Educational attainment and personality are genetically intertwined

René Mõttus *
Department of Psychology, University of Edinburgh
Institute of Psychology, University of Tartu, Estonia

Anu Realo
Department of Psychology, University of Warwick
Institute of Psychology, University of Tartu, Estonia

Uku Vainik
Montreal Neurological Institute, McGill University
Institute of Psychology, University of Tartu, Estonia

Jüri Allik
Institute of Psychology, University of Tartu, Estonia
Estonian Academy of Sciences, Tallinn, Estonia

Tõnu Esko
Estonian Genome Centre of University of Tartu, Estonia
Broad Institute, Cambridge, USA

* Corresponding author
7 George Square, EH8 9JZ Edinburgh, UK
Phone: +441316503453
E-mail: rene.mottus@ed.ac.uk

Abstract

It is possible that heritable variance in personality characteristics does not reflect (only) genetic and biological processes specific to personality *per se*. We tested the possibility that Five-Factor Model personality domains and facets, as rated by people themselves and their knowledgeable informants, reflect polygenic influences that have been previously associated with educational attainment. In a sample of over 3,000 adult Estonians, polygenic scores for educational attainment, based on small contributions from more than 150,000 genetic variants, were correlated with various personality traits, mostly from the Neuroticism and Openness domains. The correlations of personality characteristics with educational attainment-related polygenic influences reflected almost entirely their correlations with phenotypic educational attainment. Structural equation modeling of the associations between polygenic risk, personality (a weighed aggregate of education-related facets) and educational attainment lent relatively strongest support to the possibility of educational attainment mediating (explaining) some of the heritable variance in personality traits.

Keywords: personality; Big Five; education; polygenic; genetic correlation

Educational attainment and personality are genetically intertwined

Personality trait variance has a substantial genetic component (Vukasović & Bratko, 2015). However, the specific genetic variants responsible for this have largely remained elusive, possibly due to the highly polygenic nature of the traits (Chabris et al., 2013). Collectively, large numbers of common genetic variants explain up to about 15% of variance in personality traits (Mõttus, Marioni, & Deary, in press; Smith et al., 2016; Vinkhuyzen et al., 2012), but the effect of any one gene is usually too small to be reliable detectable. The same tends to be true for other psychological phenotypes such as cognitive ability (Davies et al., 2015) and subjective well-being (Okbay, Baselmans, et al., 2016). In contrast, slightly more genetic variance has been traced to specific genetic variants for some less-psychological complex phenotypes such as educational attainment (Okbay, Beauchamp, et al., 2016) and body mass index (Locke et al., 2015).

It has also been suggested that personality traits could be conceived of as mostly phenotypic phenomena with limited or even no genetic or biological architecture of their own (Turkheimer, Pettersson, & Horn, 2014). If so, observed genetic variance in personality characteristics may to some or even large extent reflect genetic influences that act broadly across the nervous system or even across the organism more generally (as a general "genetic pull"; Turkheimer et al., 2014) rather than contributing to some systems specifically responsible for what appear as personality traits. In this case, the genetic and resultant biological underpinnings of personality traits should be shared with those of other phenomena that phenotypically relate to these personality traits but fall outside of how the traits are typically defined and operationalized (Mõttus, Marioni, et al., in press).

Here, we address this possibility by investigating whether phenotypic variability in personality traits is associated with polygenic propensity for educational attainment (henceforth education) as estimated from molecular genetic data. There are numerous phenotypes that may share genetic influences with personality characteristics. We choose educational level because it is a broad behavioral phenotype that has a sizable heritable component (Colodro-Conde, Rijsdijk, Tornero-Gómez, Sánchez-Romera, & Ordoñana, 2015; Silventoinen, Krueger, Bouchard, Kaprio, & McGue, 2004), is phenotypically correlated with a spectrum of personality traits (Chapman, Fiscella, Kawachi, & Duberstein, 2010; Digman, 1989; Shiner, Masten, & Roberts, 2003) and yet is not part of how the traits are usually operationalized. Also, education has been relatively well characterized in terms of its genomic correlates: variability in this trait is known to be associated with a large number of genetic variants and these specific associations have already been quantified with a reasonable level of accuracy (Okbay, Beauchamp, et al., 2016).

Twin studies have revealed that the phenotypic correlations of several personality traits with children's and adolescence academic results can largely be accounted for by shared genetic influences (Hicks, Johnson, Iacono, & McGue, 2008; Rimfeld, Kovas, Dale, & Plomin, 2016). In addition to additive influences of genetic variants, these estimates reflect non-additive effects due to interactions between and within genetic loci, effects of rare variants and person-environment correlations (Purcell, 2002), and they are possibly confounded with the environmental effects that twins share (Vinkhuyzen et al., 2012). Recently, Belsky and colleagues (2016) showed that polygenic variance in education is associated with two childhood personality characteristics, self-control and interpersonal skills; these findings only pertain to additive genetic effects of common genetic variants. Likewise, Okbay and colleagues (Okbay, Beauchamp, et al., 2016) showed a negative polygenic correlation between education and Neuroticism. However, it is not known whether these (additive) polygenic associations generalize to other personality characteristics, including those of the omnibus personality models such as the Five-Factor Model (FFM).

Disentangling causality

It seems unlikely that education itself reflects a distinct psychobiological attribute and thereby corresponds to genetic variance that is somehow specific to this phenotype. Instead, its genetic variance is likely to be shared with that of other characteristics such as cognitive abilities or other contributors to socioeconomic success, including behaviors encompassed by personality traits. For example, genetic variance in education largely overlaps with that of cognitive abilities and (whatever leads to) social deprivation (Marioni et al., 2014). Therefore, when education shows genetic overlap with personality characteristics, this may imply different scenarios: they may both be independently influenced by the same genetic factors, one may mediate (or explain) the genetic effects of another, or the mediation may work in both ways. In case of mediation, for example, education and thereby its unique genetic influences may phenotypically contribute to certain personality characteristics (e.g., Openness or Conscientiousness), or the other way around—personality characteristics and thereby their unique genetic influences may explain some of the genetic variance in education (Rimfeld et al., 2016). This means that finding genetic correlations with education would not *inevitably* tell us something on the relative lack of the distinctive etiology of these personality traits: the traits may have their own genetic underpinnings that just happen to bleed into educational level via phenotypic causation.

In order to tackle causality, we could study people with no “exposure” to the hypothesized mediators (Kippersluis & Rietveld, 2016). For example, if the otherwise present genetic correlation between genetic propensity for education and personality traits is missing in people without educational experience (i.e., the mediation pathway is broken), this would support education being a mediator in the genetic association. To study the possible mediating role of a personality trait, we could investigate people without the trait. Alternatively, it would help to know specific genetic variants with known causal pathways to the purported mediators (approach known as Mendielian Randomizaton; Davey Smith, 2010). For example, if some genetic variants have direct causal links with education but are unlikely to be similarly directly causal to a personality trait and yet correlate with the trait, this would support education being phenotypically causal to the personality trait and thereby mediating its genetic influences.

It may be difficult to find people with no exposure to education or without a specific personality trait, and the genetic variants with clear causal pathways to these phenotypes are currently unknown. Meanwhile, statistical techniques can be used to estimate the plausibility of different causal scenarios. For example, if personality traits and education only share genetic influences with no mediation of one another, then, conditional on the genetic influences, the two should be independent. Alternatively, we may estimate statistical mediation. For instance, if personality traits appear to account for a relatively larger share of genetic propensity-education associations than education can account for genetic propensity-personality associations, this would be more consistent with personality traits mediating the genetic effects of education than the other way around.

The current study

Employing published meta-analytic associations (Okbay, Beauchamp, et al., 2016) between education and single nucleotide polymorphisms (SNP), we created polygenic scores for education (EPS) for 3,061 adult Estonians. We then correlated the EPS with the FFM domains and facets; significant correlations pointed to shared genetic influences on these personality traits and education. Next, we aggregated the personality facet correlates of education into a single aggregate variable (polyfacet scores) and did the same for the EPS. Correlation between these two variables quantifies the degree to which the personality correlates of education generally track its additive polygenic variance component. Finally, we tested whether personality traits and education would become independent after controlling for EPS, which would suggest shared genetic influences with no mediation, or whether either education or its personality correlates would account for a larger share of each other's polygenic influences. The use of both self- and informant-rated personality traits allowed us to generalize the findings across specific assessment methods.

Methods

Sample

The current sample is a subset of the Estonian Biobank cohort (approximately 52,000 individuals), a volunteer-based sample of the Estonian resident adult population (Leitsalu et al., 2014). The participants were recruited randomly by general practitioners (GPs), physicians, or other medical personnel in hospitals or private practices as well as in the recruitment offices of the Estonian Genome Centre of the University of Tartu (EGCUT). Each participant signed an informed consent form, went through a standardized health examination and donated a blood sample for DNA. From among 3,426 individuals for whom both personality and DNA data is available, we selected 3,061 individuals (1,821 women) who were at least 25 years old (mean age 49.54 years, standard deviation 15.49, maximum 91) and had thereby had a chance to complete higher education and obtain a post-graduate degree.

Measures

All but 15 participants completed the Estonian version of the NEO Personality Inventory 3 (NEO PI-3; McCrae & Costa, 2010), which is a slightly modified version of the Revised NEO Personality Inventory. The NEO PI-3 has 240 items that measure 30 personality facets, which are then grouped into the five FFM domains, each including six facets. The items were answered on a five-point scale (0 = *false/strongly disagree* to 4 = *true/strongly agree*). Personality traits of all but 2,904 participants (including the 15 participants with missing self-reports) were also rated by an informant, who was typically a spouse/partner, parent/child or friend. For cross-rater correlations, see Mõttus and colleagues (2014).

Education was based on self-reports and quantified on an eight-level scale: without any formal education (N = 6), lower basic (N = 31), basic (N = 207), high-school (N = 550), vocational plus high-school (N = 956), applied higher (N = 177), higher (N = 967) or post-graduate education (N = 167). The variable was treated as if it was continuous; as will be shown below, its association with EPS was nearly linear.

Polygenic scores aggregate the small effects of a large number of SNPs on a phenotype: the effect size for each SNP's risk allele, found in an independent sample, is multiplied by the number (0, 1, 2) of the allele for a given individual in the target sample; the sums of these products across all SNPs constitute the individual's score. The EPS were based on a meta-analysis (N = 293,723) that estimated the associations of over 8,000,000 SNPs with the number of years of formal schooling (Okbay, Beauchamp, et al., 2016). In the current sample, genotyping was completed using different Illumina platforms (CNV370-Duo BeadChip, OmniExpress BeadChips, HumanCoreExome-11 BeadChips and HumanCoreExome-10 BeadChips); the genotype data was imputed using the 1000 Genomes Project reference panel. SNPs with a minor allele frequency < 0.01, Hardy Weinberg Equilibrium p-value < .001 and info metric value < .90 were omitted. The genotypes were then linkage disequilibrium-pruned using clumping to obtain SNPs in linkage equilibrium with an $r^2 < 0.25$ within a 250 bp window. This was done so that only relatively independent SNPs would be used for calculating EPS. Clumping was carried out based on the subsample of 1,377 participants who had been genotyped using HumanCoreExome platforms in such a way that SNPs with lowest p-values in relation to education (in the meta-analysis) were retained as the index SNPs of the clumps. No p-value cutoff was used for retaining SNPs, so effectively all genetic signal was incorporated into the EPS. These procedures were carried out using PLINK (Purcell et al., 2007). The EPS were based on from 307,346 to 366,188 alleles (i.e., on over 150,000 SNPs), depending on the platform and genotyping success for particular individuals.

In order to provide a frame of reference for the predictive value of EPS for personality traits, we also calculated polygenic scores for Neuroticism (NPS) and Extraversion (EXPS) and predicted the respective traits from these scores. The NPS were based on a meta-analysis of 170,911 individuals (Okbay, Baselmans, et al., 2016) and included contributions from 264,698 to 297,022 alleles, whereas the EXPS were based on a meta-analysis of 63,030 individuals (van den Berg et al., 2016) and included contributions from 274700 to 310700 alleles; procedures were similar to the calculation of EPS. It should be noted that the current sample was part of all three meta-analyses. However, since it comprised less than 2% of the participants for the education and Neuroticism meta-analyses and less than 5% of the participants for the Extraversion meta-analysis, any bias introduced by sample overlap is likely to be very small. Ten principal components representing possible population stratification were calculated based on the genotype data.

Results

Table 1 shows the phenotypic associations between education and personality traits, as well as the associations of the EPS, NPS and EXPS with personality traits and education. The associations are adjusted for age and sex, and associations involving EPS/NPS/EXPS are also adjusted for the number of alleles used in the scores and ten principal components reflecting population stratification. Correlations |.06| or higher are significant at $p \leq .001$ (which is our criteria for highlighting an association as significant; in none of these cases did 99% of confidence intervals span zero), correlations |.07| or higher are significant at $p < .0001$, and correlations |.08| are significant at $p < 2*10^{-5}$. We do not denote statistical significance of individual associations.

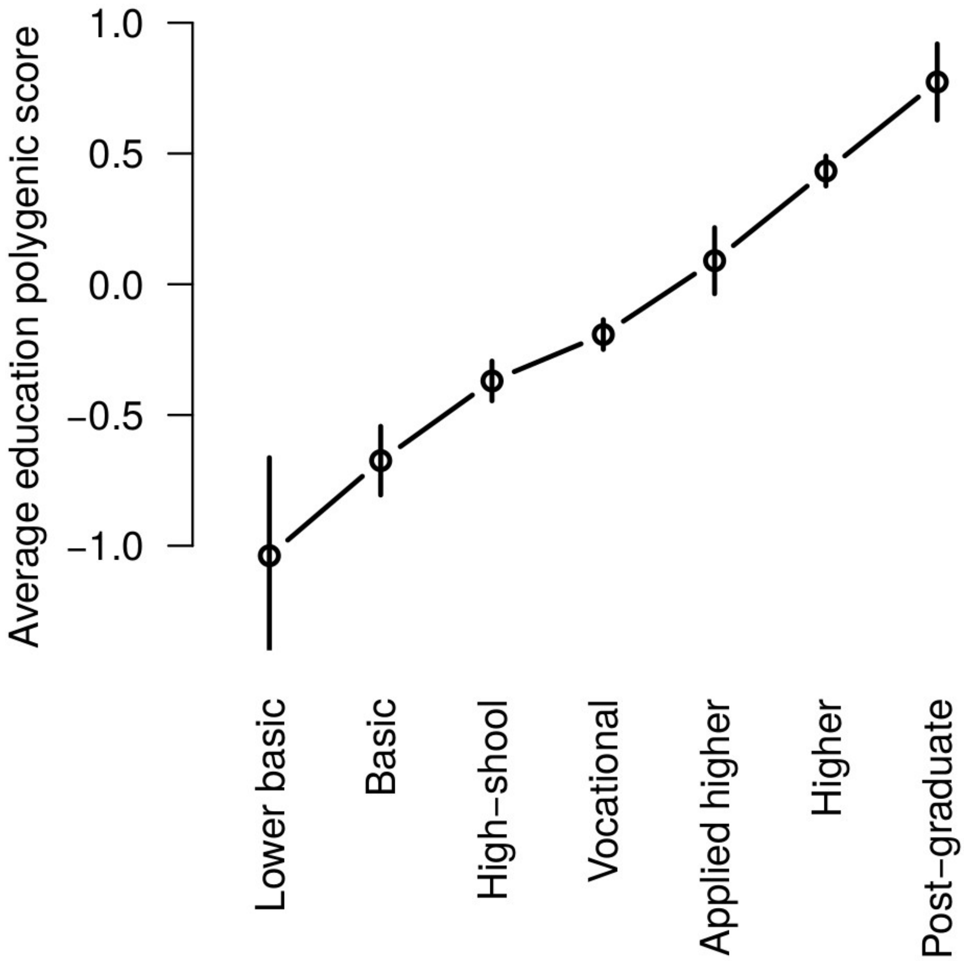
The EPS had a correlation of .41 with its target phenotype, education. Figure 1 shows that the association was nearly linear across the 7 levels of education (average for the six people with no formal education is not shown there). The finding that nearly 17% of variability in education can be traced to a selection of measured common genetic variants is remarkable in itself. To put this estimate in context, general intelligence, which is a phenomenon phenotypically close to education, accounts for 14% and 26% of variance in education, depending on age (Strenze, 2007).

Table 1. *Associations between personality domains and facets with education polygenic propensities for education and Neuroticism.*

	Education		EPS		NPS		EXPS	
	Self	Informant	Self	Informant	Self	Informant	Self	Informant
Neuroticism	-.16	-.17	-.09	-.07	.22	.17	-.07	-.06
Extraversion	.12	.11	.04	.04	-.08	-.08	.20	.15
Openness	.25	.19	.16	.13	-.04	-.02	.26	.15
Agreeableness	-.03	.05	-.02	.04	-.01	-.03	-.02	-.01
Conscientiousness	.07	.13	-.03	.04	-.09	-.06	.05	.04
N1: Anxiety	-.11	-.12	-.07	-.04	.19	.14	-.06	-.05
N2: Hostility	-.17	-.13	-.11	-.08	.18	.13	-.02	-.01
N3: Depression	-.17	-.15	-.10	-.06	.19	.14	-.07	-.06
N4: Self-Consciousness	-.09	-.12	-.05	-.04	.18	.14	-.11	-.09
N5: Impulsiveness	-.03	-.09	-.03	-.05	.12	.10	.03	.00
N6: Vulnerability to Stress	-.14	-.17	-.05	-.06	.16	.13	-.09	-.07
E1: Warmth	.04	.06	.01	.02	-.09	-.06	.17	.12
E2: Gregariousness	.05	.04	.02	.02	-.07	-.07	.12	.09
E3: Assertiveness	.20	.16	.09	.07	-.07	-.07	.15	.14
E4: Activity	.12	.14	.03	.05	-.04	-.06	.17	.13
E5: Excitement Seeking	.03	.03	.00	.00	-.02	-.02	.12	.10
E6: Positive Emotion	.07	.05	.04	.03	-.09	-.06	.17	.10
O1: Openness to Fantasy	.08	.04	.06	.04	.00	.01	.17	.09
O2: Openness to Aesthetics	.17	.12	.11	.09	.00	.00	.21	.12
O3: Openness to Feelings	.10	.07	.05	.02	-.03	.01	.19	.10
O4: Openness to Actions	.23	.18	.13	.10	-.07	-.06	.17	.10
O5: Openness to Ideas	.24	.24	.16	.16	-.03	-.04	.19	.13
O6: Openness to Values	.20	.09	.15	.06	-.04	-.02	.08	.04
A1: Trust	.22	.16	.14	.10	-.08	-.08	.06	.04
A2: Straightforwardness	.01	.07	.03	.05	-.02	-.01	-.05	-.04
A3: Altruism	-.02	.03	-.02	.02	-.05	-.04	.07	.04
A4: Compliance	-.01	.03	.01	.04	-.01	-.05	-.04	-.01
A5: Modesty	-.14	-.04	-.11	-.03	.04	.01	-.10	-.07
A6: Tendermindedness	-.15	-.06	-.12	-.03	.05	.02	.02	.01
C1: Competence	.12	.17	.02	.06	-.13	-.10	.08	.06
C2: Order	.00	.01	-.05	-.03	-.03	-.03	.03	.00
C3: Dutifulness	.06	.11	-.03	.05	-.06	-.04	.03	.02
C4: Achievement Striving	.06	.15	-.02	.04	-.02	-.03	.09	.08
C5: Self-Discipline	.02	.07	-.04	.01	-.07	-.04	.04	.02
C6: Deliberation	.08	.13	.02	.07	-.10	-.07	-.02	-.01
Education			.41		-.09		.07	

NOTE: EPS = Education Polygenic Scores; NPS = Neuroticism Polygenic Scores; EXPS = Extraversion Polygenic Scores; Self = self-ratings; Informant = informant-ratings. All associations at least |0.6| are significant at $p \leq .001$.

Figure 1. Educational levels and average education polygenic scores (means and 95% confidence intervals).



Personality and (polygenic propensity for) education

In both self- and informant-reports, EPS were significantly negatively correlated with the Neuroticism domain and positively correlated with the Openness domain, although the significance did not apply to all of their facets and the associations should therefore not be interpreted at the domain level (Mõttus, 2016; Vainik, Mõttus, Allik, Esko, & Realo, 2015). Specifically, the associations were significant in both rating types for N2: Hostility, N3: Depression, O2: Openness to Aesthetics, O4: Openness to Actions, O5: Openness to Ideas and O6: Openness to Values. The associations were also significant in both rating types for the E3: Assertiveness and A1: Trust facets of the Extraversion and Agreeableness domains, respectively. There were also associations that were only significant in one rating type. For example, people with higher EPS tended to rate themselves lower on A5: Modesty and A6: Tendermindedness, whereas this was not apparent according to informant-ratings, and those higher in EPS were rated as being more competent and deliberate by informant but not by themselves. Facet-level correlations with EPS mirrored the correlations with phenotypic education: the respective columns of Table 1 correlated .94 for both self- and informant-reports. Even the differences between self- and informant-ratings in facet-level correlations with a) EPS and b) education mirrored each other: the correlation between differences in a) and b) was .93.

The effect sizes of personality trait-education associations were small in absolute scale (e.g., .16/.16 and -.11/-.08, respectively for O5: Openness to Ideas and N2: Hostility; self-reports/informant-ratings). However, compared to the phenotypic associations between education and the personality traits they appear more sizable; for example, the respective phenotypic associations of education with O5: Openness to Idea and N2: Hostility were .24/.24 and -.17/-.13 (self-reports/informant-ratings). That is, although the correlations of personality traits with the polygenic propensity for education were lower than their correlations with phenotypic education, for some traits the effect sizes were in the similar order of magnitude. Also, some EPS-personality trait association could be interpreted as rather substantial when compared to the predictive accuracies of NPS and EXP for the very phenotypes these were tailored to, Neuroticism and Extraversion (respectively .22/.17 and .20/.15; self-reports/informant-ratings). For example, while N2: Hostility correlated with EPS -.11/-.08, it correlated with NPS .18/.13 (self-reports/informant-ratings). There are currently no equally powered meta-analyses for other FFM traits or facets, so the predictive accuracies of the polygenic scores for them could not be appropriately pitted against their associations with EPS.

The EPS and NPS were significantly associated (-.11, as opposed to phenotypic correlations of -.16 and -.17, respectively for self- and informant-ratings), also pointing to the genetic overlap between the traits. When the NPS and EPS simultaneously predicted Neuroticism in a multiple regression model, both made significant contributions (respectively .21/.16 and -.07/-.06; self-reports/informant-ratings). In contrast, the correlation between EPS and EXPS was much lower (.04, $p = .03$) and their relations with Extraversion remained identical to univariate estimates in the multiple regression model (Table 1).

As an interesting side-observation, the EXPS correlated with Openness at least as strongly as with its target phenotype Extraversion. This suggest that these traits largely overlap in their additive polygenic influences and whichever biological mechanisms the genetic influences relate to. EXPS was also significantly correlated with education (.07) and NPS (-.24).

Associations with aggregated personality

Separately in self- and informant-ratings, we treated personality facets similarly to SNPs, weighing them by their unique association with education and then aggregating them into a single composite variable—polyfacet scores for education. In order to calculate the weights for each facet, we used least absolute shrinkage and selection operator (LASSO) regression (Tibshirani, 2011) with 50-fold cross-validation and a shrinkage parameter that minimized cross-validated error. By nature, such polyfacet scores captured as much variance in education as could collectively be predicted by the 30 facets and one could therefore think of them as reflecting an education-specific personality trait. We then carried out exactly the same procedure for the EPS, yielding polyfacet scores that were maximally aligned with the polygenic propensity for education. We residualized both polyfacet scores for age and sex.

The correlations between the education polyfacet scores and education itself were .45 and .39, respectively for self- and informant-ratings, suggesting that collectively the facets accounted for between 15% and 20% of variance in education; that is, not much less than intelligence (Strenze, 2007). The correlations between the education polyfacet score and EPS were .28 and .24, respectively for self- and informant-ratings. The correlations between the polyfacet scores for education and polyfacet scores for EPS scores were .91 and .92, respectively for self- and informant-ratings. These high correlations suggest that personality correlates of education tended to be almost entirely related to the additive polygenic variance component of the outcome rather than to any other sources of its variance, pointing to largely overlapping genetic effects.

The plausibility of the three causal scenarios

Finally, in order to assess the plausibility of different causal scenarios, we fitted three structural equation models with the 'lavaan' package (Rosseel, 2012). In the first (common-cause) model, both education (adjusted for age and sex) and the polyfacet scores for education (adjusted for age and sex) were predicted by EPS (adjusted for age, sex, number of alleles used in the scores and population stratification). If the polygenic propensity for education was directly causal to both phenotypes without any mediation of genetic effects between the two phenotypes, education and the polyfacet scores should have been uncorrelated (or locally independent, in latent variable modeling terms), conditional the EPS. Of course, this reasoning was based on the assumption that EPS captured most of the *shared* genetic effects on education and personality, which may well be incorrect. For example, there may be shared non-additive effects or shared effects due to rare variants, which the EPS did not capture. As shown in the top panel of Figure 2, education and its polyfacet scores were still correlated (.38) conditional on EPS (the correlation was .33 based on informant-ratings; Figure 3), pointing to the possibility of one phenotype mediating the genetic effects of the other. This and the following two models were saturated (zero degrees of freedom) and therefore fit data perfectly, hence the models could not be compared in terms of fit and no model fit statistics are reported.

In the second model, the polyfacet scores for education were allowed to mediate the genetic effects of EPS on education. However, as shown in the middle panels of Figures 2 and 3, the direct pathways from EPS to education remained sizable (.32 and .35, respectively in self- and informant-ratings, as opposed to .42 in the common-cause model). In self-ratings, thus, only about 23% and 17% of the total effect (.42) from EPS to education could be accounted for by the indirect pathway via personality traits, respectively in self- and informant-ratings.

Figure 2. Structural equation models in self-reports. Top panel: polygenic scores for education are specified as the common cause of education and education-related personality characteristics (polyfacet scores). Middle panel: the association between polygenic scores for education and education are mediated by education-related personality characteristics (polyfacet scores). Lower panel: education mediates the association between the polygenic scores and education-related personality characteristics (polyfacet scores).

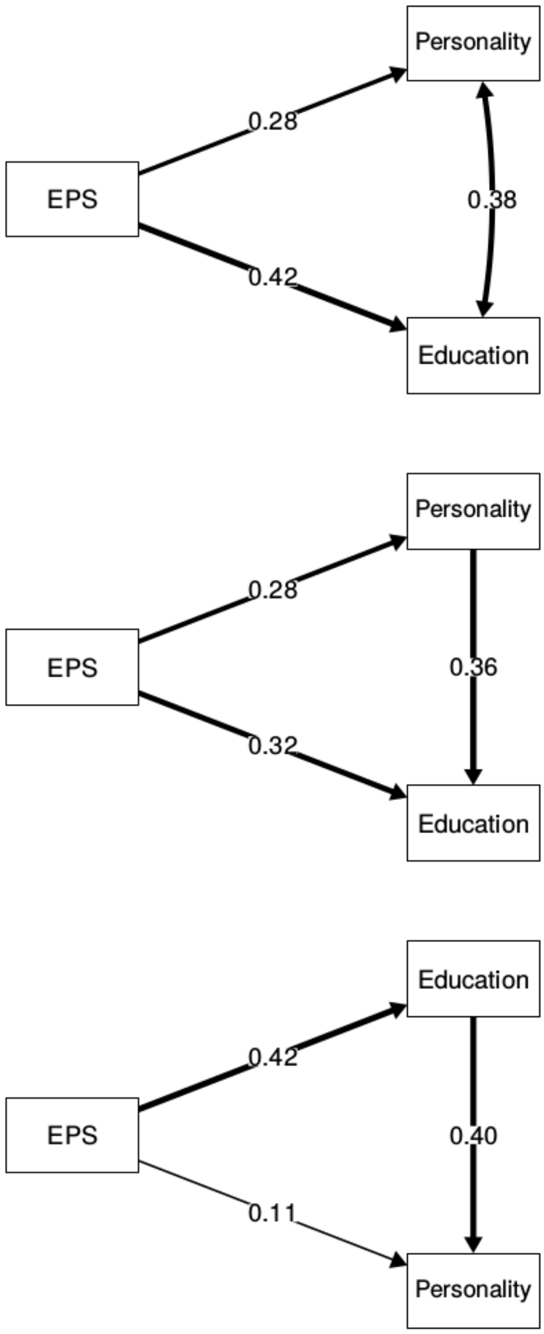
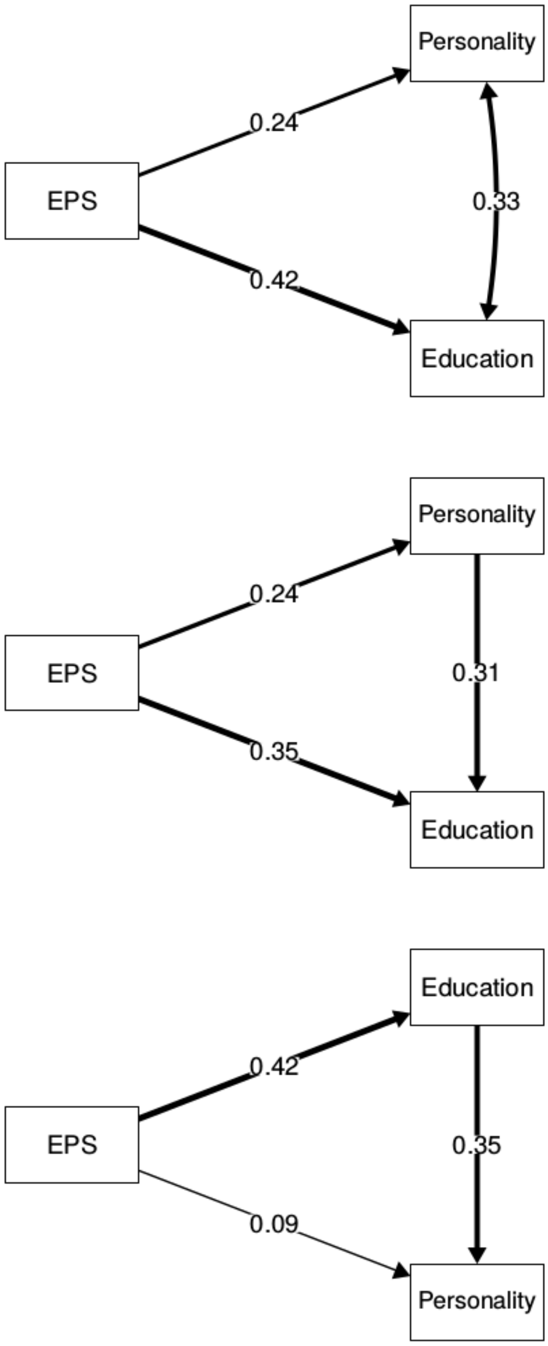


Figure 3. Structural equation models in informant-ratings (see Figure 2 for details).



In the third model, education was specified as mediator between EPS and the polyfacet scores for education (bottom panels of Figures 2 and 3). This model estimated the direct effects from EPS to polyfacet scores at .11 and .09, respectively in self-reports and informant-ratings, which were notably smaller compared to the respective estimates in the common-cause model (.28 and .24). Education could account for about 61% and 63% of the total association between EPS and the composite personality score, respectively in self- and informant-ratings.

Overall, these results lend relatively stronger support to the hypothesis that education mediates some of the genetic effects on personality characteristics than to the hypotheses of polygenic effects being a common cause of education and its related personality traits or personality traits mediating the genetic effects of education. Of course, given that personality traits could also mediate some of the genetic effects on education, the mediation may to some extent work in both ways.

Item-level analyses

We have previously argued that facet- and domain-level analyses should be supplemented with item-level analyses and where there is evidence for item-specificity in the correlations, the associations should not be generalized to aggregate traits (Mõttus, 2016; Vainik et al., 2015). Supplemental Material reports the correlations of single items with EPS (adjusted for age, sex, number of alleles used in the scores and population stratification). For some facets (e.g., O5: Openness to Ideas), the associations seemed to generalize across all of their items, whereas the associations of some facets (e.g., O6: Openness to Values) with EPS appeared to be largely driven by a subset of their items. Occasionally, items of facets that had not been significantly correlated with EPS displayed such associations. For example, although the A2: Straightforwardness facet was not significantly correlated with EPS, its item (A2.4) referring to the belief that honesty is the best policy had a highly significant correlation with polygenic propensity for education (.11 and .12, respectively in self- and informant-ratings). As another example, people with higher EPS were more likely to endorse, and be endorsed for, the item referring to thinking carefully before acting (C6.3), although the C6: Deliberation facet was only significantly associated with the EPS in informant-ratings. Overall, for 65 of the 240 items, the associations with EPS were significant in both self- and informant/ratings ($p < .05$, adjusted for false discovery rate; Benjamini & Hochberg, 1995) and these associations were always in the same direction. The items were distributed across all domains: 23 for Openness, 13 for Neuroticism, 12 for Agreeableness, 11 for Extraversion and 6 for Conscientiousness.

Such findings suggest that at least sometimes the polygenic propensity of education was linked with item's unique variance, or personality nuances (Mõttus, Kandler, Bleidorn, Riemann, & McCrae, in press), as opposed to whatever the items of the same facets share, and that sometimes facets were only associated with "education genes" (and thereby education itself) because some of their items were. This points to etiological heterogeneity within facets, which is consistent with findings such as item-specific developmental trends (Mõttus et al., 2015).

Discussion

The findings showed genetic overlap between education and several personality traits, especially the facets and items of Openness. The associations of personality traits with polygenic scores for education were sometimes not much lower than the associations of personality traits with their own polygenic scores. The aggregated associations of personality traits with education largely reflected shared genetic effects. This may suggest that the same genetic influences act on both education and its associated personality traits. If so, attempts to delineate the genetic underpinnings of education (Okbay, Beauchamp, et al., 2016) may incidentally also reveal the genetic mechanisms of phenotypically related personality characteristics. Also, education could be used as a proxy to narrow the range of potentially personality-related SNPs (Rietveld et al., 2014). An alternative explanation is that particular personality traits partially mediate the genetic variance in education (Rimfeld et al., 2016): these traits may predispose people to seek out more formal education and thereby the genetic influences on these traits can account for some of the genetic variance in this important life-outcome and whatever down-stream consequences it may have.

But the mediation may also work in the opposite direction: experiences related to education may be causal to various personality traits and therefore genetic influences on education can account for some of the genetic variance in these traits. For example, certain genetic variants may predispose people to completing more years of formal schooling (e.g., via faster information processing or better physical health that allows for good school attendance), which in turn enhances interest in aesthetic and intellectual experiences or contributes to disapproval of dishonesty. If so, a better understanding of the genetic etiology of education may, again, eventually help to account for some of the heritability in personality traits and thereby shed light on their etiology. It is also possible that other non-personality traits such as occupational choices or physical fitness can account for some of the heritable personality variance.

We tried to statistically estimate the plausibility of the three scenarios and found the third scenario—mediation through education—to be relatively more plausible than the others. This is consistent with the possibility that at least some of the genetic variance in education-related personality traits does not reflect distinctive genetic and thereby biological mechanisms for these traits (Turkheimer et al., 2014). Until proven otherwise, any attempts to delineate the etiology of these personality traits would need to consider this possibility. Of course, it will ultimately require knowing relevant genetic and biological mechanisms of the phenotypes, or studying groups who have had no exposure to purported mediators, to appropriate disentangle the causal pathways—our findings are only suggestive.

In either case, the genetic overlap will need to be factored into attempts to interpret the phenotypic associations between personality traits and education. Turkheimer and colleagues (2014) argue that when associations of personality traits with other variables are investigated "our scientific hypotheses are usually phenotypic in nature" (p. 533): one phenotype causes the other.

When the phenotypic associations reflect genetic overlap, they may need to be interpreted accordingly. Naturally, the implications of our findings stretch beyond the associations between personality traits and education: genetic overlaps should be considered for *any* other variables that are hypothesized to be either causal to personality traits or among their downstream consequences. For example, personality traits are shown to be phenotypically associated with obesity (Sutin, Ferrucci, Zonderman, & Terracciano, 2011), but these links may at least to some extent reflect genetic overlaps.

With molecular genetic data becoming widely accessible, researchers will be increasingly interested in using them to decompose phenotypic associations into genetic and non-genetic components. The present study highlights one possible methodology for doing this. Although other techniques that allow for estimating genetic correlations from molecular genetic data are available (Bulik-Sullivan et al., 2015; Yang, Lee, Goddard, & Visscher, 2011), they typically require very large samples. Polygenic scores can also be used in smaller samples. This approach does require SNP-outcome associations from an independent large sample but, in the era of genome-wide association studies, such information is becoming available for an ever-increasing number of phenotypes.

In sum, current study systematically examined the polygenic overlap between education and personality traits, and found clear evidence for this. Mediation analysis suggested that education could mediate some of the genetic influences on personality, suggesting that genetic studies of education could also provide useful insights into the genetic underpinnings of personality variability.

Acknowledgements

The study was supported by the University of Tartu (SP1GVARENG) and by institutional research funding (IUT2-13) from the Estonian Ministry of Education and Science to Jüri Allik.

References

Belsky, D. W., Moffitt, T. E., Corcoran, D. L., Domingue, B., Harrington, H., Hogan, S., … Caspi, A. (2016). The genetics of success: How single-nucleotide polymorphisms associated with educational attainment relate to life-course development. *Psychological Science*, 956797616643070. http://doi.org/10.1177/0956797616643070

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300.

Bulik-Sullivan, B., Finucane, H. K., Anttila, V., Gusev, A., Day, F. R., Loh, P.-R., … Neale. (2015). An atlas of genetic correlations across human diseases and traits. *Nature Genetics*, *47*(11), 1236–1241. http://doi.org/10.1038/ng.3406

Chabris, C. F., Lee, J. J., Benjamin, D. J., Beauchamp, J. P., Glaeser, E. L., Borst, G., … Laibson, D. I. (2013). Why it is hard to find genes associated with social science traits: Theoretical and empirical considerations. *American Journal of Public Health*, *103*(S1), S152–S166. http://doi.org/10.2105/AJPH.2013.301327

Chapman, B. P., Fiscella, K., Kawachi, I., & Duberstein, P. R. (2010). Personality, socioeconomic status, and all-cause mortality in the United States. *American Journal of Epidemiology*, *171*(1), 83–92. http://doi.org/10.1093/aje/kwp323

Colodro-Conde, L., Rijsdijk, F., Tornero-Gómez, M. J., Sánchez-Romera, J. F., & Ordoñana, J. R. (2015). Equality in educational policy and the heritability of educational attainment. *PLOS ONE*, *10*(11), e0143796. http://doi.org/10.1371/journal.pone.0143796

Costa, P. T., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, Fl.: Psychological Assessment Resources.

Davey Smith, G. (2010). Mendelian Randomization for strengthening causal inference in observational studies: Application to gene × environment interactions. *Perspectives on Psychological Science*, *5*(5), 527–545. http://doi.org/10.1177/1745691610383505

Davies, G., Armstrong, N., Bis, J. C., Bressler, J., Chouraki, V., Giddaluru, S., … Deary, I. J. (2015). Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949). *Molecular Psychiatry*, *20*(2), 183–192. http://doi.org/10.1038/mp.2014.188

Digman, J. M. (1989). Five robust trait dimensions: development, stability, and utility. *Journal of Personality*, *57*(2), 195–214.

Hicks, B. M., Johnson, W., Iacono, W. G., & McGue, M. (2008). Moderating effects of personality on the genetic and environmental influences of school grades helps to explain sex differences in scholastic achievement. *European Journal of Personality*, *22*(3), 247–268.

Kippersluis, H. van, & Rietveld, C. A. (2016). Pleiotropy-robust Mendelian Randomization. *bioRxiv*, 72603. http://doi.org/10.1101/072603

Leitsalu, L., Haller, T., Esko, T., Tammesoo, M.-L., Alavere, H., Snieder, H., … Metspalu, A. (2014). Cohort Profile: Estonian Biobank of the Estonian Genome Center, University of Tartu. *International Journal of Epidemiology*, *44*, 1137–1147. http://doi.org/10.1093/ije/dyt268

Locke, A. E., Kahali, B., Berndt, S. I., Justice, A. E., Pers, T. H., Day, F. R., … Speliotes, E. K. (2015). Genetic studies of body mass index yield new insights for obesity biology. *Nature*, *518*(7538), 197–206. http://doi.org/10.1038/nature14177

Marioni, R. E., Davies, G., Hayward, C., Liewald, D., Kerr, S. M., Campbell, A., … Deary, I. J. (2014). Molecular genetic contributions to socioeconomic status and intelligence. *Intelligence*, *44*, 26–32. http://doi.org/10.1016/j.intell.2014.02.006

McCrae, R. R., & Costa, P. T. (2010). *NEO Inventories professional manual*. Odessa, FL: Psychological Assessment Resources.

McCrae, R. R., & John, O. P. (1992). An introduction to the five-factor model and its applications. *Journal of Personality*, *60*(2), 175–215. http://doi.org/10.1111/j.1467-6494.1992.tb00970.x

Mõttus, R. (2016). Towards more rigorous personality trait–outcome research. *European Journal of Personality*, *30*(4), 292–303. http://doi.org/10.1002/per.2041

Mõttus, R., Kandler, C., Bleidorn, W., Riemann, R., & McCrae, R. R. (in press). Personality traits below facets: The consensual validity, longitudinal stability, heritability, and utility of personality nuances. *Journal of Personality and Social Psychology*. http://doi.org/10.1037/pspp0000100

Mõttus, R., Marioni, R., & Deary, I. J. (in press). Markers of psychological differences and social and health inequalities: Possible genetic and phenotypic overlaps. *Journal of Personality*. http://doi.org/10.1111/jopy.12220

Mõttus, R., McCrae, R. R., Allik, J., & Realo, A. (2014). Cross-rater agreement on common and specific variance of personality scales and items. *Journal of Research in Personality*, *52*, 47–54. http://doi.org/10.1016/j.jrp.2014.07.005

Mõttus, R., Realo, A., Allik, J., Esko, T., Metspalu, A., & Johnson, W. (2015). Within-trait heterogeneity in age group differences in personality domains and facets: Implications for the development and coherence of personality traits. *PLoS ONE*, *10*(3), e0119667. http://doi.org/10.1371/journal.pone.0119667

Okbay, A., Baselmans, B. M. L., De Neve, J.-E., Turley, P., Nivard, M. G., Fontana, M. A., … Cesarini, D. (2016). Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nature Genetics*, *48*(6), 624–633. http://doi.org/10.1038/ng.3552

Okbay, A., Beauchamp, J. P., Fontana, M. A., Lee, J. J., Pers, T. H., Rietveld, C. A., … Benjamin, D. J. (2016). Genome-wide association study identifies 74 loci associated with educational attainment. *Nature*, *533*(7604), 539–542. http://doi.org/10.1038/nature17671

Purcell, S. (2002). Variance components models for gene-environment interaction in twin analysis. *Twin Research*, *5*(6), 554–571. http://doi.org/10.1375/136905202762342026

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A. R., Bender, D., … Sham, P. C. (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, *81*(3), 559–575. http://doi.org/10.1086/519795

Rietveld, C. A., Esko, T., Davies, G., Pers, T. H., Turley, P., Benyamin, B., … Koellinger, P. D. (2014). Common genetic variants associated with cognitive performance identified using the proxy-phenotype method. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13790–13794. http://doi.org/10.1073/pnas.1404623111

Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology*, No Pagination Specified. http://doi.org/10.1037/pspp0000089

Rosseel, I. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*, 1–36.

Shiner, R. L., Masten, A. S., & Roberts, J. M. (2003). Childhood personality foreshadows adult personality and life outcomes two decades later. *Journal of Personality*, *71*(6), 1145–1170. http://doi.org/10.1111/1467-6494.7106010

Silventoinen, K., Krueger, R. F., Bouchard, T. J., Kaprio, J., & McGue, M. (2004). Heritability of body height and educational attainment in an international context: Comparison of adult twins in Minnesota and Finland. *American Journal of Human Biology*, *16*(5), 544–555. http://doi.org/10.1002/ajhb.20060

Smith, D. J., Escott-Price, V., Davies, G., Bailey, M. E., Colodro Conde, L., Ward, J., … O’Donovan, M. (2016). Genome-wide analysis of over 106,000 individuals identifies 9 neuroticism-associated loci. *bioRxiv*. http://doi.org/10.1101/032417

Strenze, T. (2007). Intelligence and socioeconomic success: A meta-analytic review of longitudinal research. *Intelligence*, *35*(5), 401–426.

Sutin, A. R., Ferrucci, L., Zonderman, A. B., & Terracciano, A. (2011). Personality and obesity across the adult life span. *Journal of Personality and Social Psychology*, *101*(3), 579–592. http://doi.org/10.1037/a0024286

Tibshirani, R. (2011). Regression shrinkage and selection via the lasso: a retrospective. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, *73*(3), 273–282. http://doi.org/10.1111/j.1467-9868.2011.00771.x

Turkheimer, E., Pettersson, E., & Horn, E. E. (2014). A phenotypic null hypothesis for the genetics of personality. *Annual Review of Psychology*, *65*(1), 515–540. http://doi.org/10.1146/annurev-psych-113011-143752

Vainik, U., Mõttus, R., Allik, J., Esko, T., & Realo, A. (2015). Are trait–outcome associations caused by scales or particular items? Example analysis of personality facets and BMI. *European Journal of Personality*, *29*, 622–634. http://doi.org/10.1002/per.2009 van den Berg, S. M., de Moor, M. H. M., Verweij, K. J. H., Krueger, R. F., Luciano, M., Arias Vasquez, A., … Boomsma, D. I. (2016). Meta-analysis of genome-wide association studies for Extraversion: Findings from the Genetics of Personality Consortium. *Behavior Genetics*, *46*, 170–182. http://doi.org/10.1007/s10519-015-9735-5

Vinkhuyzen, A. A. E., Pedersen, N. L., Yang, J., Lee, S. H., Magnusson, P. K. E., Iacono, W. G., … Wray, N. R. (2012). Common SNPs explain some of the variation in the personality dimensions of neuroticism and extraversion. *Translational Psychiatry*, *2*(4), e102. http://doi.org/10.1038/tp.2012.27

Vukasović, T., & Bratko, D. (2015). Heritability of personality: A meta-analysis of behavior genetic studies. *Psychological Bulletin*, *141*(4), 769–785. http://doi.org/10.1037/bul0000017

Yang, J., Lee, S. H., Goddard, M. E., & Visscher, P. M. (2011). GCTA: A Tool for Genome-wide Complex Trait Analysis. *American Journal of Human Genetics*, *88*(1), 76–82. http://doi.org/10.1016/j.ajhg.2010.11.011