggesting that 1% of the variance explained by EA3 GPS is unique to its association with

educational decisions. By comparison, 9% of the EA3 effect (see model 1) overlapped with the predictors of family SES and GCSE scores.

Compared to adolescents who went to university, a one-SD increase in EA3 was associated with a 15% decrease in the odds of seeking full-time employment (OR=0.85;  $p < .01$ ), and a 21% decrease in the odds of doing an apprenticeship (OR=0.79;  $p < .01$ ) when family SES and previous academic achievement were added to the model. Similar to model 1, the odds of taking a gap year (OR=1.00;  $p = 0.82$ ) compared to pursuing university did not differ when GCSE and family SES were accounted for. In contrast to model 1, the odds of becoming NEET (OR=1.05;  $p = 0.73$ ) no longer differed from the odds of going to university once family SES and GCSE achievement were accounted for.

**Table 2.** Odds ratios for educational trajectories compared to entering university based on individuals' genome-wide polygenic score, socioeconomic status and past academic achievement.

	Intercept (SE)	B (SE)	Odds Ratio (CI 95%)
<b>Gap year</b>			
EA3	-1.69 (0.06)**	0.01 (0.06)	1.01(0.91-1.12)
SES		0.22 (0.06)**	1.25 (1.11-1.40)
GCSE		-0.25 (0.06)**	0.78 (0.69-0.88)
<b>Working</b>			
EA3	-1.50 (0.06)**	-0.16(0.06)**	0.85 (0.76-0.95)
SES		-0.51 (0.06)**	0.60 (0.53-0.68)
GCSE		-2.04 (0.08)**	0.13 (0.11-0.15)
<b>Apprenticeship</b>			
EA3	-1.76 (0.06)**	-0.24 (0.06)**	0.79 (0.69-0.89)
SES		-0.37 (0.07)**	0.69 (0.60-0.79)
GCSE		-1.61 (0.08)**	0.20 (0.17-0.24)
<b>NEET</b>			
EA3	-4.31 (0.21)**	-0.05 (0.15)	1.06 (0.78-1.42)
SES		-0.66 (0.17)**	0.51 (0.36-0.73)
GCSE		-2.34 (0.19)**	0.10 (0.07-0.14)

*Note:* SES= socioeconomic status; EA3= genome-wide polygenic score; NEET= not in education, employment or training. All comparisons are made using university as the reference group;  $R^2 = 0.40$ (Nagelkerke). Model  $\chi^2(12) = 2061.87$ ,  $p < .001$ ; \*\*= $p < .001$

## Discussion



We report a polygenic score investigation of the factors contributing to early adult educational and occupational trajectories in a UK sample of over 4,000 young adults. Understanding the antecedents to decisions on educational and occupational paths made at the end of compulsory education is important for improving societal provisions for educational opportunities. Our findings also have general implications for developmental psychology, because they elucidate aspects of the interplay between genetic propensities and environment, called gene-environment correlation.

Here, we showed that genome-wide polygenic scores (EA3) from the largest GWA of educational attainment, based on 1.1 million individuals (Lee et al., 2018) explain 10% of the variance in educational trajectories at age 18, specifically, attending university, pursuing full-time employment or an apprenticeship, or becoming NEET.

Although 10% represents only a small fraction of the known SNP heritability of educational attainment (Krapohl & Plomin, 2015) this is still a meaningful amount of variance in the psychological literature, where typical effect sizes are even smaller (Open Science Collaboration, 2015). SNP-based heritability can be considered the upper bound limit for polygenic scores derived from GWA analysis. For this reason, any discrepancy between polygenic score prediction and SNP-based estimates reiterates that polygenic scores are a noisy measure of the true additive genetic factor based on measured SNPs. Considerably larger sample sizes are needed to detect the thousands of genetic variants each of which has a miniscule effect. As a result, the application of polygenic scores to shape educational policy is currently premature. Nonetheless, the unique predictive status of polygenic scores, which are not subject to reverse causality, make them attractive variables in the social sciences for investigating the complex contribution of genetic and environmental factors to human behavior (Belsky et al., 2016).

The present study adds to previous investigations of the genetic contributions to labour market outcomes (Belsky et al., 2016; Domingue, Belsky, Conley, Harris, & Boardman, 2015; Papageorge & Thom, 2017). Emerging adulthood provides a unique and novel opportunity to investigate GE correlation because it is the time when young adults gain more autonomy in their educational trajectories. Our mediation analyses allowed us to investigate the mechanisms driving these genetic influences on emerging adult trajectories.

Even when controlling for family SES and past academic achievement, EA3 still differentiated those children who pursue a university degree after compulsory education from those who end their formal studies to pursue full-time work or an apprenticeship. These results suggest that the choice of whether or not to continue with formal education at age 18 cannot be completely explained by genetic influences that are common to academic ability or family SES. It is possible that the remaining genetic influence on education decisions is due to other genetically influenced traits, like personality, psychopathology and behavior problems (Krapohl et al., 2014).

The substantial reduction in polygenic score prediction after accounting for SES and academic achievement (10% to 1%) is noteworthy. This reduction is not surprising given the substantial heritability of both SES and academic achievement and the genetic correlations between SES and academic achievement (Marioni et al., 2014) and between early and later academic achievement (Rimfeld et al., 2018). However, understanding the developmental processes by which these genetic correlations emerge remains for future research. That said, the odds associated with EA3 for becoming NEET no longer differed from the odds of going to university, once family SES and GCSE achievement were controlled for. These results suggest that this

EA3 effect can be explained entirely by differences in past achievement and family SES.

Finding that children's genetic variation correlates with their family SES may seem perplexing because a child's genotype cannot affect their family's SES. Because parents and their offspring are genetically related, children's EA3 can be considered an approximation of their parents' genotypes. For this reason, we would predict that the effect of EA3 on SES would be even stronger for the parents' own EA3, although these data are not available in TEDS.

It is important to reiterate that genetic effects happen both between and within families. It is not the case that children of highly educated and high SES parents automatically inherit the genetic variants for more years of education and also grow up in an educationally endowed environment. For example among pairs of siblings, the sibling with the higher polygenic score typically goes on to complete more years of schooling as compared to their lower-scoring co-sibling, despite being born to the same parents and raised in the same family home (Belsky et al., 2018; Domingue et al., 2015). Although the present study did not include sibling difference analyses, manuscripts are in preparation using TEDS data to investigate within-family differences in educational and health outcomes as a function of sibling differences in polygenic scores. These future publications represent an important direction for polygenic score analyses, which can elucidate the role of gene-environment interplay in the mechanisms of intergenerational transmission of success (Belsky et al., 2018; Lee et al., 2018).

We also stress that polygenic scores are probabilistic, not deterministic, with some individuals performing better or worse than would be expected given their genetic propensities (see spread of data in Figure 1). These deviations open the possibility

for an exciting application of polygenic scores for studies of resilience and positive genetics (Plomin et al., 2009).

Our polygenic score estimates are substantially lower when compared to SNP and twin heritability estimates of educational phenotypes. Polygenic scores capture only a fraction of the known heritability of academic achievement as measured by traditional pedigree-based methods like the twin design (Polderman et al., 2015). The gap between predictive validity of EA3 for educational phenotypes and other methods is characteristic of the missing heritability gap, which is likely to be due to rare variants, gene-gene interaction and gene-environment interaction, as well as measurement error. We chose not to correct for any of these factors because the predictive utility of polygenic scores depends on the polygenic score as measured with these shortcomings.

There are no necessary policy implications of finding genetic influence on educational trajectories. Genetic influences on behavioral phenotypes reflect what is, not what can be. Sweeping changes in educational policy can have profound changes on young people's educational trajectories despite genetic influence. For example, policies such as a required minimum number of years of education or tailored help for individuals with difficulties in learning can increase educational attainment in the entire population and even reduce differences among individuals. Genetic influences do not imply immutability.

Technological advancements and related cost reductions have brought us to the cusp of remarkable scientific opportunity for studying genetic influences. With these advancements come important ethical considerations. Although predictions about a person's educational attainment are limited with current GPS, a careful and considerate discussion of the future applications of GPS is urgent (Plomin, 2018).

Our findings support gene-environment correlation as a mechanistic pathway for genetic influence. Indeed, finding genetic influences on decisions at the end of compulsory education is indicative of active gene-environment correlation, in which individuals select and modify their environmental experiences in part based on their genetic propensities. Active GE correlation is key to understanding how genotypes use the environment to develop into phenotypes. As young adults gain autonomy in their life decisions and have the opportunity to select into environments that match their genetic propensities, genetic effects are expected to amplify. Our results indicate genome-wide polygenic scores offer a novel approach to elucidate this complex interplay and to study genetic influence on broad life trajectories.

### **Limitations**

Our results must be considered in light of the following limitations. First, academic decisions at the end of compulsory education were based on participants' intended trajectories at age 18. It is possible that some individuals may change their minds later in adulthood. For example, some young people may choose to return to education after full-time employment or raising a family while others may drop out of university before completing their degree. However, in the UK only 6% of 2013-14 university enrollees, who previously obtained their A-levels, failed to complete their degree, suggesting that the majority of young adults follow through their initial educational decisions (<https://www.hesa.ac.uk/data-and-analysis/performance-indicators/summary>).

Second, although we considered the influence of family social status on children's educational choices, there are other socio-structural variables that were not accounted for. It is possible, for example, that educational choices are influenced by

differences in how geographically or financially accessible university is for people, as well as the degree of implicit bias in university admission procedures.

Third, the results from this study provide a snapshot of the predictive potential of the current polygenic scores for years of education in a European sample. The TEDS sample is representative of the British population on ethnicity, family SES and parental occupation (Haworth et al., 2013) but the genotyped subsample is limited to the 93 percent of the sample with European ancestry. Our findings need to be tested for generalization to other countries with different educational policies, and to non-European samples.

## **Conclusions**

Applying DNA-based methods to understanding the developmental structure of real-world outcomes represents a tipping point for research on complex human behavior. Unlike classical family designs that require special relatives such as twins or adoptees, polygenic scores can be calculated on any group of unrelated individuals with genomic data available. Polygenic scores will become widely used in developmental psychology as indices of genetic propensities for many traits (Plomin et al., 2018). Their benefit lies in their unique predictive status: polygenic scores are not subject to reverse causality and can be measured from DNA taken during any time from infancy to old age, although the strength of prediction will depend on the developmental stage at which the investigated phenotype is measured (Plomin & von Stumm, 2018). In the current study, we used EA3 to elucidate genetic influences on decisions on educational trajectories in young adulthood. Our findings highlight the advantages of using genetically sensitive methods to investigate and understand individual differences in life course development.

## **ACKNOWLEDGEMENTS**

We gratefully acknowledge the ongoing contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant to RP from the UK Medical Research Council (MR/M021475/1 and previously G0901245), with additional support from the US National Institutes of Health (AG046938) and the European Commission (602768; 295366). RP is supported by a Medical Research Council Professorship award (G19/2). SvS is supported by a Jacobs Foundation Research Fellowship award (2017-2019).

### References

- Aldrich, J. H., & Nelson, F. D. (1984). *Linear probability, logit, and probit models* (Vol. 45): Sage.
- Arnett, J. J. (2014). *Emerging adulthood: The winding road from the late teens through the twenties*: Oxford University Press.
- Ayorech, Z., Krapohl, E., Plomin, R., & von Stumm, S. (2017). Genetic influence on intergenerational educational attainment. *Psychological science*, 1302-1310.
- Belsky, D. W., Domingue, B. W., Wedow, R., Arseneault, L., Boardman, J. D., Caspi, A., . . . Herd, P. (2018). Genetic analysis of social-class mobility in five longitudinal studies. *Proceedings of the National Academy of Sciences*, 201801238.
- Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B. W., Harrington, H. S., Houts, R., . . . Caspi, A. (2016). The Genetics of Success How Single-Nucleotide Polymorphisms Associated With Educational Attainment Relate to Life-Course Development. *Psychological science*, 957-972.
- Bradley, R. H., & Corwyn, R. F. (2002). Socioeconomic status and child development. *Annual review of psychology*, 53(1), 371-399.
- Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S. H., . . . Visscher, P. M. (2012). Multivariate genetic analyses of cognition and

- academic achievement from two population samples of 174,000 and 166,000 school children. *Behavior genetics*, 42(5), 699-710.
- Crowley, L., Jones, K., Cominetti, N., & Gulliford, J. (2013). International Lessons: Youth unemployment in the global context. *Lancaster University*.
- Domingue, B. W., Belsky, D. W., Conley, D., Harris, K. M., & Boardman, J. D. (2015). Polygenic influence on educational attainment. *AERA Open*, 1(3), 2332858415599972.
- Durbin, R. (2014). Efficient haplotype matching and storage using the positional Burrows–Wheeler transform (PBWT). *Bioinformatics*, 30(9), 1266-1272.
- Feng, Z., Ralston, K., Everington, D., & Dibben, C. (2017). Long term health effects of NEET experiences: evidence from a longitudinal analysis of young people in Scotland. *International Journal for Population Data Science*, 1(1).
- Fuchsberger, C., Abecasis, G. R., & Hinds, D. A. (2015). minimac2: faster genotype imputation. *Bioinformatics*, 31(5), 782-784.
- Goldman-Mellor, S., Caspi, A., Arseneault, L., Ajala, N., Ambler, A., Danese, A., . . . Williams, T. (2016). Committed to work but vulnerable: self-perceptions and mental health in NEET 18-year olds from a contemporary British cohort. *Journal of Child Psychology and Psychiatry*, 57(2), 196-203.
- Hart, B., & Risley, T. R. (1995). *Meaningful differences in the everyday experience of young American children*: Paul H Brookes Publishing.
- Haworth, C. M., Davis, O. S., & Plomin, R. (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res Hum Genet*, 16(1), 117-125. doi:10.1017/thg.2012.91
- Kelly, I. R., Dave, D. M., Sindelar, J. L., & Gallo, W. T. (2014). The impact of early occupational choice on health behaviors. *Review of Economics of the Household*, 12(4), 737-770.



- Knopik, V. S., Neiderhiser, J., DeFries, J. C., & Plomin, R. (2017). *Behavioral genetics* (7th ed.). New York: New York: Worth.
- Krapohl, E., & Plomin, R. (2015). Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol Psychiatry*. doi:10.1038/mp.2015.2
- 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.
- Lee, J. J., Wedow, R., Okbay, A., Kong, E., Maghziyan, O., Zacher, M., . . . Linnér, R. K. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nature genetics*, 1112–1121.
- Leppel, K., Williams, M. L., & Waldauer, C. (2001). The impact of parental occupation and socioeconomic status on choice of college major. *Journal of Family and Economic issues*, *22*(4), 373-394.
- Li, H. (2011). A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*, *27*(21), 2987-2993.
- Loh, P.-R., Danecek, P., Palamara, P. F., Fuchsberger, C., Reshef, Y. A., Finucane, H. K., . . . Abecasis, G. R. (2016). Reference-based phasing using the Haplotype Reference Consortium panel. *Nature genetics*, *48*(11), 1443.
- Macmillan, L., Tyler, C., & Vignoles, A. (2015). Who gets the top jobs? The role of family background and networks in recent graduates' access to high-status professions. *Journal of Social Policy*, *44*(3), 487-515.
- Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., . . . Hocking, L. J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence*, *44*, 26-32.

- McCarthy, S., Das, S., Kretzschmar, W., Durbin, R., Abecasis, G., & Marchini, J. (2016). A reference panel of 64,976 haplotypes for genotype imputation. *bioRxiv*.
- Nagelkerke, N. (1991). A note on a general definition of the coefficient of determination. *Biometrika*, *78*, 691-692.
- Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., . . . Oskarsson, S. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*, 539-542.
- Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716.
- Papageorge, N., & Thom, K. (2017). Genes, education, and labor market outcomes: evidence from the health and retirement study. (Social Science Research Network, Rochester, NY)
- Parker, P. D., Schoon, I., Tsai, Y.-M., Nagy, G., Trautwein, U., & Eccles, J. S. (2012). Achievement, agency, gender, and socioeconomic background as predictors of postschool choices: A multicontext study. *Developmental psychology*, *48*(6), 1629
- Patterson, N., Price, A. L., & Reich, D. (2006). Population structure and eigenanalysis. *PLoS genetics*, *2*(12), e190
- Plomin, R. (2018). *Blueprint: How DNA Makes Us Who We Are*. London: Allen Lane (Penguin Press)
- Plomin, R., DeFries, J., Knopik, V., & Neiderhiser, J. (2016). Top 10 replicated findings from behavioral genetics. *Perspectives on Psychological Science*, *11*(1), 3-23
- Plomin, R., Haworth, C. M., & Davis, O. S. (2009). Common disorders are quantitative traits. *Nature reviews genetics*, *10*(12), 872-878.
- Plomin, R., & von Stumm, S. (2018). The new genetics of intelligence. *Nature Reviews Genetics*, 148-159.

- Polderman, T. J., Benyamin, B., de Leeuw, C. A., Sullivan, P. F., van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nat Genet*. doi:10.1038/ng.3285
- Price, A. L., Patterson, N. J., Plenge, R. M., Weinblatt, M. E., Shadick, N. A., & Reich, D. (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, *38*(8), 904.
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., . . . Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American Journal of Human Genetics*, *81*(3), 559-575.
- Rietveld, C. A., Medland, S. E., Derringer, J., Yang, J., Esko, T., Martin, N. W., . . . Koellinger, P. D. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Science*, *340*(6139), 1467-1471.
- Rimfeld, K., Ayorech, Z., Dale, P., Kovas, Y., & Plomin, R. (2016). Genetics affects choice of academic subjects as well as achievement. *Scientific Reports*, *6*.
- Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2015). Pleiotropy across academic subjects at the end of compulsory education. *Scientific reports*, *5*, 11713.
- Rimfeld, K., Malanchini, M., Krapohl, E., Hannigan, L., Dale, P., & Plomin, R. (2018). The stability of educational achievement across school years is largely explained by genetic factors. *NPJ Science of Learning*, *3*(1)
- Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika*, *66*(4), 507-514.
- Selzam, S., Krapohl, E., von Stumm, S., O'reilly, P., Rimfeld, K., Kovas, Y., . . . Plomin, R. (2017). Predicting educational achievement from DNA. *Molecular psychiatry*, *22*(2), 267-272.

- Team, R. C. (2015). R: A language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2014.
- Trzaskowski, M., Harlaar, N., Arden, R., Krapohl, E., Rimfeld, K., McMillan, A., . . . Plomin, R. (2014). Genetic influence on family socioeconomic status and children's intelligence. *Intelligence*, *42*, 83-88.
- Van de Werfhorst, H. (2001). Field of study and social inequality. *Four Types of Educational Resources in the Process of Stratification in the Netherlands. ICS-dissertation, Nijmegen*.
- Vattikuti, S., Guo, J., & Chow, C. C. (2012). Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits. *PLoS genetics*, *8*(3), e1002637.
- Vilhjálmsón, B. J., Yang, J., Finucane, H. K., Gusev, A., Lindström, S., Ripke, S., . . . Do, R. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *The american journal of human genetics*, *97*(4), 576-592.
- Werfhorst, H. G., Sullivan, A., & Cheung, S. Y. (2003). Social class, ability and choice of subject in secondary and tertiary education in Britain. *British Educational Research Journal*, *29*(1), 41-62.
- White, I. R., Blane, D., Morris, J., & Mourouga, P. (1999). Educational attainment, deprivation-affluence and self reported health in Britain: a cross sectional study. *Journal of Epidemiology & Community Health*, *53*(9), 535-541.
- White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological bulletin*, *91*(3), 461.