Published in final edited form as:

J Pers Soc Psychol. 2019 December ; 117(6): 1145–1163. doi:10.1037/pspp0000241.

Polygenic score for educational attainment captures DNA variants shared between personality traits and educational achievement

Emily Smith-Woolley[1,*], **Saskia Selzam**[1,*], **Robert Plomin**[1]

[1]King's College London, MRC Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, London. SE5 8AF, UK.

Abstract

Genome-wide polygenic scores (GPS) can be used to predict individual genetic risk and resilience. For example, a GPS for years of education (*EduYears*) explains substantial variance in cognitive traits such as general cognitive ability and educational achievement. Personality traits are also known to contribute to individual differences in educational achievement. However, the association between the *EduYears* GPS and personality traits remains largely unexplored. Here, we test the relation between GPS for *EduYears*, neuroticism and wellbeing, and six personality and motivation domains: Academic motivation, Extraversion, Openness, Conscientiousness, Neuroticism and Agreeableness. The sample was drawn from a UK-representative sample of up to 8,322 individuals assessed at age 16. We find that *EduYears* GPS was positively associated with Openness, Conscientiousness, Agreeableness and Academic motivation, predicting between 0.6% and 3% of the variance. In addition, we find that *EduYears* GPS explains between 8% and 16% of the association between personality domains and educational achievement at the end of compulsory education. In contrast, both the neuroticism and wellbeing GPS significantly accounted for between 0.3% and 0.7% of the variance in a subset of personality domains and did not significantly account for any of the covariance between the personality domains and achievement, with the exception of the neuroticism GPS explaining 5% of the covariance between Neuroticism and achievement. These results demonstrate that the genetic effects of educational attainment relate to personality traits, highlighting the multifaceted nature of *EduYears* GPS.

Keywords

polygenic score; personality; motivation; academic achievement

Introduction

Education is one of society's most expensive intervention programmes. Among the member countries of the Organisation for Economic Cooperation and Development (OECD), education accounts for between 6–15% of annual gross domestic product (OECD, 2017) and the average young person in these countries will stay in education until the age of 22

Corresponding authors: Emily Smith-Woolley: emily.smith-woolley@kcl.ac.uk, Saskia Selzam: saskia.selzam@kcl.ac.uk.
*Joint first authors

(OECD, 2007). Given its societal value, great importance is placed on succeeding in education, both in terms of educational attainment (education level) and education achievement (education grade).

For a century, psychologists have attempted to unravel the major predictors of individual differences in educational success. Early work showed that ‘cognitive capacity’ played a substantial role in education performance (Binet & Simon, 1916), a term that now many refer to as general cognitive ability or ‘g’. However, it did not tell the whole story. Around the same time, Webb (1915) proposed that in addition to g, academic performance was also influenced by a ‘w’ or ‘will’ factor, representing drive or motivation (Webb, 1915). This led the way for ‘psychological’ explanations of educational success. Most now accept a more complex model of academic performance that comprises both what a person *can do* (general cognitive ability) and *how* a person will do it (personality, motivation and other psychosocial influences).

One important factor influencing both the *can* and the *how*, is genetics. Inherited DNA differences play an important role in explaining individual differences in personality traits, general cognitive ability and educational outcomes. Decades of research using twin studies have shown substantial heritability for personality traits, general cognitive ability and educational outcomes (Polderman et al., 2015). To estimate genetic and environmental influences based on twin studies, the relative similarities between identical (monozygotic; ‘MZ’) twins, who share 100% of their inherited DNA, are compared to the relative similarities between fraternal (dizygotic; ‘DZ’) twins, who share on average 50% of their inherited DNA differences (Knopik, Neiderhiser, DeFries, & Plomin, 2017). Because both sets of twins grow up in equally similar environments (Derks, Dolan, & Boomsma, 2006; Kendler, Neale, Kessler, Heath, & Eaves, 1993), the influence of genetics and the environmental on traits can be unpacked: if MZ twins correlate higher for a trait than DZ twins, then genetic influence is inferred. However, twin studies can only tell us about the relative genetic influence on differences in traits within a population, rather than the influence of measured DNA differences on traits. In the current study, we use a more recent, genetically sensitive method – genome-wide polygenic scoring – to predict a broad range of personality and motivation traits directly from DNA. Furthermore, we estimate the role of measured DNA in the association between these personality traits and academic achievement at age 16.

General cognitive ability and educational performance

Educational achievement represents a cumulative process of acquiring many skills, gradually over time. Although it is influenced by a multitude of different factors, one of the most powerful and parsimonious predictors is general cognitive ability. General cognitive ability captures the communalities within a diverse set of cognitive measures, such as memory, verbal-reasoning and non-verbal reasoning (Plomin & Deary, 2015). It is highly correlated with academic achievement at age 9 ($r = 0.45$) (Spinath, Spinath, Harlaar, & Plomin, 2006), school performance at the end of compulsory education at age 16 ($r = 0.81$) (Deary, Strand, Smith, & Fernandes, 2007) and also later with university achievement ($r = .48$) (Frey & Detterman, 2004).

In addition to educational achievement, general cognitive ability is also strongly associated with years spent in full-time education (Deary & Johnson, 2010; Jencks, 1979; Ritchie & Tucker-Drob, 2018). However, although general cognitive ability explains more than half of the variance in academic outcomes (Deary et al., 2007), it still leaves a substantial portion of the variance unexplained. Therefore, it is important to consider other explanatory factors influencing educational performance.

Personality and educational performance

The most widely researched personality correlates of educational performance are dimensions of the Five-Factor Model (FFM) (McCrae & Costa, 1987). The FFM comprises Conscientiousness (dependability and drive to achieve), Extraversion (sociability and activity), Openness to Experience (curiosity and broadmindedness), Agreeableness (compassion and kindness) and Neuroticism (stress and anxiety). These broad domains have been linked both positively (conscientiousness, openness and agreeableness) and negatively (neuroticism and extraversion) to academic performance (Busato, Prins, Elshout, & Hamaker, 2000; Chamorro-Premuzic & Furnham, 2003; Conard, 2006; De Raad & Schouwenburg, 1996; O'Connor & Paunonen, 2007; Petrides, Chamorro-Premuzic, Frederickson, & Furnham, 2005; Poropat, 2009; Richardson, Abraham, & Bond, 2012). In addition, their underlying, specific facets (most notably dutifulness, achievement-striving and anxiety) have also been associated with differences in academic performance (Chamorro-Premuzic & Furnham, 2003).

Many studies have explored the reasons for observed associations between FFM dimensions and academic performance – both in terms of attainment and achievement. Conscientiousness is comparable to the 'w' factor described by Webb (1915) and has been linked to academic effort (Trautwein, Lüdtke, Roberts, Schnyder, & Niggli, 2009) through time spent on homework (Trautwein & Lüdtke, 2007) and time use efficiency (Kelly & Johnson, 2005). It has been shown to predict academic performance at high-school (Heaven & Ciarrochi, 2008; Laidra, Pullmann, & Allik, 2007), undergraduate (Chamorro-Premuzic & Furnham, 2003; Conard, 2006; Wagerman & Funder, 2007) and even at postgraduate level (Hirschberg & Itkin, 1978). Agreeableness and Openness have also been linked to academic performance: Agreeableness through following teacher instructions and learning style (Busato, Prins, Elshout, & Hamaker, 1998) and Openness through critical thinking (Bidjerano & Dai, 2007) and intelligence (Holland, Dollinger, Holland, & Macdonald, 1995; McCrae & Costa, 1997). Like Conscientiousness, Openness is also related to success in school and at university, showing positive correlations with undergraduate and postgraduate examination scores (Geramian, Mashayekhi, & Ninggal, 2012; Laidra et al., 2007). In contrast, Neuroticism and Extraversion have been negatively linked to academic achievement; Extraversion through distractibility, sociability and problems regulating effort devoted to academic tasks (Bidjerano & Dai, 2007) and Neuroticism through stress linked with exams and poor impulse control (Zeidner & Matthews, 2000).

Because there are intercorrelations between personality traits, general cognitive ability and academic achievement, an important question to consider is how these personality traits link to achievement over and above cognitive ability. Conscientiousness has consistently been linked to academic achievement over and above general cognitive ability. For example it was demonstrated (Poropat, 2009) that Conscientiousness was largely independent of intelligence and that when academic achievement at secondary school was accounted for, Conscientiousness continued to predict achievement at university. This is in line with another study also showing that once prior achievement on SATs were accounted for, Conscientiousness incrementally predicted later achievement (Conard, 2006). However, there have been few studies looking at personality and general cognitive ability concurrently at secondary school level.

Motivation and educational performance

In addition to personality dimensions, other explanations of academic performance have been put forward. In a systematic review of psychological traits, Richardson and colleagues (Richardson et al., 2012) suggest five 'non-intellective' domains influencing educational success: 1) personality traits 2) motivational factors 3) self-regulatory strategies 4) student's approaches to learning and 5) psychosocial influences. Although the authors note that these domains are 'conceptually overlapping', they argue that it is important to consider a wide variety of 'non-intellective' factors when predicting academic performance.

One of these factors, which has consistently been linked to academic performance, is motivation. Although aspects of motivation correlate moderately with the FFM dimensions, for example extraversion (positively) and neuroticism (negatively) (Komarraju & Karau, 2005), many argue that elements of motivation, such as self-efficacy beliefs, may influence achievement over and above these dimensions (Caprara, Vecchione, Alessandri, Gerbino, & Barbaranelli, 2011).

Self-efficacy beliefs are an individual's beliefs about their capabilities to produce effects (Bandura, 1997). Self-efficacy and related traits, such as self-perceived ability, engagement and academic self-concept are important constructs which help to explain students' learning and progress (Multon, Brown, & Lent, 1991; Schunk, 1989). In one study specifically looking at math self-efficacy and self-concept (Parker, Marsh, Ciarrochi, Marshall, & Abduljabbar, 2014), moderate correlations with achievement in math and science were found ($r = .17 – .58$), and math self-efficacy was also a significant predictor of university entry. Similarly to personality dimensions, self-efficacy beliefs have also been shown to predict academic achievement over and above general cognitive ability; self-perceptions of ability explained an extra 8% of the variance in math achievement and 9% in English achievement at age 9 after accounting for general cognitive ability (Spinath et al., 2006).

Heritability of personality traits

The heritability of personality traits has been well established. Estimates of the genetic influence on variance in the Big Five personality traits range from 40–60% (Bouchard Jr & McGue, 2003; Jang, Livesley, & Vemon, 1996; Polderman et al., 2015). In line with twin study heritability estimates of personality traits, one twin study using the same sample as in the present study, found that at age 16, heritability ranged from 35% for wellbeing to 40% for self-efficacy and up to 46% for aspects of personality (Krapohl et al., 2014). Furthermore, in the same study, they found that inherited DNA differences explained a large portion of the observed correlation between personality and general cognitive ability and academic achievement. Consistent with this, a study using twins from the US also found that genetically influenced variation accounted for the associations between personality traits and both academic achievement and verbal knowledge (Tucker-Drob, Briley, Engelhardt, Mann, & Harden, 2016). Furthermore, they found that part of these genetically-mediated associations were shared with general cognitive ability. This suggests that some of the genetic factors driving variation in personality and general cognitive ability are also explaining variance in achievement. This concept is known as 'pleiotropy' – the finding that single genetic variants affect multiple traits (Solovieff, Cotsapas, Lee, Purcell, & Smoller, 2013).

Although twin studies are not able to point to specific genetic variants that are responsible for covariation between traits, the extent to which the phenotypic correlation between traits can be explained by genetics (the genetic correlation) is an index of pleiotropy. Why might genetic variants associated with personality and general cognitive ability also be related to achievement? Doing well in exams requires more than just intelligence; it requires motivation, concentration, diligence, good mental health, as well as many other factors. Furthermore, these heritable traits might also lead individuals to choose certain environments for themselves, for example, individuals high on Conscientiousness may choose to attend optional revision classes and complete homework on time. These decisions may in turn lead to better educational outcomes, such as higher grades. This illustrates a concept known as gene-environment correlation (rGE) (Knopik et al., 2017; Plomin, DeFries, & Loehlin, 1977). rGE is the idea that an individual's genetically influenced behaviour may elicit specific reactions from others (evocative rGE), or lead individuals to choose experiences and environments that correlate with their genotype (active rGE). A third type of rGE is passive rGE, whereby children are exposed to family environments that are partly created by, and therefore correlated with, their parents' genetic propensities. If passive rGE is at play, these 'inherited' environments reinforce children's own genetic propensities, driving development, or co-development of traits. Indeed, recent studies have shown that passive rGE is a likely mechanism in the development of educational achievement (Kong et al., 2018; Lee et al., 2018). Presented in this context, finding that much of the correlation between personality and educational achievement is explained by genetic factors, may therefore be partly reflecting a developmental pattern induced by rGE.

Using DNA to predict personality traits

In addition to family studies, such as twin designs, DNA-based methods have also shed light on genetic influence on personality traits. Genome-wide association (GWA) studies test associations between millions of known DNA variants, called single nucleotide polymorphisms (SNPs), and phenotypic traits in large samples comprising thousands of individuals. GWA studies have shown that effect sizes between individual SNPs and complex traits are usually very small, with single SNPs generally explaining less than 0.1% of the variance each (Gratten, Wray, Keller, & Visscher, 2014). However, because it is assumed that most of these genetic effects are additive, more phenotypic variance can be explained when considering these SNPs jointly (Purcell et al., 2009). By summing up the number of trait-increasing alleles, which are weighted by the GWA SNP effect sizes across thousands of SNPs, it is possible to generate a genetic score for each individual in an independent sample. These genetic scores, referred to as genome-wide polygenic scores (GPS), allow DNA-based prediction for any complex trait.

One of the largest published GWA studies for a behavioural trait is years of education (*EduYears*) (Lee et al., 2018; Okbay, Baselmans, et al., 2016; Rietveld et al., 2013). This study, which had a sample size of 1.1 million adults, tested associations between SNPs and total years in education. It is possible to use the results from this study, indicating which SNPs are associated with years of education and how large the association is, to create GPS in an independent, genotyped sample. Genome-wide polygenic scores for years of education have been shown to explain 11–13% of the variance in the target trait years of education (Lee et al., 2018), 7–10% in cognitive performance (Lee et al., 2018), up to 5% in reading ability (Selzam, Dale, et al., 2017) and up to 15% in educational achievement at 16 (Allegrini et al., 2018).

Although 'cognitive' GPS such as years of education and intelligence appear to be explaining variance in their target traits, and related traits such as achievement (Plomin & von Stumm, 2018), personality GPS have been less predictive. For example, a GPS for wellbeing explains 0.9% of the variance in wellbeing and 0.7% in neuroticism (Okbay, Baselmans, et al., 2016). In the current study, we sought to investigate whether a polygenic score for years of education could predict variance in a range of personality and motivation domains, how this prediction compared to personality polygenic score prediction, and whether personality polygenic scores relate to educational achievement.

Why might a genome-wide polygenic score for education link to personality? Similarly to achievement, educational attainment (years in education), is influenced by a multitude of heritable traits in both the cognitive ability and personality domains (Fredricks, Blumenfeld, & Paris, 2004). So far, only one study (Mõttus, Realo, Vainik, Allik, & Esko, 2017) has related *EduYears* GPS to personality traits. This study investigated the link between *EduYears* GPS and the Big Five personality traits in an Estonian sample of ~3,000 adults of a wide age range. *EduYears* GPS predicted 0.5% of the variance in Neuroticism and 1.2% in Openness to experience, suggesting that the polygenic score for educational attainment tags genetic variants that also relate to personality domains. However so far, no study has investigated links to other personality traits aspects, such as the underlying, more specific facets of personality (e.g. wellbeing or anxiety), as well as motivation traits such as self-efficacy beliefs.

The present study

Given the genetic links between personality traits and educational achievement, the current study sought to explore these associations further by testing the extent to which *EduYears* GPS correlated with personality and motivation domains, as well as their sub-traits. In addition, using a neuroticism GPS and wellbeing GPS, we contrasted the association between these personality GPS and educational achievement to *EduYears* GPS. We also tested whether associations remained after accounting for general cognitive ability. Finally, given previous quantitative genetics findings, we tested the extent to which the *EduYears*, neuroticism and wellbeing GPS explain the covariance between a range of personality traits and educational achievement at age 16.

Methods

Ethics

Ethical approval for this study was received from King's College London Ethics Committee, Reference Number: PNM/09/10–104.

Sample

The sampling frame for the present study was the Twins Early Development Study (TEDS) (Haworth, Davis, & Plomin, 2013). TEDS includes 16,000 twin pairs born between 1994 and 1996 and followed from birth to the present day. Although there has been some attrition, approximately 10,000 twin pairs are still enrolled in the study, providing behavioral, cognitive and psychological data. The TEDS sample is representative of families with children in England and Wales (Haworth et al., 2013). The current study uses a genotyped subsample of TEDS which comprises 10,346 Caucasian individuals, including 7,026 unrelated individuals (i.e., one member of a twin pair), and 3,320 DZ co-twins. Written informed consent was obtained from parents before data collection.

Genotyping

Two genotyping platforms were used to genotype TEDS individuals because these genotyping efforts were separated by 5 years. AffymetrixGeneChip 6.0 SNP arrays were used to genotype 3,747 individuals at Affymetrix, Santa Clara (California, USA) based on buccal cell DNA samples. Genotypes were generated at the Wellcome Trust Sanger Institute (Hinxton, UK) as part of the Wellcome Trust Case Control Consortium 2 (https://www.wtccc.org.uk/ccc2/). Additionally, 8,122 individuals, including 3,607 dizygotic twin pairs, were genotyped on HumanOmniExpressExome-8v1.2 arrays at the Molecular Genetics Laboratories of the Medical Research Council Social, Genetic Developmental Psychiatry Centre, based on DNA that was extracted from saliva samples. A total sample of 10,346 samples (including 3,320 dizygotic twin pairs and 7,026 unrelated individuals), with 7,289 individuals and 559,772 SNPs genotyped on Illumina and 3,057 individuals and 635,269 SNPs genotyped on Affymetrix remained after quality control. Both samples were imputed separately to the Haplotype Reference Consortium (release 1.1) reference genotypes using the Sanger Imputation Server (McCarthy et al., 2016), before merging genotype data obtained from both platforms. Following post-imputation quality control and platform harmonisation, 7,363,646 SNPs were retained for the analyses (for full details, see Selzam et al., 2018).

To calculate genomic principal components to account for population stratification, we performed principal component analysis on a subset of 39,353 common (MAF > 5%), perfectly imputed (info = 1) autosomal SNPs, after stringent pruning to remove markers in linkage disequilibrium ($r^2 > 0.1$) and exclusion of high linkage disequilibrium genomic regions.

Measures

GCSE.—The General Certificate of Secondary Education (GCSE) is a standardized UK-based examination at the end of compulsory education at age 16. Students are required to take three core subjects: English, mathematics and science. For 7,325 genotyped individuals, these results were obtained from questionnaires sent via mail, in addition to telephone interviews with twins and their parents. We also obtained subject grades for an additional 1,227 genotyped participants that had missing self-reported data from the National Pupil database (NPD: https://www.gov.uk/government/collections/national-pupil-database. Written consent was given before accessing this data. The total sample included 8,552 genotyped individuals (M = 16.30 years; SD = 0.29 years), including 2,799 DZ twin pairs. Subjects were graded from 4 (G; the minimum pass grade) to 11 (A*; the best possible grade). We used a mean of the three z-standardized compulsory subjects because other subjects are taken by only subsamples of the students. English, mathematics and science performance correlated highly with each other (r = 0.70 – 0.81). Furthermore, self-reported GCSE grades of TEDS participants show high accuracy, correlating 0.98 English and 0.99 for mathematics grades with data obtained for a subsample from the NPD.

General cognitive ability.—Individuals were measured on multiple cognitive tests including verbal and non-verbal abilities at age 7 (M = 7.12, SD = 0.24, N = 5,612), 12 (M = 11.44, SD = 0.65, N = 5,284) and 16 (M = 16.47, SD = 0.278, N = 2,840). Age specific mean score composites were derived from four tests at age 7: Conceptual Grouping (McCarthy, 1972), Similarities, Vocabulary and Picture Completion (Wechsler, Golombok, & Rust, 1992); three tests at age 12: Raven's Progressive Matrices (Raven & Raven, 1998), General Knowledge (Kaplan, Fein, Kramer, Delis, & Morris, 1999) and Picture Completion (Wechsler et al., 1992) and two tests at age 16: Raven's Progressive Matrices (Raven & Raven, 1998) and Mill Hill Vocabulary test (Raven, Raven, & Court, 1989). A general cognitive ability composite was created by taking the arithmetic mean of the z-standardized cognitive ability composites, requiring data to be present for at least two ages (N = 3,939; including 1,261 DZ twin pairs).

Personality and motivation measures.—We included 28 self-report measures collected at age 16 (M = 16.48 years; SD = 0.27 years) via self-reports using paper booklet (b) and web-based (w) assessment:

(w) PISA maths self-efficacy – 8 items—(PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org): This scale was selected from the PISA 2000, 2003 and 2006 student questionnaires, comprising 8 items asking participants to rate how confident they feel about having to do mathematical tasks on a 4-point scale from 'Not at all confident' to 'Very confident'. For example, solving an equation like: $2(x + 3) = (x + 3)(x - 3)$. The total score was created by taking the mean of the 8 items, requiring at least 4 to be present. The scale has an average reliability of 0.83 across OECD countries (Ray & Margaret, 2003). We find similar reliability estimates in the present sample (α = 0.90).

(w) PISA math interest – 3 items—(PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org): This scale was selected from the PISA 2000, 2003 and

2006 student questionnaires. The scale asked participants to rate how interested they were in mathematics on a 4-point scale from ‘Strongly disagree’ to ‘Strongly agree’. For example rating statements such as: I look forward to my mathematics lessons. The total score was created by taking the mean of the 3 items, requiring at least 2 to be present. Reliability for this measure. The mean reliability across OECD countries is 0.75 for this measure (Ray & Margaret, 2003). We find a slightly better reliability estimate in the present study than that previously reported ($\alpha = 0.93$)

(w) PISA time spent on math – 3 items—(PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org): This scale was selected from the PISA 2000, 2003 and 2006 student questionnaires. The scale asked participants to rate how much time they typically spent per week studying mathematics from ‘No time’ to ‘6 hours or more’. For example ‘Regular lessons in mathematics at my school’. The total score was created by taking the mean of the 3 items, requiring at least 2 to be present. The mean reliability across OECD countries is 0.76 for this measure (Ray & Margaret, 2003). We find slightly lower reliability estimates ($\alpha = 0.53$) in the current sample.

(w) Academic self-concept – 11 items—(Burden, 1998). This scale aims to assess children’s perceptions of themselves as learners and problem solvers by asking children to rate themselves on a 5 point scale from ‘Very much like me’ to ‘Not at all like me’ to statements such as ‘I know the meaning of lots of words’. The total score was created by taking the mean of the 11 items, requiring at least 5 to be present. The mean reliability across OECD countries is 0.79 for this measure (Ray & Margaret, 2003). We find similar reliability estimates ($\alpha = 0.84$) in the current sample.

(w) Total attitude towards key subjects – 3 items—(PISA, OECD Programme for International Student Assessment; www.pisa.oecd.org): This scale was selected from the PISA 2000, 2003 and 2006 student questionnaires. Participants were asked to answer the question ‘In general, how important do you think it is for you to do well in the subjects below?’ on a 4 point scale from ‘Not at all important’ to ‘Very important’ for the subjects English, mathematics and science. The total score was created by taking the mean of the 3 items, requiring at least 2 to be present. The mean reliability across OECD countries is 0.79 for this measure (Ray & Margaret, 2003). We find lower reliability in our sample ($\alpha = 0.45$).

(w) School engagement – 19 items—(Appleton, Christenson, Kim, & Reschly, 2006): This scale aims to assess children’s engagement with the school environment, including teacher-student relations, control and relevance of school work, peer support and family support for learning. Participants were required to answer questions such as ‘I enjoy talking to the teachers at my school’ and ‘Students at my school respect what I have to say’ on a 4 point scale from ‘Strongly disagree’ to ‘Strongly agree’. The total score was created by taking the mean of the 19 items, requiring at least 10 to be present. The reliability of factors in this measure range from 0.76 to 0.88 (Appleton, Christenson, Kim, & Reschly, 2006). We find high reliability ($\alpha = 0.99$) in the current sample.

(w) Big five personality (Extraversion, Openness, Agreeableness, Conscientiousness, neuroticism) – 30 items—(Mullins-Sweatt, Jamerson, Samuel, Olson, & Widiger, 2006): We used the subscales from this measure, tapping into Extraversion, Openness, Agreeableness, Conscientiousness and Neuroticism.

***Extraversion* – 6 items:** participants were asked to rate were they were on a scale that varied for each item. For example for the trait 'Activity' they had to rate were they were on a scale from 'vigorous, energetic, active' to 'passive, lethargic'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension has been estimated to be between 0.60 – 0.76. In the current sample, the reliability is within the range of previous studies (α = 0.68).

***Openness* – 6 items:** participants were asked to rate were they were on a scale that varied for each item. For example for the trait 'Fantasy' they had to rate were they were on a scale from 'dreamer, unrealistic, imaginative' to 'practical, concrete'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension ranged between 0.51 – 0.69. In the current sample, the reliability is within the range of previous studies (α = 0.61).

***Agreeableness* – 6 items:** For example for the trait 'Compliance' they had to rate were they were on a scale from 'docile, cooperative' to 'oppositional, combative, aggressive'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension ranged between 0.56 – 0.72. In the current sample, the reliability is within the range of previous studies (α = 0.65).

***Conscientiousness* – 6 items:** participants were asked to rate were they were on a scale that varied for each item. For example for the trait 'Self-discipline' they had to rate were they were on a scale from 'dogged, devoted' to 'hedonistic, negligent'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension ranged between 0.73 – 0.78. In the current sample, the reliability is within the range of previous studies (α = 0.77).

***Neuroticism* – 6 items:** participants were asked to rate were they were on a scale that varied for each item. For example for the trait 'Angry hostility' they had to rate were they were on a scale from 'angry, bitter' to 'even-tempered'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. Across five studies, the reliability of this dimension ranged between 0.62 – 0.69. The reliability is in line with previous estimates (α = 0.70).

(w) Ambition – 5 items—(Duckworth & Quinn, 2009): This measure required participants to rate statements such as 'I aim to be the best in the world at what I do' and 'I am ambitious' on a 5-point scale from 'very much like me' to 'Not like me at all'. The total score was created by taking the mean of the 5 items, requiring at least 3 to be present. The questionnaire from which these questions were drawn has good reliability, with Cronbach's alphas ranging from 0.83 – 0.84 (Duckworth & Quinn, 2009). The reliability in the present sample is slightly lower than estimates from previous studies, but is still considered acceptable (α = 0.74).

(w) Grit – 9 items—(Duckworth & Quinn, 2009): This measure required participants to rate statements such as 'I am driven to succeed' on a 5-point scale from 'Very much like me' to 'Not like me at all'. The total score was created by taking the mean of the 9 items, requiring at least 5 to be present. The questionnaire has good reliability, with Cronbach's alphas ranging from 0.83 – 0.84 (Duckworth & Quinn, 2009). The reliability in the present sample is slightly lower than estimates from previous studies, but is still considered acceptable (α = 0.74).

(w) Curiosity - 7 items—(Kashdan, Rose, & Fincham, 2004): This measure required participants to rate statements such as 'everywhere I go, I am looking out for new things or experiences' and 'I would describe myself as someone who actively seeks as much information as I can in a new situation' on a 7-point scale from 'Strongly agree' to 'Strongly disagree'. The total score was created by taking the mean of the 7 items, requiring at least 4 to be present. Across five studies, the Cronbach's alpha ranged from 0.72 – 0.80 (Kashdan et al., 2004). In the current sample, the reliability is within the range of previous studies (α = 0.74).

(w) Hopefulness – 6 items—(Snyder et al., 1997): This measure required participants to rate sentences about themselves, such as: 'I think I am doing pretty well' and 'I think the things I have done in the past will help me in the future' from 'All of the time' to 'None of the time'. The total score was created by taking the mean of the 6 items, requiring at least 3 to be present. Across eight studies, Cronbach's alpha ranged from 0.72 to 0.86, with a median alpha of 0.77 (Snyder et al., 1997). In the current sample, the reliability is within the range of previous studies (α = 0.83).

(b) Strengths and Difficulties Questionnaire: Behavior Problems – 20 items—(Goodman, 1997): This is a dimensional and developmental measure of child mental health for children aged 3–16 years. Children are required to answer statements on a 3-point Likert scale (Not true; Quite true; Very true). It taps into 4 domains, each of which are measured by 5 items, requiring at least three to be present form the subscale:

Conduct problems: For example: 'I get very angry and often lose my temper'. Reliability estimates across studies range from 0.44 – 0.62 (Mieloo et al., 2012). We found reliability estimates in line with those from other studies (α = 0.53).

Hyperactivity/inattention: For example: 'I am easily distracted, I find it difficult to concentrate'. Reliability estimates across studies range from 0.75 – 0.87 (Mieloo et al., 2012). Our reliability estimate was in line with those reported in previous studies (α = 0.73).

Peer relations: For example: 'I have one good friend or more'. Reliability estimates across studies range from 0.40 – 0.58 (Mieloo et al., 2012). In the current sample, the reliability is within the range of previous studies (α = 0.56).

Prosocial behaviour: For example: 'I try to be nice to other people. I care about their feelings'. Reliability estimates across studies range from 0.59 – 0.82 (Mieloo et al., 2012). In the current sample, the reliability is within the range of previous studies (α = 0.67).

(b) Strengths and Weaknesses of ADHD Symptoms and Normal Behaviour Scale – 18 items—(Swanson et al., 2012): This behavior rating scale is based on DSM-5 criteria for ADHD diagnosis measuring inattentive, hyperactive, and impulsive behaviors. Children are asked to compare themselves to other people of their age on a 7-point scale from 'Far below average' to 'Far above average':

Inattention scale: Derived from 9 items. Item example: 'I sustain attention on tasks or leisure activities' requiring at least half of the items to be present. This scale is scored so that higher scores mean better attention. The reliability for this subscale is 0.91 in one English study and 0.92 in a Spanish study, with good test re-test reliability as well (r = 0.72 and 0.49) (Lakes, Swanson, & Riggs, 2012). Our reliability estimate was in line with those reported in previous studies ($\alpha = 0.88$).

Hyperactivity scale: Derived from 9 items. Item example: 'I sit still (control movement of hands/feet)' requiring at least half of the items to be present. This scale is scored so that higher scores indicate calm and controlled behavior. The reliability for this subscale is 0.93 in one English study and 0.95 in a Spanish study, with good test re-test reliability (r = 0.71 and 0.61) (Lakes et al., 2012). Our reliability estimate was in line with those reported in previous studies ($\alpha = 0.90$).

(w) Gratitude - 6 items—(McCullough, Emmons, & Tsang, 2002): This measure required participants to rate statements such as 'I am grateful to a wide variety of people' and 'I have so much in life to be thankful for' on a 7-point scale from 'Strongly agree' to 'Strongly disagree'. The total score was created by taking the mean of the 6 items, requiring at least 3 to be present. The internal consistency reliability of this scale is 0.82 (McCullough et al., 2002). The reliability is slightly lower than estimates from previous studies, but is still considered acceptable ($\alpha = 0.75$).

(b) Cognitive Disorganisation for cognitive disorganization – 11 items—(Mason, Linney, & Claridge, 2005): This scale, measuring poor attention and concentration, requires individuals to answer 11 items by answering either 'Yes' or 'No'. For example: 'Do you frequently have difficulty in starting to do things?'; 'Do you find it difficult to keep interested in the same thing for a long time?'; 'Is it hard for you to make decisions?' A total score is derived by taking the mean of the 11 items, requiring at least 6 items to be non-missing. Reliability of this scale is good, with Cronbach alpha estimates of 0.77 (Mason et al., 2005). We found the reliability of this scale to be the same as reported previously ($\alpha = 0.77$).

(b) Childhood Anxiety Sensitivity Index – 18 items—(Silverman, Fleisig, Rabian, & Peterson, 1991): This is a child-reported questionnaire measuring anxiety sensitivity (i.e., the belief that anxiety symptoms have negative consequences). Responses are rated on a 3-point Likert scale from 'Not true' to 'Very true'. For example: 'I don't want other people to know when I feel afraid'; 'I get scared when I feel nervous'. A total score is derived by taking the mean of the 18 items, requiring at least 9 items to be non-missing. Reliability of this scale has been tested in clinical and non-clinical samples, both showing good Cronbach alpha's of 0.87 (Silverman et al., 1991). We found the reliability of this scale to be very similar to previous reports of reliability (α = 0.86).

(b) Moods and Feelings Questionnaire (MFQ) Short version – 11 items—(Angold, Costello, Messer, & Pickles, 1995): A brief questionnaire based on DSM-III-R criteria for depression. It is measured on a 3-point Likert scale (Not true; Quite true; Very true) and includes a series of descriptive phrases regarding how the participant has been feeling or acting recently. For example: 'I felt I was no good anymore'; 'I felt lonely'; 'I hated myself'. A total score is derived by taking the mean of the 11 items, requiring at least 6 items to be non-missing. This scale was reversed so that higher scores meant participants felt fewer depressive traits. The reliability of this scale is good, for both the child version (α = 0.85) and the adult version (α = 0.87) (Angold et al., 1995). We found the reliability of this scale to be in line with previous reports of reliability of this scale (α = 0.86).

(w) Life satisfaction – 21 items (Huebner, 1994): This measure taps into different elements of life satisfaction, such as family, school, environment and life satisfaction from friends. It is measured on a 6-point scale from 'Strongly agree' to 'Strongly disagree' and asks participants to rate statements such as: 'I enjoy being at home with my family' and 'I like where I live'. A total score is derived by taking the mean of the 21 items, requiring at least 11 items to be non-missing. Previous studies have shown the reliability of this measure to be good, estimated at α = 0.92 (Huebner, 1994). In the present sample, we found a similar estimate (α = 0.86).

(w) Subjective happiness – 4 items—(Lyubomirsky & Lepper, 1999): These questions tap into perceived happiness, asking participants to complete a sentence. For example: 'In general, I consider myself…' with a 7-point response option from '…Not a very happy person' to '…A very happy person'. A total score is derived by taking the mean of the 4 items, requiring at least 2 items to be non-missing. Reliability estimates from 14 samples ranged from 0.79 – 0.94 (Lyubomirsky & Lepper, 1999). We found the reliability of this scale in our sample to be similar to previously reported estimates (α = 0.78).

(w) Optimism – 6 items—(Scheier, Carver, & Bridges, 1994): This measure required participants to rate statements such as 'In uncertain times, I usually expect the best' and 'I'm always optimistic about my future' on a 5-point scale from 'Very much like me' to 'Not like me at all'. The total score was created by taking the mean of the 6 items, requiring at least 3 to be present. The reliability of this measure is good, estimated at α = 0.82 (Scheier et al., 1994). We found the reliability of this scale in our sample to be similar to previously reported estimates (α = 0.76).

Supplementary Table S1 shows that for most measures, there were small but significant gender differences, and that for some measures there were small effects of age. Prior to any further analyses, all variables were corrected for the effects of gender and age using the regression method to obtain z-standardized residuals.

Due to the large number of measures and the widespread correlations (Supplementary Figure S1), we looked at empirical studies of personality structure and conducted factor analysis

(FA) in order to reduce the large number of measures to six domains. These comprised: Extraversion, Neuroticism, Openness to Experience, Conscientiousness, Agreeableness and Academic Motivation.

Before conducting factor analysis, we performed parallel analysis to guide factor extraction. In parallel analysis, FA is repeatedly applied to sets of randomly generated, uncorrelated data. These data contain the same sample parameters as in the study sample, and by simulating numerous FAs, produces a distribution of eigenvalues. If the component eigenvalue in the study sample is greater than the 95th percentile of the simulated eigenvalues, the retention of this component is justified (Oconnor, 2000). Results from parallel analysis based on our sample parameters (N = 603, based on the total number of individuals with no missing data; number of variables = 28; number of iterations = 1000) indicated the retention of five factors (see Figure S2). To guide our decision-making in creating personality domains, we performed oblique rotation (promax) to allow for correlated factors.

The five-factor FA accounted for 42% of the total variance. Factor loadings revealed an underlying structure representing the FFM. However, instead of an Extraversion factor, there was a factor representing Academic Motivation. The measure of extraversion loaded substantially onto factors of Openness (0.59), Neuroticism (−0.25) and Conscientiousness (−0.26) instead of forming a separate factor. This is presumably because there were no other scales that served as indicators of Extraversion. Based on existing scientific knowledge of personality structure, we decided to re-run the FA excluding extraversion and instead have extraversion as its own separate personality domain. Repeated parallel analysis confirmed the selection of the top five factors for rotation. The final FA, without extraversion also explained 42% of the total variance (Table 1) and item loadings revealed 5 factors: Neuroticism (e.g. cognitive disorganisation and anxiety), Openness to Experience (e.g. ambition and curiosity), Conscientiousness (e.g. attention and focus), Agreeableness (e.g. prosocial behaviour and gratitude) and Academic Motivation (e.g. maths self-efficacy and engagement with key subjects). Item loading are shown in Table 2.

Rather than extracting factor loadings to create personality domains for subsequent analysis, which would lead to a substantial loss of data due to listwise deletion, we created variables by taking the arithmetic mean of the standardized subscales, requiring at least half to be present and reversing measures when they correlated negatively with a factor. Composites based on factor loading extraction and mean composite calculation correlated highly (average $r = 0.91$). Descriptive statistics of the six personality and motivation domains and the 28 subscales are shown in Supplementary Table S1, and correlations between the domains can be found in Supplementary Figure S3.

To test whether there were any meaningful differences between those with missing and non-missing personality and motivation composites, we conducted sensitivity analysis. We assessed mean differences in socio-economic status assessed at first contact (mean composite of parental education, occupation, and maternal age at the birth of the first child), general cognitive ability and GCSE results between missing and non-missing personality and motivation composites scores. We found small differences between those with missing and non-missing data, accounting for an average of 1% (range 0.1% – 2.6%) of the phenotypic variance (see Supplementary Table S2).

Statistical Analyses

Genome-wide polygenic score calculation

For the 10,346 individuals in our sample, we calculated three polygenic scores. The first was based on the summary statistics for a GWA meta-analysis for years of education (N = 766,345 after removal of all 23andme participants) (Lee et al., 2018). The second and third were based on the two largest GWA meta-analyses for personality traits to date, Neuroticism (N = 329,821) (Luciano et al., 2018) and Wellbeing (N = 298,420) (Okbay, Baselmans, et al., 2016).

The first wave of TEDS genotyped samples (N = 2,148) (Trzaskowski et al., 2013) was included in the discovery sample of the Wellbeing GWA meta-analysis. Therefore, we performed a statistical correction on the summary statistic effect size coefficients and *p*-values (Socrates et al., 2017) to account for the overlap between the discovery and target sample. We first replicated the genome-wide association study on Wellbeing using genotypes from the 2,148 TEDS individuals that were included in the meta-analysis, following the GWA protocol applied in the discovery analysis (Okbay, Baselmans, et al., 2016). Secondly, the obtained beta coefficients and standard errors for each SNP were then used to adjust the meta-analyses beta coefficients and standard errors. These adjusted values are analogous to the effects for each SNP if the TEDS sample would have been removed in the discovery meta-analysis (Socrates et al., 2017). Third, we calculated new *p*-values based on the adjusted beta coefficients and standard errors. The adjusted summary statistics for wellbeing were used for polygenic score calculation in the full TEDS sample.

A GPS is calculated by using information from GWA study summary statistics about the strength of association between a genetic variant and a trait, to score individuals' genotypes in independent target samples such as TEDS. Here, we used a Bayesian approach to polygenic score calculation, implemented in the software *LDpred* (Vilhjálmsson et al., 2015). In comparison to conventional clumping and *p*-value thresholding approaches, *LDpred* has demonstrated an improvement in predictive accuracy (Vilhjálmsson et al., 2015). Through this method, a posterior effect size is calculated for each single SNP that is present in both the GWA study summary statistics and the target genotype sample. To calculate this, the original summary statistic effect size estimates are adjusted based on two factors: (1) the relative influence of a SNP given its level of LD with surrounding SNPs in the target sample (here TEDS), and (2) a prior on the effect size of each SNP. This prior depends on the SNP-heritability of the discovery (i.e. GWA study) trait and an assumption on the fraction of causal markers believed to influence the discovery trait. For this study, we set the LD radius to a 2 Megabase window and used a prior based on a fraction of causal markers of 1, meaning that we apply the assumption that all SNPs are causally influencing the discovery trait. Therefore, the prior re-weights the beta effect sizes such that the effects are spread out amongst the SNPs across the whole genome in proportion to the LD present amongst these SNPs. To accommodate the high computational demands of these calculations, we reduced our genotype data set to SNPs that had perfect imputation scores (info = 1), leaving 515,100 SNPs for analysis.

In the next step, all trait-associated alleles were counted (0,1, or 2 for each SNP), weighted by the posterior SNP effect size obtained through *LDpred*, and summed across the genome to calculate a GPS for each individual in TEDS. Although we use a prior based on a fraction of causal markers of 1 to create a GPS for the main analysis, we calculated two more scores with fractions 0.01 and 0.10 for comparison.

To control for platform effects (Affymetrix vs Illumina) and plate effects, as well as effects of population stratification, we regressed all GPS used in this study on platform and plate data, and the first ten principal components. For all subsequent analyses, we used *z*-standardized residuals.

Trait prediction based on regression analysis

To test the extent to which *EduYears* GPS, neuroticism GPS and wellbeing GPS can predict personality traits that are related to GCSE, we used regression analysis. Because these traits are associated with general cognitive ability, we repeated these analyses using the residuals obtained from regressing our personality and motivation traits onto general cognitive ability. We performed bootstrapping with 1000 bootstrap samples, to obtain 95% bootstrap percentile intervals for each coefficient of determination (R^2). To identify whether prediction estimates between the three GPS differed significantly, we used the Williams modification of the Hotelling test (Williams, 1959), which takes into account non-independence of the predictor variables. Additionally, we performed three multiple regression analyses with the polygenic scores as outcomes to assess the relative contributions of general cognitive ability and the personality and motivation phenotypes to polygenic score variation.

Sensitivity analyses for GPS trait prediction

We carried out two types of sensitivity analyses. Firstly, by virtue of the considerable GWA study sample sizes differences between EduYears ($N \sim 760{,}000$) and the personality association studies (neuroticism: $N \sim 330{,}000$; wellbeing; $N \sim 300{,}000$), it is possible that differences in GPS predictions are a product of differences in power to detect effect sizes. We therefore repeated our association analyses between *EduYears* GPS and personality measures using the 2016 GWA study summary statistics based on a sample of ~300,000 individuals to assess any gains in prediction as a result of the steep sample size increase. Supplementary Figure S7.

Secondly, it is a common concern that regression coefficients from GPS analyses are biased due to overfit to the data (Choi, Mak, & O’Reilly, 2018; Wray et al., 2013). Due to the lack of an independent validation sample to test model performance, we carried out internal validation by applying repeated 5 fold cross-validation in our sample to reduce model bias and variability of cross-validation prediction estimates (Kim, 2009). Furthermore, we restricted our sample to unrelated individuals only to simultaneously assess a potential bias due to the inclusion of relatives in our target sample (For descriptive statistics of the unrelated sample, see Supplementary Table S3). For each of the folds, the sample was randomly partitioned into 80% training samples, used to train the model, and 20% validation samples, where each individual appeared only once in the validation sample, used to evaluate the model performance. The 5-fold cross-validation procedure was repeated 50 times with random data splits, and the final cross-validated R^2 estimates were calculated as the average of all model estimates.

GPS prediction of covariance

Finally, we calculated the extent to which each GPS accounts for the relation between personality and motivation domains and GCSE grades using structural equation modelling. We estimated (i) GPS effect on the personality/motivation traits and GCSE grades ($a * b$), (ii) the residual correlation between personality/motivation traits and GCSE results after accounting for the mutual effect of the GPS on both traits (c') and (iii), the total covariance explained by the model ($a * b + c'$). Using this information, it is possible to calculate the extent to which a GPS explains the association between personality/motivation domains and GCSE results ($a * b / a * b + c'$) (see Supplementary Methods S1).

Alpha correction for multiple testing

Multiple testing was accounted for by adjusting the significance threshold by the effective number of tests in accordance with the Nyholt-Šidák correction, which accounts for correlation among the variables. For the Nyholt approach, eigenvalue decomposition is applied to a correlation matrix containing the variables used for analysis, and the eigenvalue variance in relation to the absolute number of variables is used to calculate the effective number of variables (D_{eff}) (Nyholt, 2004). For our analyses, we calculated an effective number of variables based on seven input variables (GCSE results and six personality variables) before and after correcting these variables for general cognitive ability, resulting in D_{eff} of 6.27 and 6.34, respectively. These derived values are then used to calculate the Šidák corrected (Sidak, 1971) significance threshold (alpha = $1 - 0.95^{1/Deff}$). We calculated a total number of 58.83 tests performed for our main analyses. This was calculated by adding together the number of tests: 18.81 tests for comparing each of the three GPS with the seven variables (3 × 6.27), 19.02 tests for comparing the three GPS with the seven variables whilst accounting for general cognitive ability (3 × 6.34), 18 tests to calculate the extent to which the three GPS account for the covariance between GCSE grades and personality traits (3 × 6) and 3 multiple regressions (3). This resulted in a corrected *p*-value threshold of 8.72×10^{-4}.

All analyses were performed in the statistical software R (R Core Team, 2017). Parallel analysis was performed using the ‘parallel’ function in the package *nFactors* (Raiche & Magis, 2010). Factor analysis was performed using the ‘factanal’ function in the *stats* package. Bootstrapping was performed using the ‘boot’ function in the *boot* package (Canty & Ripley, 2012). Robust standard errors were calculated using the ‘coeftest’ function implemented in the *lmtest* package (Zeileis & Hothorn, 2002). Significance of difference between correlation coefficients was tested using the ‘r.test’ function in the *psych* package (Revelle, 2017). Repeated cross-validation was performed using the ‘trainControl’ and ‘train’ function (method ‘lm’) in the package *caret* (Kuhn, 2015). Structural equation modelling analyses were performed using the package *lavaan* (Rosseel et al., 2011), selecting the robust standard error option to account for the clustering in our data due to the inclusions of DZ twin pairs.

Results

Correlations between personality domains and academic achievement

Phenotypic correlations between academic achievement (GCSE results) and the six personality and motivation domains were examined to evaluate the strength of associations between these measures. Pearson's correlation coefficients were statistically significant and absolute values ranged from 0.13 to 0.45 (see Supplementary Figure S3). For correlations between all underlying personality facets and motivation traits and GCSE results, see Supplementary Figure S3.

Polygenic score prediction of personality and academic motivation

To test the predictive validity of the polygenic score for years of education (*EduYears* GPS) and the six personality and motivation domains that contribute to educational success, we performed association analyses. Figure 1A shows that *EduYears* GPS was a significant predictor of all personality/motivation domains but Neuroticism and Extraversion, which did not withstand correction for multiple testing. *EduYears* GPS was significantly positively associated with Agreeableness ($\beta = 0.098$, $p = 2.17 \times 10^{-16}$, $R^2 = 0.010$), Conscientiousness ($\beta = 0.077$, $p = 5.59 \times 10^{-5}$, $R^2 = 0.006$), Openness ($\beta = 0.141$, $p = 5.09 \times 10^{-16}$, $R^2 = 0.021$), and Academic Motivation ($\beta = 0.167$, $p = 3.99 \times 10^{-21}$, $R^2 = 0.029$). The direction of associations indicated that higher *EduYears* GPS scores related to higher Academic motivation, Openness, Conscientiousness and Agreeableness. We also tested the association with GCSE grades, finding EduYears GPS significantly predicted GCSE results ($\beta = 0.370$, $p = 3.36 \times 10^{-288}$, $R^2 = 0.137$), as reported in Allegrini et al., 2018.

The GPS for neuroticism significantly negatively related to GCSE results ($\beta = -0.067$, $p = 1.51 \times 10^{-9}$, $R^2 = 0.044$), Openness ($\beta = -0.65$, $p = 4.37 \times 10^{-3}$, $R^2 = 0.039$) and Academic Motivation composites ($\beta = -0.088$, $p = 6.43 \times 10^{-7}$, $R^2 = 0.074$), and was as expected positively associated with the Neuroticism composite ($\beta = 0.087$, $p = 2.21 \times 10^{-11}$, $R^2 = 0.073$) (Figure 1A). Associations with the Conscientiousness, Extraversion and Agreeableness composite did not survive multiple testing corrections. Overall, the direction of effects indicated that individuals that carry more genetic variants that are related to Neuroticism (i.e. individuals with a higher Neuroticism GPS) scored higher on Neuroticism, had significantly lower GCSE grades, and showed a significant decrease in Openness and Academic Motivation.

The wellbeing GPS was a significant predictor of the Neuroticism composite ($\beta = -0.076$, $p = 1.74 \times 10^{-8}$, $R^2 = 0.056$) and the Agreeableness composite ($\beta = 0.053$, $p = 2.97 \times 10^{-5}$, $R^2 = 0.027$), such that a higher wellbeing GPS related to lower Neuroticism scores, and higher Agreeableness scores. No correlation was found with GCSE score (Figure 1A). Results for other GPS thresholds are reported in Supplementary Figures S4–6.

With the exception of the Neuroticism composite and Extraversion, the magnitudes of the correlation coefficients between *EduYears* GPS and the personality measures were at least twice as high as of those relating to the Neuroticism and Wellbeing GPS. Formal comparisons between correlation coefficients showed that *EduYears* GPS was a significantly stronger predictor than the Neuroticism and Wellbeing GPS for GCSE results ($p = 1.00 \times 10^{-109}$; $p = 1.90 \times 10^{-138}$, respectively), Openness ($p = 8.8 \times 10^{-4}$; $p = 3.00 \times 10^{-6}$, respectively) and Academic Motivation ($p = 3.80 \times 10^{-4}$; $p = 1.40 \times 10^{-10}$, respectively). For Agreeableness, *EduYears* GPS was a better predictor than the Neuroticism GPS ($p = 2.30 \times 10^{-6}$), but not the Wellbeing GPS ($p = 0.006$). The contrasts between the Neuroticism and the Wellbeing GPS showed that the Neuroticism GPS significantly predicted more variance in academic motivation ($p = 7.90 \times 10^{-4}$) and GCSE results ($p =3.20 \times 10^{-5}$).

Controlling for general cognitive ability

General cognitive ability correlated with personality and motivation facets and composites, as well as GCSE grades (Supplementary Figure S3). Therefore, we corrected the composites and GCSE results for variance explained by general cognitive ability and repeated the association analyses as shown in Figure 1B. We found that *EduYears* GPS was still a significant, albeit attenuated, predictor of GCSE grades, Agreeableness, Openness and Academic motivation. For the Neuroticism GPS, previously significant correlations with Academic Motivation and Openness did not reach the multiple-testing corrected *p*-value threshold after accounting for general cognitive ability, and the strength of associations was mostly attenuated for GCSE results. In contrast, the associations with Extraversion and Neuroticism remained significant and of similar strength after correction for general cognitive ability. The correlation between the Wellbeing GPS and the Neuroticism composite remained statistically significant, with no attenuation of effect size. These results suggest that the covariance shared between the GPS and the personality and motivation domains is partly tagged by general cognitive ability, but not solely explained by it. Attenuations were substantially more pronounced for *EduYears* GPS associations (71.3% including GCSE; 73.9% excluding GCSE) than for the neuroticism (50.9% including GCSE; 43.2% excluding GCSE) and Wellbeing GPS (4.5% including GCSE; 5.2% excluding GCSE), indicating that as expected, the *EduYears* GPS tags more genetic variants related to general cognitive ability.

Sensitivity analyses

Associations between the 2016 *EduYears* GPS and personality measures—To assess the extent to which the considerably larger GWA study sample size had on *EduYears* GPS predictions of personality traits relative to the neuroticism and wellbeing GPS, we repeated our analyses using a GPS that is based on the 2016 *EduYears* GPS that has a similar sample size to the neuroticism and wellbeing GWA study. We found that for the personality domains, Pearson's correlation coefficients using the 2016 and the 2018 *EduYears* GPS were almost identical (Supplementary Figure S7), indicating that GWA study power differences between *EduYears and* neuroticism and wellbeing are not likely to explain the differences in predictions of personality measures.

Repeated cross-validation of prediction estimates—To test whether our regression model estimates were biased, potentially due to overfit data or relatedness within the sample, we contrasted them to more robust estimates obtained from repeated 5-fold cross-validation in unrelated samples (Figure 2). Model estimates derived from our previous analyses using the full sample were very similar to the mean of all cross-validated predictions, and without exception fell within the 95% cross-validated R^2 percentile ranges. Moreover, where prediction estimates from our full sample differed, the values were generally more conservative than the mean cross-validated R^2 values. Overall, these comparisons indicate that our model predictions in our full sample are not inflated due to overfitting.

Multiple regression analyses predicting polygenic scores from cognitive ability, personality and academic motivation

To further assess the contributions of cognitive ability and the personality/motivation domains in the polygenic score variation, we performed multiple regression analyses with the polygenic scores as dependent variables. Table 3 shows the beta coefficients for each measure in the joint prediction models. Results for Model 1 indicated that a significant proportion of variance in *EduYears* GPS was explained by the predictors ($F(7,2149) = 29.00$, $p = 1.94 \times 10^{-38}$, $R^2_{adjusted} = 0.083$). The effects were predominantly driven by general cognitive ability and the Agreeableness composite. The overall multiple regression model predicting neuroticism GPS was significant $F(7,2149) = 6.29$, $p = 2.49 \times 10^{-7}$, $R^2 = 0.017$), with the largest effect sizes from individual contributors stemming from general cognitive ability and Neuroticism. The multiple regression model predicting the wellbeing GPS was not statistically significant ($F(7,2149) = 3.11$, $p = 2.87 \times 10^{-3}$, $R^2 = 0.007$), and most of the variance was, albeit not significantly, accounted for by the Neuroticism composite.

Polygenic score prediction of covariation

Because GCSE grades, *EduYears* GPS and the personality and motivation domains are intercorrelated (Supplementary Figure S1), we tested the extent to which *EduYears* GPS accounted for the association between GCSE grades and the personality and motivation domains. Figure 2 and Table 4 show that *EduYears* GPS significantly accounted for a significant amount of covariation between GCSE and Academic Motivation (12.2%, $p = 1.24 \times 10^{-12}$), Openness (14%, $p = 6.06 \times 10^{-11}$), Conscientiousness (7.7%, $p = 8.72 \times 10^{-4}$) and Agreeableness (15.6%, $p = 8.69 \times 10^{-13}$). For comparison, we performed the same analyses using the Neuroticism and Wellbeing GPS. The Neuroticism GPS only accounted for a significant amount of covariance between Neuroticism and GCSE (5%, $p = 1.92 \times 10^{-4}$) (Figure 2; Table 4). No significant covariance was explained by the Wellbeing GPS.

Discussion

Summary of findings

Our results show that a genome-wide polygenic score (GPS) for educational attainment predicts a number of personality and motivation domains, including Agreeableness, Openness, Conscientiousness and Academic Motivation. We find that the educational attainment GPS (*EduYears*) is more predictive of Academic Motivation, Openness and Agreeableness than personality GPS themselves, and that *EduYears* GPS explains between 8–16% of the covariance between personality and motivation domains and educational achievement at age 16. These findings suggest that DNA variants contributing to educational attainment are also important predictors of personality and motivation.

Much of the previous research using *EduYears* GPS has focused on its relation with ‘cognitive’ traits, such as general cognitive ability and educational outcomes (Belsky et al., 2018; Lee et al., 2018; Okbay, Beauchamp, et al., 2016; Rietveld et al., 2013; Selzam, Dale, et al., 2017; Selzam, Krapohl, et al., 2017). In contrast, our findings demonstrate the broad, multifaceted nature of *EduYears* GPS, which is also associated with a variety of personality and motivation traits. Indeed, we show that *EduYears* GPS significantly predicts four out of six personality and motivation domains: Academic motivation, Openness, Conscientiousness, and Agreeableness, explaining between 0.6% and 2.9% of the variance. Our formal comparisons show that for Academic motivation and Openness, *EduYears* GPS was a better predictor than the neuroticism and wellbeing GPS, as well as for Agreeableness in comparison to the neuroticism GPS. In predicting Neuroticism and Extraversion, *EduYears* GPS achieves comparable effect sizes to the neuroticism and wellbeing GPS. Our sensitivity analyses showed that the larger prediction estimates for *EduYears* GPS are not a function of the larger GWA study sample size in comparison to the neuroticism and wellbeing GWA study, as a GPS for *EduYears* based on the 2016 GWA study with a comparably large sample produced almost identical results.

We find that even once we accounted for general cognitive ability, *EduYears* GPS still predicted significant variance in Agreeableness (0.6%), Openness (0.4%), Academic Motivation (0.7%), and GCSE results (6.1%). Correcting for general cognitive ability substantially attenuated associations between the personality traits and *EduYears* GPS (74%), compared to neuroticism GPS (43%), and even less for the wellbeing GPS (5%). Attenuation patterns are also mirrored in the multiple regression analyses. We found that general cognitive ability remains a significant predictor for *EduYears* GPS and neuroticism GPS but not the wellbeing GPS when controlling for all personality measures, and the beta effect sizes are larger for the prediction of *EduYears* than for the neuroticism GPS. One likely explanation for this finding is that the GWA study on years of education tags more general cognitive ability related variants than the neuroticism and wellbeing GWA study. Therefore, statistically controlling for general cognitive ability in the prediction of personality traits would have a greater impact on *EduYears* GPS compared to either neuroticism or wellbeing GPS. The findings that *EduYears* GPS is correlated with personality and motivation traits, even after accounting for general cognitive ability are particularly interesting for two reasons. Firstly, they show that a polygenic score for years of education not only tags genetic variance associated with its target trait, but also many other traits that contribute to how long a person stays in education. And secondly, our findings illustrate that staying in education depends on more than just intelligence; many cognitive and non-cognitive genetically-influenced traits contribute to educational attainment.

In addition to showing that *EduYears* GPS explains significant variance in personality and motivation domains, we also show that it explains between 8 – 16% of the association between personality and motivation domains and educational achievement at age 16. In contrast, the wellbeing GPS did not significantly account for any covariance between these traits and GCSE results, and the neuroticism GPS accounted for a significant amount of variance only in Neuroticism (5%). As previously mentioned, a possible explanation for this finding is that GWA studies performed on personality traits may tag variants specific to the target trait, rather than capturing trait-related variants that also contribute to the development of skills important for educational achievement. In contrast, a GWA study performed on educational attainment is likely to capture genetic variants that are important contributors to many down-stream educationally relevant traits. For example, if motivation is a genetically influenced trait and an important factor for higher educational attainment, a GWA study on years of education will indirectly capture some of the genetic effects relating to motivation if individuals with higher motivation levels are likely to stay in education for longer on average. Another possible mechanism to explain these associations may be that passive rGE is more pronounced for educational attainment than for neuroticism and wellbeing. It has been shown that non-transmitted genetic variants related to educational attainment in parents predict their children's educational achievement, in addition to their children's inherited genetic propensities for educational attainment (Kong et al., 2018). This finding points towards a source of passive rGE, where parents provide a family environment based on their own genetics, which in turn contributes to their children's development, even if they do not share these same markers with their parents. A GWA study on educational attainment might therefore pick up on both the direct effects between the individuals' genetic markers and their educational attainment, and also the effects of the family environment that covaries with their parental non-transmitted genotypes. Therefore, part of the associations we find could be reflecting passive rGE.

Overall, our results demonstrate the substantial genetic pleiotropy (i.e. one DNA marker affects several traits) across educational achievement and educationally relevant traits, although it is not possible to distinguish between biological pleiotropy (i.e. one DNA marker directly affects several traits) and mediated pleiotropy (i.e. one DNA marker directly affects one trait, which then in turn affects another trait (Solovieff et al., 2013). The findings of this study support previous twin research, showing that between 8 – 37% of the covariance between personality traits and GCSE is explained by shared genetic factors (Krapohl et al., 2014). Although the difference between the magnitudes of effect sizes from GPS and twin method results seem large, the GPS effect sizes are substantial given the limitations of the polygenic score method. In contrast to the twin method, which captures all types of genetic variation, GPS results are based on common DNA markers only. Furthermore, the predictive power of polygenic scores is directly related to the power of GWA studies to detect the small SNP effect sizes to begin with, which is one of the main difficulties faced in genetic research (Cesarini & Visscher, 2017). Due to lack of statistical power attributed to sample size and other factors, such as genotyping error or measurement error of the target phenotype, effect size estimates of specific SNPs include measurement error (Dudbridge, 2003; Mark et al., 2008; Van Der Sluis, Verhage, Posthuma, & Dolan, 2010). Therefore, these estimates are not entirely representative of the "true" genetic effect, further contributing to a downward bias of the GPS prediction.

Limitations

Despite the broad range of phenotypes used within the present study, there were limitations to our measures. The first limitation concerns our personality dimension reduction analysis.

Although the five dimensions that emerged from this analysis were closely aligned with the literature on personality, instead of a fifth factor for Extraversion, we found a factor tapping into motivation. There are two reasons for this finding. Firstly, the measures captured by the Academic motivation dimension are not typically included within factor analysis of personality dimensions. These measures, (e.g. academic self-concept, self-efficacy and attitudes towards subjects) correlate with the Conscientiousness dimension ($r = 0.18 – 0.47$), as would be expected given its underlying facets of ‘productive’ and ‘self-discipline’, however most of the variance is left unexplained. Secondly, the underlying facets of Extraversion (e.g. ‘gregarious’, ‘excitement seeking’ and ‘warmth’) were not well covered within our measures. For these reasons, it is not surprising that a separate factor of Extraversion did not emerge. Therefore, we excluded Extraversion from the factor analysis and used this measure by itself in an effort to maintain consistency with the wealth of existing literature describing the distinct factor structure of personality that includes Extraversion.

The second limitation with our measures was the missing data. Because not everyone in our study completed all of the personality and motivation measures, there was missing data for each of our broad dimensions. To make sure that this did not affect the representativeness of the sample, we compared those with missing and non-missing data on socio-economic status, general cognitive ability and achievement at age 16. We found that missingness accounted for 1–3% of the variance in these outcome variables, suggesting that those with missing and non-missing data were not substantially different on these traits.

Another limitation was that we did not have access to parental DNA. This meant that we were unable test the effect of non-transmitted alleles that are related to years of education, neuroticism and wellbeing on offspring personality measures. This would make it possible to estimate the extent to which the associations between the three GPS and the personality domains are influenced by passive rGE. We were also not able to estimate potential effects of active or evocative rGE, which are difficult to investigate because of the lack of adequate measures.

A final limitation concerns a potential overfit to our data. Especially in GPS analyses where parameters for GPS construction are often chosen based on the best prediction of the outcome, prediction estimates can be inflated due to this optimisation. To reduce the chance of overfit, we applied a threshold of 1 to the GPS construction, meaning that all genetic variants are retained (albeit adjusted due to linkage disequilibrium in the sample and the SNP-heritability of the trait). In a further attempt to validate our prediction estimates, we performed internal validation via repeated cross-validation as we had no access to external, independent data. We found that the more stable estimates obtained from repeated cross-validation were largely consistent with our prediction estimates, therefore indicating that our findings were comparably robust.

Conclusion

Despite the limitations to this study, it is the most comprehensive study to date investigating the link between *EduYears* GPS and personality traits. Our findings indicate the pleiotropic nature of the *EduYears* GPS and illustrate that, at a genetic level, staying in education is associated with a multitude of different traits – personality, motivation and intelligence. Although the predictions from polygenic scores are relatively small for personality measures (between 0.6% and 2.9%), this study goes some way in starting to unpack the genetic architecture of educational achievement and associated traits, beyond what we have learnt from twin studies. As GPS prediction improves thanks to the increasing sample sizes of GWA studies and methodological advances, GPS will become more powerful for prediction of education-related measures.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

References

Allegrini A, Selzam S, Rimfeld K, von Stumm S, Pingault J-B, & Plomin R (2018). Genomic prediction of cognitive traits in childhood and adolescence. bioRxiv, 418210.

Angold A, Costello EJ, Messer SC, & Pickles A (1995). Development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. International Journal of Methods in Psychiatric Research, 4(5), 237–249.

Appleton JJ, Christenson SL, Kim D, & Reschly AL (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. Journal of School Psychology, 44(5), 427–445.

Bandura A (1997). Self-efficacy: The exercise of control. New York: Macmillan.

Belsky DW, Domingue BW, Wedow R, Arseneault L, Boardman JD, Caspi A, et al. (2018). Genetic analysis of social-class mobility in five longitudinal studies. Proceedings of the National Academy of Sciences, 115(31), E7275–E7284.

Bidjerano T, & Dai DY (2007). The relationship between the big-five model of personality and self-regulated learning strategies. Learning and individual differences, 17(1), 69–81.

Binet A, & Simon T (1916). The development of intelligence in children: The Binet-Simon Scale: Williams & Wilkins Company.

Bouchard TJ Jr, & McGue M (2003). Genetic and environmental influences on human psychological differences. Journal of neurobiology, 54(1), 4–45. [PubMed: 12486697]

Burden R (1998). Assessing children's perceptions of themselves as learners and problem-solvers: The construction of the Myself-as-Learner Scale (MALS). School Psychology International, 19(4), 291–305.

Busato VV, Prins FJ, Elshout JJ, & Hamaker C (1998). The relation between learning styles, the Big Five personality traits and achievement motivation in higher education. Personality and Individual Differences, 26(1), 129–140.

Busato VV, Prins FJ, Elshout JJ, & Hamaker C (2000). Intellectual ability, learning style, personality, achievement motivation and academic success of psychology students in higher education. Personality and Individual Differences, 29(6), 1057–1068.

Canty A, & Ripley B (2012). boot: Bootstrap R (S-Plus) functions. R package version 1.3–20, 1(7).

Caprara GV, Vecchione M, Alessandri G, Gerbino M, & Barbaranelli C (2011). The contribution of personality traits and self-efficacy beliefs to academic achievement: A longitudinal study. British Journal of Educational Psychology, 81(1), 78–96. [PubMed: 21199485]

Cesarini D, & Visscher PM (2017). Genetics and educational attainment. npj Science of Learning, 2(1), 4. [PubMed: 30631451]

Chamorro-Premuzic T, & Furnham A (2003). Personality traits and academic examination performance. European journal of Personality, 17(3), 237–250.

Choi SW, Mak TSH, & O'Reilly P (2018). A guide to performing Polygenic Risk Score analyses. bioRxiv, 416545.

Conard MA (2006). Aptitude is not enough: How personality and behavior predict academic performance. Journal of Research in Personality, 40(3), 339–346.

De Raad B, & Schouwenburg HC (1996). Personality in learning and education: A review. European Journal of personality, 10(5), 303–336.

Deary IJ, & Johnson W (2010). Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. International journal of epidemiology, 39(5), 1362–1369. [PubMed: 20504860]

Deary IJ, Strand S, Smith P, & Fernandes C (2007). Intelligence and educational achievement. Intelligence, 35(1), 13–21.

Derks EM, Dolan CV, & Boomsma DI (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. Twin Research and Human Genetics, 9(3), 403–411. [PubMed: 16790150]

Duckworth AL, & Quinn PD (2009). Development and validation of the Short Grit Scale (GRIT–S). Journal of personality assessment, 91(2), 166–174. [PubMed: 19205937]

Dudbridge F (2003). Pedigree disequilibrium tests for multilocus haplotypes. Genetic Epidemiology, 25(2), 115–121. [PubMed: 12916020]

Fredricks JA, Blumenfeld PC, & Paris AH (2004). School engagement: Potential of the concept, state of the evidence. Review of Educational Research, 74(1), 59–109.

Frey MC, & Detterman DK (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. Psychological Science, 15(6), 373–378. [PubMed: 15147489]

Geramian SM, Mashayekhi S, & Ninggal MTBH (2012). The relationship between personality traits of international students and academic achievement. Procedia-Social and Behavioral Sciences, 46, 4374–4379.

Goodman R (1997). The Strengths and Difficulties Questionnaire: a research note. Journal of child psychology and psychiatry, 38(5), 581–586. [PubMed: 9255702]

Gratten J, Wray NR, Keller MC, & Visscher PM (2014). Large-scale genomics unveils the genetic architecture of psychiatric disorders. Nature Neuroscience, 17(6), 782. [PubMed: 24866044]

Haworth CM, Davis OS, & Plomin R (2013). Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. Twin Research and Human Genetics, 16(1), 117–125. [PubMed: 23110994]

Heaven PC, & Ciarrochi J (2008). Parental styles, conscientiousness, and academic performance in high school: A three-wave longitudinal study. Personality and Social Psychology Bulletin, 34(4), 451–461. [PubMed: 18340031]

Hirschberg N, & Itkin S (1978). Graduate student success in psychology. American Psychologist, 33(12), 1083.

Holland DC, Dollinger SJ, Holland CJ, & Macdonald DA (1995). The relationship between psychometric intelligence and the five-factor model of personality in a rehabilitation sample. Journal of Clinical Psychology, 51(1), 79–88. [PubMed: 7782479]

Huebner ES (1994). Preliminary development and validation of a multidimensional life satisfaction scale for children. Psychological Assessment, 6(2), 149.

Jang KL, Livesley WJ, & Vemon PA (1996). Heritability of the big five personality dimensions and their facets: a twin study. Journal of personality, 64(3), 577–592. [PubMed: 8776880]

Jencks C (1979). Who Gets Ahead? The Determinants of Economic Success in America. New York, NY: Basic Books.

Kaplan E, Fein D, Kramer J, Delis D, & Morris R (1999). The WISC-III as a process instrument. San Antonio, TX: The Psychological Corporation.

Kashdan TB, Rose P, & Fincham FD (2004). Curiosity and exploration: Facilitating positive subjective experiences and personal growth opportunities. Journal of Personality Assessment, 82(3), 291–305. [PubMed: 15151805]

Kelly WE, & Johnson JL (2005). Time Use Efficiency and the Five-Factor Model of Personality. Education, 125(3).

Kendler KS, Neale MC, Kessler RC, Heath AC, & Eaves LJ (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. Behavior Genetics, 23(1), 21–27. [PubMed: 8476388]

Kim J-H (2009). Estimating classification error rate: Repeated cross-validation, repeated hold-out and bootstrap. Computational statistics & data analysis, 53(11), 3735–3745.

Knopik VS, Neiderhiser JM, DeFries JC, & Plomin R (2017). Behavioral Genetics (7th ed.). London, UK: Worth Publishers.

Komarraju M, & Karau SJ (2005). The relationship between the big five personality traits and academic motivation. Personality and Individual Differences, 39(3), 557–567.

Kong A, Thorleifsson G, Frigge ML, Vilhjalmsson BJ, Young AI, Thorgeirsson TE, et al. (2018). The nature of nurture: Effects of parental genotypes. Science, 359(6374), 424–428. [PubMed: 29371463]

Krapohl E, Rimfeld K, Shakeshaft NG, Trzaskowski M, McMillan A, Pingault J-B, et al. (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.

Kuhn M (2015). caret: Classification and Regression Training. Astrophysics Source Code Library.

Laidra K, Pullmann H, & Allik J (2007). Personality and intelligence as predictors of academic achievement: A cross-sectional study from elementary to secondary school. Personality and Individual Differences, 42(3), 441–451.

Lakes KD, Swanson JM, & Riggs M (2012). The reliability and validity of the English and Spanish strengths and weaknesses of ADHD and normal behavior rating scales in a preschool sample: Continuum measures of hyperactivity and inattention. Journal of Attention Disorders, 16(6), 510–516. [PubMed: 21807955]

Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. (2018). Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. Nature genetics, 50(8), 1112. [PubMed: 30038396]

Luciano M, Hagenaars SP, Davies G, Hill WD, Clarke T-K, Shirali M, et al. (2018). Association analysis in over 329,000 individuals identifies 116 independent variants influencing neuroticism. Nature Genetics, 50(1), 6. [PubMed: 29255261]

Lyubomirsky S, & Lepper HS (1999). A measure of subjective happiness: Preliminary reliability and construct validation. Social Indicators Research, 46(2), 137–155.

Mark M, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JP, et al. (2008). Genome-wide association studies for complex traits: consensus, uncertainty and challenges. Nature reviews genetics, 9(5), 356.

Mason O, Linney Y, & Claridge G (2005). Short scales for measuring schizotypy. Schizophrenia Research, 78(2), 293–296. [PubMed: 16054803]

McCarthy D (1972). McCarthy Scales of Children's Abilities. New York: Psychological Corporation.

McCarthy S, Das S,;, K. W, Delaneau O, Wood AR, Teumer A, et al. (2016). A reference panel of 64,976 haplotypes for genotype imputation. Nature Genetics, 48(10), 1279. [PubMed: 27548312]

McCrae RR, & Costa PT (1987). Validation of the five-factor model of personality across instruments and observers. Journal of Personality and Social Psychology, 52(1), 81. [PubMed: 3820081]

McCrae RR, & Costa PT (1997). Conceptions and correlates of openness to experience In Handbook of Personality Psychology (pp. 825–847): Elsevier.

McCullough ME, Emmons RA, & Tsang J-A (2002). The grateful disposition: a conceptual and empirical topography. Journal of Personality and Social Psychology, 82(1), 112. [PubMed: 11811629]

Mieloo C, Raat H, van Oort F, Bevaart F, Vogel I, Donker M, et al. (2012). Validity and reliability of the Strengths and Difficulties Questionnaire in 5–6 year olds: differences by gender or by parental education? PloS one, 7(5), e36805. [PubMed: 22629332]

Mõttus R, Realo A, Vainik U, Allik J, & Esko T (2017). Educational attainment and personality are genetically intertwined. Psychological science, 28(11), 1631–1639. [PubMed: 28910230]

Mullins-Sweatt SN, Jamerson JE, Samuel DB, Olson DR, & Widiger TA (2006). Psychometric properties of an abbreviated instrument of the five-factor model. Assessment, 13(2), 119–137. [PubMed: 16672728]

Multon KD, Brown SD, & Lent RW (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. Journal of Counseling Psychology, 38(1), 30.

Nyholt D (2004). A simple correction for multiple testing for single-nucleotide polymorphisms in linkage disequilibrium with each other. The American Journal of Human Genetics, 74(4), 765–769. [PubMed: 14997420]

O'Connor MC, & Paunonen SV (2007). Big Five personality predictors of post-secondary academic performance. Personality and Individual Differences, 43(5), 971–990.

Oconnor BP (2000). SPSS and SAS programs for determining the number of components using parallel analysis and Velicer's MAP test. Behavior Research Methods, Instruments, & Computers, 32(3), 396–402.

OECD. (2007). Education at a Glance 2007. Paris, France: Organisation for Economic Co-operation and Development.

OECD. (2017). Education at a Glance 2017. Paris, France: Organisation for Economic Co-operation and Development.

Okbay A, Baselmans BM, De Neve J-E, Turley P, Nivard MG, Fontana MA, et al. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. Nature Genetics, 48(6), 624. [PubMed: 27089181]

Okbay A, Beauchamp JP, Fontana MA, Lee JJ, Pers TH, Rietveld CA, et al. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. Nature, 533(7604), 539. [PubMed: 27225129]

Parker PD, Marsh HW, Ciarrochi J, Marshall S, & Abduljabbar AS (2014). Juxtaposing math self-efficacy and self-concept as predictors of long-term achievement outcomes. Educational Psychology, 34(1), 29–48.

Petrides K, Chamorro-Premuzic T, Frederickson N, & Furnham A (2005). Explaining individual differences in scholastic behaviour and achievement. British Journal of Educational Psychology, 75(2), 239–255. [PubMed: 16033665]

Plomin R, & Deary IJ (2015). Genetics and intelligence differences: five special findings. Molecular Psychiatry, 20(1), 98. [PubMed: 25224258]

Plomin R, DeFries JC, & Loehlin JC (1977). Genotype-environment interaction and correlation in the analysis of human behavior. Psychological bulletin, 84(2), 309. [PubMed: 557211]

Plomin R, & von Stumm S (2018). The new genetics of intelligence. Nature Reviews Genetics.

Polderman TJ, Benyamin B, De Leeuw CA, Sullivan PF, Van Bochoven A, Visscher PM, et al. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. Nature genetics, 47(7), 702. [PubMed: 25985137]

Poropat AE (2009). A meta-analysis of the five-factor model of personality and academic performance. Psychological Bulletin, 135(2), 322. [PubMed: 19254083]

Purcell SM, Wray NR, Stone JL, Visscher PM, O'Donovan MC, Sullivan PF, et al. (2009). Common polygenic variation contributes to risk of schizophrenia and bipolar disorder. Nature, 460(7256), 748. [PubMed: 19571811]

R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: Foundation for Statistical Computing.

Raiche G, & Magis D (2010). nFactors: An R package for parallel analysis and non graphical solutions to the Cattell scree test. R package version 2, 2(3).

Raven J, & Raven J (1998). Court JH (1998). Manual for Raven's progressive matrices and vocabulary scales.

Raven JC, Raven JE, & Court JH (1989). Mill Hill vocabulary scale: Psychological Corporation.

Ray A, & Margaret W (2003). PISA Programme for international student assessment (PISA) PISA 2000 technical report: PISA 2000 technical report: oecd Publishing.

Revelle WR (2017). psych: Procedures for Personality and Psychological Research. Northwestern University Scholars

Richardson M, Abraham C, & Bond R (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. Psychological Bulletin, 138(2), 353. [PubMed: 22352812]

Rietveld CA, Medland SE, Derringer J, Yang J, Esko T, Martin NW, et al. (2013). GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. Science, 340(6139), 1467–1471. [PubMed: 23722424]

Ritchie SJ, & Tucker-Drob EM (2018). How much does education improve intelligence? A meta-analysis. Psychological science, 29(8), 1358–1369. [PubMed: 29911926]

Rosseel Y, Oberski D, Byrnes J, Vanbrabant L, Savalei V, Merkle E, et al. (2011). The lavaan package for R.

Scheier MF, Carver CS, & Bridges MW (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): a reevaluation of the Life Orientation Test. Journal of personality and social psychology, 67(6), 1063. [PubMed: 7815302]

Schunk DH (1989). Self-efficacy and achievement behaviors. Educational psychology review, 1(3), 173–208.

Selzam S, Dale PS, Wagner RK, DeFries JC, Cederlöf M, O'Reilly PF, et al. (2017). Genome-wide polygenic scores predict reading performance throughout the school years. Scientific Studies of Reading, 21(4), 334–349. [PubMed: 28706435]

Selzam S, Krapohl E, von Stumm S, O'reilly P, Rimfeld K, Kovas Y, et al. (2017). Predicting educational achievement from DNA. Molecular psychiatry, 22(2), 267. [PubMed: 27431296]

Selzam S, McAdams TA, Coleman JR, Carnell S, O'Reilly PF, Plomin R, et al. (2018). Evidence for gene-environment correlation in child feeding: Links between common genetic variation for BMI in children and parental feeding practices. PLoS Genetics, 14(11), e1007757. [PubMed: 30457987]

Sidak Z (1971). On probabilities of rectangles in multivariate Student distributions: their dependence on correlations. The Annals of Mathematical Statistics, 42(1), 169–175.

Silverman WK, Fleisig W, Rabian B, & Peterson RA (1991). Childhood anxiety sensitivity index. Journal of Clinical Child and Adolescent Psychology, 20(2), 162–168.

Snyder CR, Hoza B, Pelham WE, Rapoff M, Ware L, Danovsky M, et al. (1997). The development and validation of the Children's Hope Scale. Journal of pediatric psychology, 22(3), 399–421. [PubMed: 9212556]

Socrates A, Bond T, Karhunen V, Auvinen J, Rietveld C, Veijola J, et al. (2017). Polygenic risk scores applied to a single cohort reveal pleiotropy among hundreds of human phenotypes. bioRxiv, 203257.

Solovieff N, Cotsapas C, Lee PH, Purcell SM, & Smoller JW (2013). Pleiotropy in complex traits: challenges and strategies. Nature Reviews Genetics, 14(7), nrg3461.

Spinath B, Spinath FM, Harlaar N, & Plomin R (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. Intelligence, 34(4), 363–374.

Swanson JM, Schuck S, Porter MM, Carlson C, Hartman CA, Sergeant JA, et al. (2012). Categorical and Dimensional Definitions and Evaluations of Symptoms of ADHD: History of the SNAP and the SWAN Rating Scales. The International journal of educational and psychological assessment, 10(1), 51. [PubMed: 26504617]

Trautwein U, & Lüdtke O (2007). Students' self-reported effort and time on homework in six school subjects: Between-students differences and within-student variation. Journal of educational psychology, 99(2), 432.

Trautwein U, Lüdtke O, Roberts BW, Schnyder I, & Niggli A (2009). Different forces, same consequence: Conscientiousness and competence beliefs are independent predictors of academic effort and achievement. Journal of personality and social psychology, 97(6), 1115. [PubMed: 19968422]

Trzaskowski M, Eley TC, Davis OS, Doherty SJ, Hanscombe KB, Meaburn EL, et al. (2013). First genome-wide association study on anxiety-related behaviours in childhood. PloS one, 8(4), e58676. [PubMed: 23565138]

Tucker-Drob EM, Briley DA, Engelhardt LE, Mann FD, & Harden KP (2016). Genetically-mediated associations between measures of childhood character and academic achievement. Journal of personality social psychology, 111(5), 790. [PubMed: 27337136]

Van Der Sluis S, Verhage M, Posthuma D, & Dolan CV (2010). Phenotypic complexity, measurement bias, and poor phenotypic resolution contribute to the missing heritability problem in genetic association studies. PloS one, 5(11), e13929. [PubMed: 21085666]

Vilhjálmsson BJ, Yang J, Finucane HK, Gusev A, Lindström S, Ripke S, et al. (2015). Modeling linkage disequilibrium increases accuracy of polygenic risk scores. 97(4), 576–592.

Wagerman SA, & Funder DC (2007). Acquaintance reports of personality and academic achievement: A case for conscientiousness. Journal of Research in Personality, 41(1), 221–229.

Webb E (1915). Character and intelligence: An attempt at an exact study of character: University Press.

Wechsler D, Golombok S, & Rust J (1992). WISC-III UK Wechsler intelligence scale for children: UK manual. Sidcup, UK: The Psychological Corporation.

Williams EJ (1959). The comparison of regression variables. Journal of the Royal Statistical Society, 396–399.

Wray NR, Yang J, Hayes BJ, Price AL, Goddard ME, & Visscher PM (2013). Pitfalls of predicting complex traits from SNPs. Nature Reviews Genetics, 14(7), 507.

Zeidner M, & Matthews G (2000). Intelligence and personality.

Zeileis A, & Hothorn T (2002). Diagnostic checking in regression relationships. R News 2 (3): 7–10. R News, 2(3), 7–10.

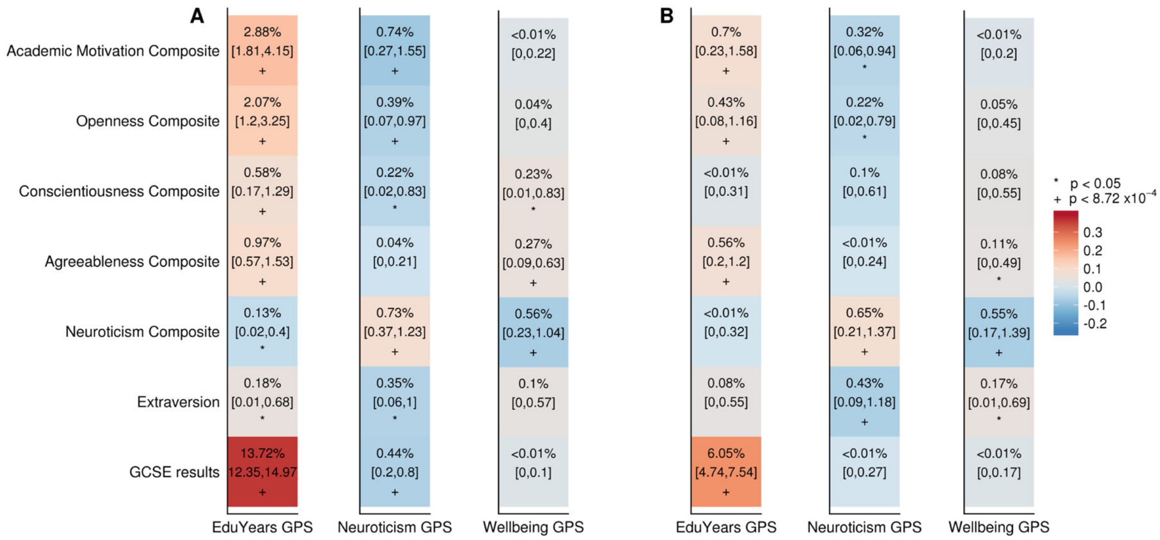

Figure 1.

Genome-wide polygenic scores (GPS) as a predictor of six personality and motivation domains and GCSE (General Certificate of Secondary Education) results (**A**) before and (**B**) after accounting for general cognitive ability. The colour shading represents magnitude of Pearson Correlation coefficients and the values in each cell represent the amount of phenotypic variance explained by the polygenic scores. Values in square brackets represent the lower and upper bounds of the 95% bootstrapped percentile intervals based on 1000 bootstrap samples. '+' = *p*-value threshold for significance after correction for multiple testing (8.72×10^{-4})

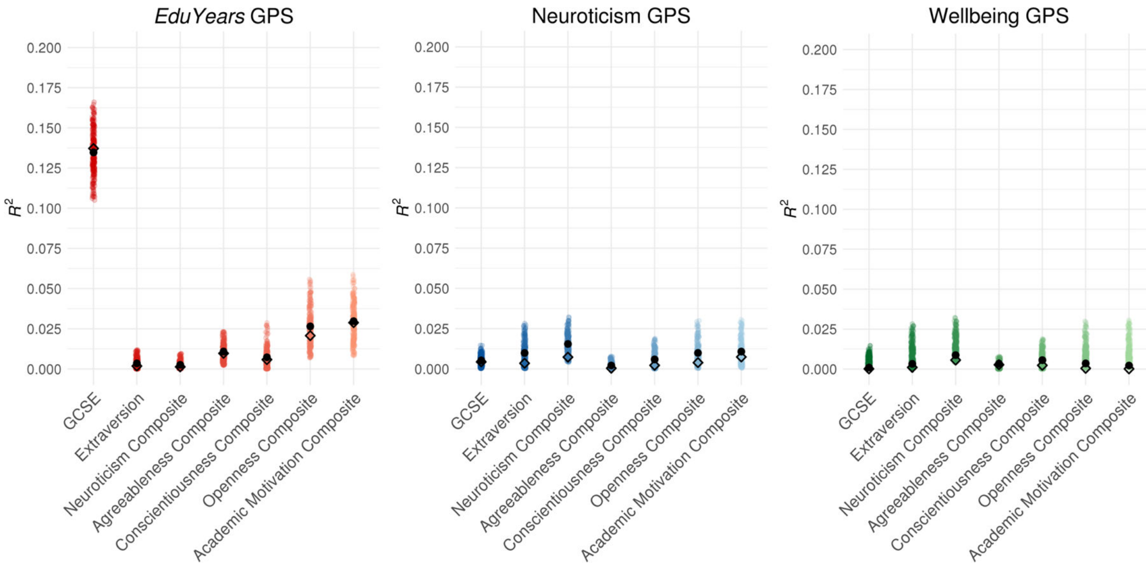

Figure 2.

Comparison of prediction in full sample including DZ twin pairs to cross-validated predictions in unrelated individuals. We performed 5 fold cross-validation with 50 repetitions. R^2 estimates that fall within the 95% percentile of all cross-validation prediction estimates are represented by the coloured dots. The black dots indicate the mean of these the cross-validated R^2 and the diamond shaped symbols indicate the prediction estimates obtained from the original regression models using the full sample.

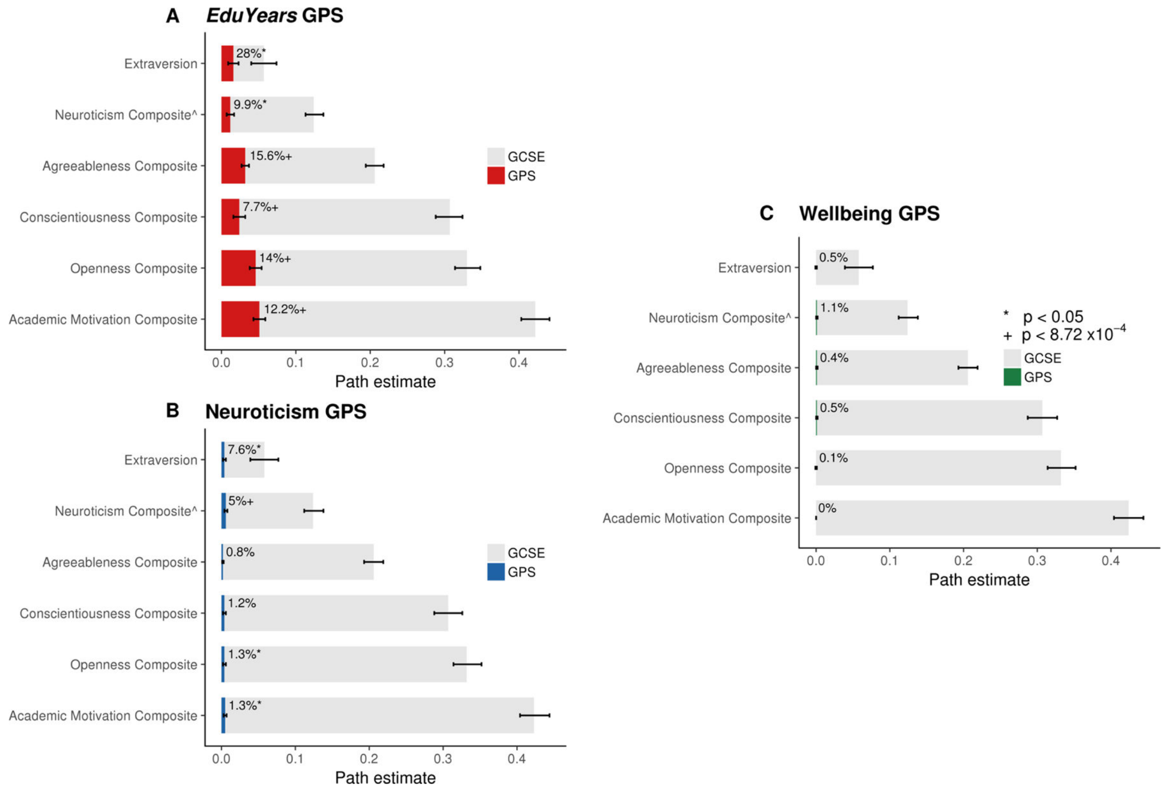

Figure 3.

Standardized path estimates for the association between GCSE (General Certificate of Secondary Education) grades and personality/motivation domains, and the proportion of these associations accounted for by (**A**) *EduYears* GPS (genome-wide polygenic scores) (**B**) Neuroticism GPS and (**C**) Wellbeing GPS. Error bars represent robust standard errors. Path estimates presented are estimated based on maximum likelihood (see Table 4 for all path estimates). '^' = the direction of association between the Neuroticism composite and GCSE grades was negative; '+' = *p*-value threshold for significance after correction for multiple testing (8.72×10^{-4}).

Table 1.

Factor Analysis

Extracted factors	Rotation Sums of Squared Loadings		
	Total	**% of Variance**	**Cumulative %**
1	3.085	11.426	11.426
2	2.824	10.459	21.885
3	2.288	8.472	30.358
4	1.586	5.875	36.233
5	1.566	5.799	42.032

Note: Individual and cumulative variance explained by the five factors

Table 2.

Rotated item factor loadings

	Neuroticism	Openness	Conscientiousness	Agreeableness	Academic motivation
Cognitive disorganisation	0.52		0.49		
Anxiety	0.62				
MFQ	0.80				
Subjective happiness	−0.67			0.21	
Life satisfaction	−0.68			0.34	
Peer Problems	0.54				
Academic self-concept		0.50			0.36
Ambition		0.68			
Curiosity		0.71			
Hopefulness	−0.25	0.72			
SDQ Hyperactivity			0.80		
SWAN Hyperactivity			−0.67		
SWAN Inattention		0.21	−0.61		
Agreeableness				0.68	
SDQ Prosocial				0.51	
Maths Self-efficacy					0.96
Maths interest					0.54
Maths time spent					0.25
Attitudes key subjects		0.20			0.18
School engagement				0.21	
Openness		0.24	0.22		
Conscientiousness	0.22	0.35	−0.39		
SDQ conduct	0.20	0.34	0.27	−0.48	
GRIT		0.44	−0.41		
Gratitude	−0.21	0.26		0.43	
Neuroticism	0.45	−0.31			
Optimism	−0.29	0.46			

Note: Oblique (promax) rotation was applied. Only variables with factor loadings of ≥ 0.20 are shown. Items included in the same composite are highlighted in grey.

Table 3.

Results from multiple regression analyses: Cognitive and personality/motivation composites predicting genome-wide polygenic scores

	Model 1:EduYears GPS			Model 2: Neuroticism GPS			Model 3: Wellbeing GPS		
predictors	***β***	***SE***	***p***	***β***	***SE***	***p***	***β***	**SE**	***p***
general cognitive ability	**0.252**	**0.024**	**$7.54\times10^{-25+}$**	**−0.079**	**0.023**	**$5.56\times10^{-4+}$**	0.045	0.024	0.058
Academic motivation	0.071	0.026	0.007^{*}	−0.042	0.026	0.102	−0.045	0.027	0.093
Openness	0.041	0.029	0.158	−0.010	0.030	0.734	0.020	0.031	0.506
Conscientiousness	−0.059	0.027	0.032	0.013	0.027	0.627	−0.012	0.028	0.668
Agreeableness	**0.115**	**0.031**	**$1.71\times10^{-4+}$**	0.066	0.031	0.038^{*}	−0.019	0.034	0.570
Neuroticism	0.041	0.027	0.120	**0.095**	**0.028**	**$6.05\times10^{-4+}$**	−0.096	0.030	$1.37\times10^{-3*}$
Extraversion	0.009	0.025	0.719	−0.030	0.024	0.213	<0.001	0.026	0.997

Note. Beta coefficients, standard errors and p-values are presented for each of the predictors in the regression models.

* $p < 0.05$,

'+' $= 8.72 \times 10^{-4}$ (*p*-value threshold for significance after correction for multiple testing).

Table 4.

Path estimates and standard errors

	parameter	*EduYears* GPS β	*EduYears* GPS *robust SE*	Neuroticism GPS β	Neuroticism GPS *robust SE*	Wellbeing GPS β	Wellbeing GPS *robust SE*
Academic motivation Composite	GPS effect	0.051^{+}	0.008	0.005*	0.002	<0.001	<0.001
	resid cor	0.371^{+}	0.019	0.418^{+}	0.020	0.424^{+}	0.020
	total effect	0.422^{+}	0.020	0.424^{+}	0.020	0.424^{+}	0.020
	proportion	**0.122**$^{+}$	**0.017**	0.013*	0.005	<0.001	<0.001
Openness Composite	GPS effect	0.046^{+}	0.008	0.004*	0.002	<0.001	<0.001
	resid cor	0.284^{+}	0.017	0.328^{+}	0.019	0.332^{+}	0.019
	total effect	0.331^{+}	0.018	0.333^{+}	0.019	0.232^{+}	0.019
	proportion	**0.139**$^{+}$	**0.021**	0.013	0.006	<0.001	<0.001
Neuroticism Composite	GPS effect	−0.012*	0.005	−0.006^{+}	0.002	−0.001	0.001
	resid cor	−0.112^{+}	0.012	−0.118^{+}	0.013	−0.123^{+}	0.013
	total effect	−0.125^{+}	0.013	−0.125^{+}	0.013	−0.123^{+}	0.013
	proportion	0.099*	0.036	**0.050**$^{+}$	**0.013**	0.011	0.008
Conscientiousness Composite	GPS effect	0.024*	0.009	0.004	0.002	0.001	0.001
	resid cor	0.283^{+}	0.029	0.303^{+}	0.019	0.306^{+}	0.020
	total effect	0.306^{+}	0.031	0.307^{+}	0.020	0.306^{+}	0.020
	proportion	**0.077**$^{+}$	**0.023**	0.013	0.007	0.004	0.004
Agreeableness Composite	GPS effect	0.032^{+}	0.005	0.002*	0.001	<0.001	<0.001
	resid cor	0.174^{+}	0.012	0.204^{+}	0.013	0.205^{+}	0.013
	total effect	0.206^{+}	0.013	0.206^{+}	0.013	0.205^{+}	0.013
	proportion	**0.156**$^{+}$	**0.022**	0.008	0.005	0.003	0.003
Extraversion	GPS effect	0.016*	0.007	0.004*	0.002	<0.001	<0.001

parameter	*EduYears* GPS β	*EduYears* GPS *robust SE*	Neuroticism GPS β	Neuroticism GPS *robust SE*	Wellbeing GPS β	Wellbeing GPS *robust SE*
resid cor	0.041*	0.017	0.054+	0.019	0.058+	0.019
Total effect	0.057*	0.018	0.058+	0.019	0.058+	0.019
proportion	0.278*	0.116	0.076*	0.038	0.004	0.012

Note. GPS effect = effect of the genome-wide polygenic score (GPS) on both traits; resid cor = residual correlation between phenotypes after mutually adjusting for the effects of the GPS, total effect = effect accounted for by the model (resid cor + GPS effect); proportion = the proportion of the total effect that is accounted for by the GPS effect (GPS effect / total effect). Statistically significant proportions of variance explained are in bold.

* $p < 0.05$,

'+' $= 8.72 \times 10^{-4}$ (p-value threshold for significance after correction for multiple testing)