Investigating the phenotypic consistency and genetic architecture of noncognitive skills.

Tim T Morris* [1,2], George Davey Smith [1,2], Gerard van Den Berg [3], and Neil M Davies [1,2].

1. MRC Integrative Epidemiology Unit, University of Bristol, BS8 2BN, United Kingdom.

2. Population Health Sciences, Bristol Medical School, University of Bristol, Barley House, Oakfield Grove, Bristol, BS8 2BN, United Kingdom.

3. Department of Economics, University of Bristol, Priory Road Complex, Bristol BS8 1TU, United Kingdom.

Corresponding author: Tim T Morris, MRC Integrative Epidemiology Unit, University of Bristol, Oakfield House, Oakfield Grove, BS8 2BN, UK. Email: Tim.Morris@bristol.ac.uk.

Keywords: noncognitive; cognitive; education; genetics; ALSPAC.

Summary

Noncognitive skills have been demonstrated to predict a range of socioeconomic outcomes including educational attainment and employment. There is however little evidence of how consistent and reliable noncognitive skills are over time or across different measures. Using data from a UK cohort, we show that some key indicators of noncognitive skills are inconsistent, that phenotypic relationships between them are generally weak, and that they associate with educational and labour market outcomes inconsistently and less strongly than cognitive skills. Genomewide analyses reveal that noncognitive skills exhibit low heritability and no clear shared genetic architecture with cognitive skills or outcomes. Our results implicate a high noise to signal ratio and suggest caution in the use of noncognitive measures as reliable indicators of underlying traits. Some noncognitive skills, particularly behavioural difficulties, may provide malleable targets for interventional educational programmes, but many current measurements may be too imprecise and inconsistent to be reliably used.

Noncognitive skills have long been argued as an important mechanism through which a range of outcomes are differentiated between individuals[1–7]. They include characteristics such as persistence, motivation and temperament and are intended to capture a complementary set of skills to cognitive measures such as intelligence and abstract reasoning[1]. However, the definition of noncognitive skills as "personality traits, goals, character, motivations, and preferences"[8] is broad to the point of vagueness; there is no standard scientific definition of noncognitive skills and studies have used a wide range of measures (Box 1). This disparate measurement may reflect that previous studies have used samples or measures of convenience[9]. Modest correlations between different noncognitive skills have been observed for grit and conscientiousness[10]; academic effort and academic problems[11]; and internalising and externalising behaviours[12]. However, the bulk of reported correlations between noncognitive skills have been weak[10,11], questioning any notion that they are one dimensional constructs[13].

Box 1: A non-exhaustive list of noncognitive skills

agreeableness[14]	effort[14]	locus of control[15]	risk aversion[6]
ambition[16]	emotional stability[9]	metacognitive strategies[4]	self-control[6]
attention[17]	emotional intelligence[14]	motivation[18]	self-discipline[19]
behavioural problems[7]	empathy[8]	neuroticism[7]	self-esteem[5]
charm[5]	externalising[20]	openness[8]	self-efficacy[8]
clumsiness[7]	extraversion[7]	optimism[14]	self-perceptions[4]
communication[4]	executive function[21]	organisation[14]	self-regulation[9]
confidence[4]	friendliness[4]	patience[6]	sociability[18]
consistency[1]	goals[18]	perseverance[1]	social skills[17]
conscientiousness[10]	greed[16]	personality[18]	temperament[6]
curiosity[18]	gregariousness[22]	preferences[18]	tenacity[19]
creativity[12]	grit[10]	procrastination[14]	time management[23]
delay of gratification[9]	humility[8]	reliability[19]	trustworthiness[19]
dependability[1]	impulsivity[24]	responsibility[25]	work habits[26]
docility[5]	leadership[3]	resilience/coping[8]	

A large body of research suggests that people with higher measurements of noncognitive skills have better education, health and labour market outcomes[2,3,5,8,11,14,18,25,27]. For example, Borghans and colleagues report that personality explains 16% of the variance in achievement scores in a US sample[27]. This has led some to argue that noncognitive skills are as important as cognitive skills: "For many dimensions of behaviour… noncognitive ability is as important, if not more important, than cognitive ability"[5]. However, the evidence supporting this is diverse, inconsistent and inconclusive. A recent systematic review demonstrated small and heterogenous effects of noncognitive skills on a range of outcomes including academic achievement, with evidence of substantial publication bias (Table 1)[9].There is a lack of robust causal evidence and interventions have generally had only weak and short-lived effects[4].

Table 1: Meta-analysis results from systematic review of the effect of noncognitive skills on a range of outcomes.

	Experimental studies		Observational studies	
	Effect sizes (95% CI)	95% prediction interval	Effect sizes (95% CI)	95% prediction interval
Literacy	0.33 (0.19, 0.47)	-0.13, 0.79	0.16 (0.12, 0.20)	-0.01, 0.33
Numeracy	0.37 (0.16, 0.57)	-0.27, 1.01	0.17 (0.12, 0.21)	-0.04, 0.37
Academic achievement & school readiness	0.16 (-0.02, 0.34)	-1.03, 1.35	0.22 (0.14, 0.31)	-0.07, 0.52
Internalising	0.37 (0.14, 0.60)	-0.34, 1.08	0.15 (0.09, 0.21)	-0.07, 0.37
Externalising	0.23 (0.15, 0.30)	0.07, 0.39	0.21 (0.15, 0.28)	-0.08, 0.51
General behaviour problems	0.32 (0.18, 0.46)	-0.06, 0.70	0.19 (0.12, 0.27)	0.00, 0.39
Social skills	0.46 (0.34, 0.61)	0.01, 0.92	0.13 (0.07, 0.18)	-0.06, 0.31
Psychosocial school readiness	0.42 (0.22, 0.63)	-0.16, 1.00	0.14 (0.06, 0.21)	-0.09, 0.36
Total IQ			0.20 (0.11, 0.30)	
Verbal IQ			0.09 (-0.02, 0.20)	
Performance IQ	0.40 (0.12, 0.69)			
Expressive vocabulary	0.27 (0.01, 0.53)		0.09 (0.04, 0.14)	
Receptive vocabulary	0.49 (0.24, 0.70)		0.16 (0.05, 0.27)	-0.17, 0.50
General language skills	0.37 (-0.25, 0.99)		0.11 (0.04, 0.18)	-0.12, 0.33
General cognitive decline	0.56 (0.16, 0.99)		0.08 (-0.01, 0.17)	

Prediction interval indicates dispersion of effects across the studies that generated the meta-analysis effect size. It estimates where the true effects are to be expected for 95% of similar studies. Note that some effects were inestimable due to low study numbers. Reproduced from [9].

Noncognitive skills have been argued as a valuable target for interventions to improve outcomes because they are more malleable than cognitive skills[3,8,28,29]. Malleability relates to causation; the way that intervention can lead to change. But the extent to which the instruments to detect noncognitive skills are accurately measuring a genuine underlying trait is unknown and it is possible that the perceived malleability of noncognitive skills reflects measurement error or inconsistency. To be ideal candidates for policy interventions, measures of noncognitive skills must reliably pick up a consistent signal of an underlying skill, as displayed by high phenotypic correlations. It has been argued that reliable measures of noncognitive skills are available[8,30], though this has been questioned by a lack of consistent longitudinal evidence[15,31]. Consistency has varied widely between studies and skills; for example, test-retest reliability (calculated as the correlation coefficient between the same measure at two occasions) has been estimated at 0.12[32] and 0.46 to 0.61[33] for risk aversion, and 0.49 for locus of control. These values are far lower than those reported for cognitive skills (c.f. 0.52 for digit span to 0.82 for reading)[32].

The extent to which noncognitive skill measures detect signal over noise can be investigated with genome-wide data. Most human traits are at least partly heritable[34], but there has been relatively little research into the genetics of noncognitive skills compared to cognitive skills[35]. Personality types are the most studied of the noncognitive skills with heritability estimated at between 0.4 and 0.5[37,38], longitudinal variability[36], and evidence of a genetic contribution to their stability over time[39].

Heritability estimates for other noncognitive skills range from 0.36 for alienation to 0.83 for academic effort[11]; 0.31 to 0.56 for aspects of openness[40]; 0.18 to 0.49 for aspects of conscientiousness[41]; and 0.40 for enjoyment and self-perceived ability[42]. Sibling correlations are lower for noncognitive (0.09 to 0.46) than cognitive (0.50 to 0.62) skills[43,44], supporting a smaller genetic component, but the extent of genetic correlation between different noncognitive skills has not been widely investigated[45].

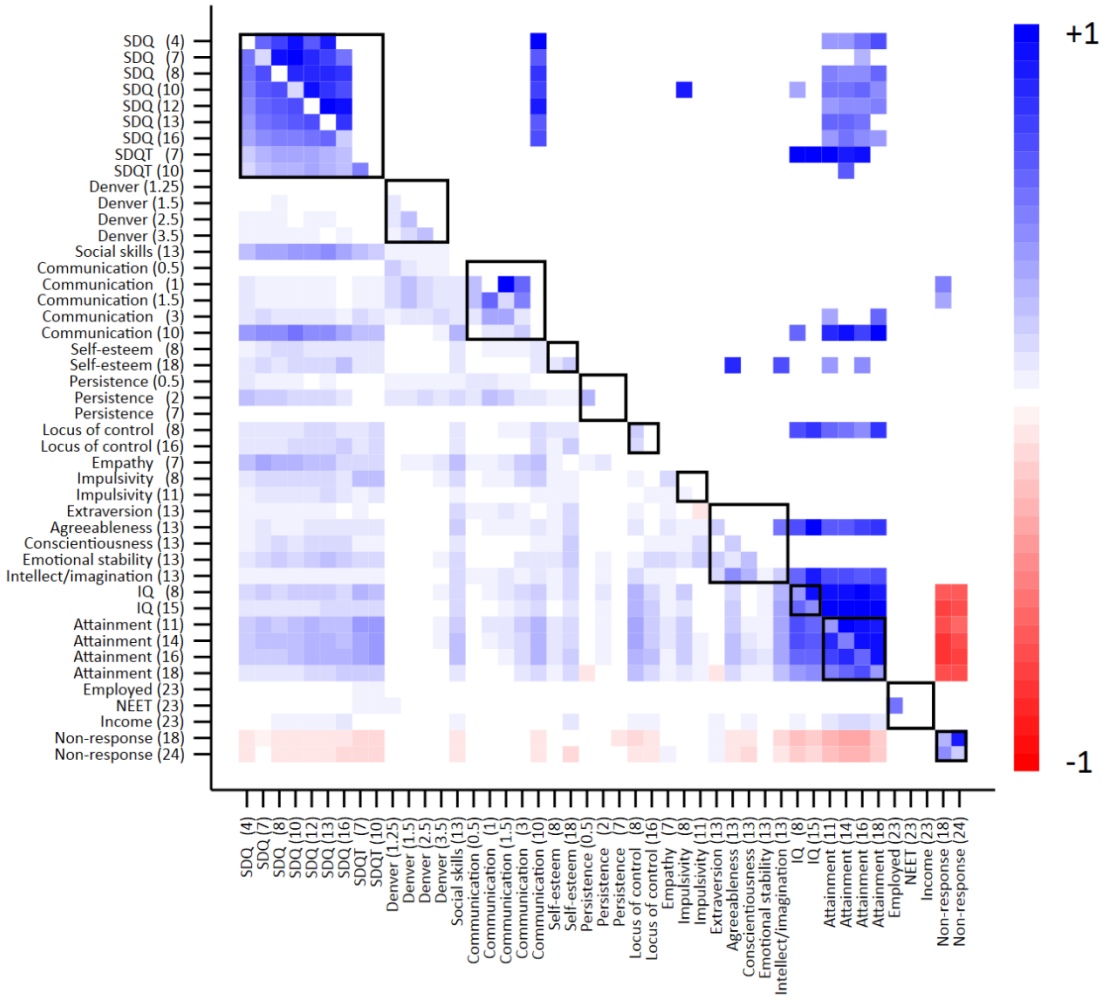
In this study, we contribute to the literature with a comprehensive analysis into the phenotypic and genotypic structure of noncognitive skills. We investigate two related research questions: 1) are noncognitive skills phenotypically and genotypically correlated over time and across different skills? and; 2) Do noncognitive skills associate with socioeconomic outcomes?

Results

How consistent are noncognitive skills over time?

With the exception of the SDQ scale (r=0.18 to 0.73), most noncognitive measures were inconsistent over time (Figure 1). This suggests that most skills are highly variable (inconsistent) or poorly measured (unreliable). Correlations amongst the Big 5 personality types were low except between the intellect/imagination and agreeableness subscales (r=0.46). While the phenotypic correlations for repeat measures of noncognitive skills were low, they were generally positive where distinguishable from zero and so do not appear to capture conflicting information at different measurement occasions. The weak temporal phenotypic correlations of noncognitive measures contrasts sharply to those for cognitive measures of IQ (r=0.60) and measures of educational attainment (r>0.78 for compulsory education). The high correlation of the education measures will have been partly socially organised by the streaming of children in schools.

Figure 1: Heritability, phenotypic correlations and genetic correlations of skills and outcomes



Age in parentheses. SDQ, strengths and difficulties; NEET, not in education, employed or training. Values on the diagonal represent univariate heritabilities; values below the diagonals represent phenotypic correlations; values above the diagonal represent genetic correlations. Multiple testing was handled using an FDR threshold of 5%. Empty cells display correlations that fell below FDR threshold. Black outline boxes indicate the same skills measured at different occasions.

How correlated are different noncognitive skills?

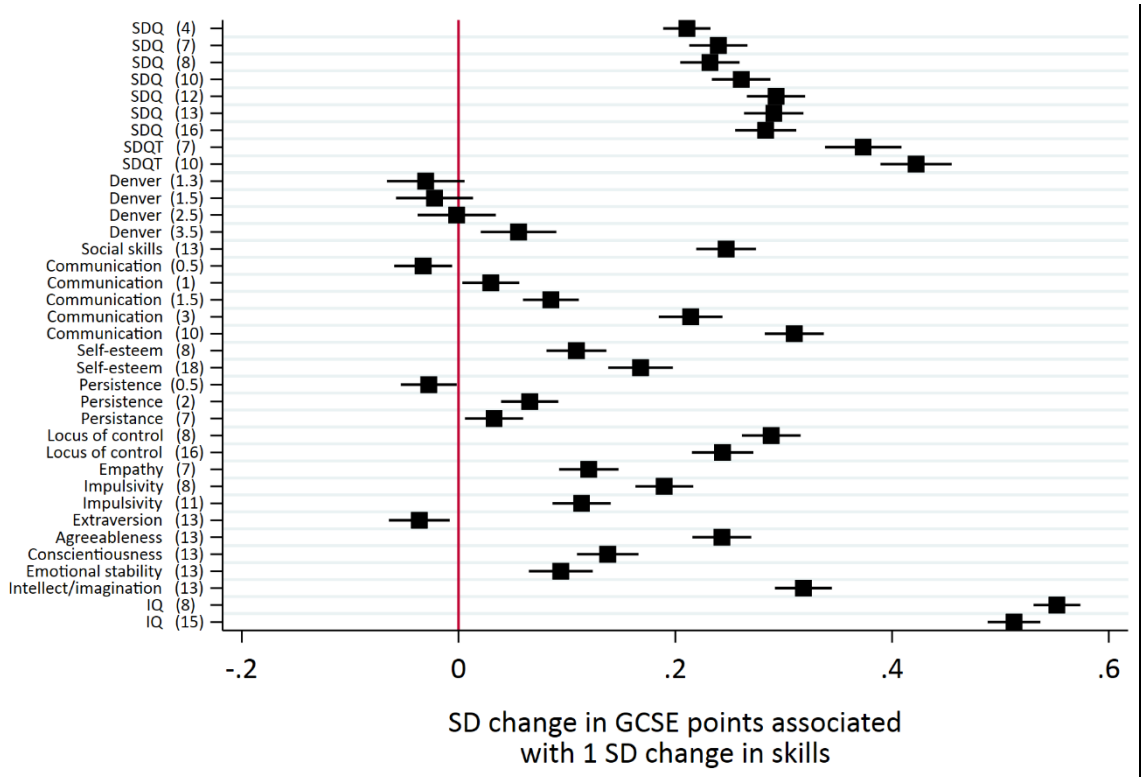
Phenotypic correlations between different measures of noncognitive skills are mostly very low (Figure 1). This suggests that these different measures are capturing empirically distinct phenomena. The SDQ scale is the only measure that correlates consistently with other noncognitive skills at r>0.2. This between-trait correlation is strongest for social skills (r =0.24 to 0.49), communication at age 10 (r =0.29 to 0.56) and empathy (r =0.17 to 0.38). Between-skill phenotypic correlations are almost exclusively positive, suggesting that where patterns are observed, children who score highly on one noncognitive skill are also more likely to score highly on another. All noncognitive skills except for the teacher reported measures of SDQ, locus of control at age 8 and the intellect/imagination personality type correlate weakly with cognitive skills (r<0.25 for IQ at age 8 and r<0.20 for IQ at age 15). The highest phenotypic correlation is between IQ and the intellect/imagination personality type (r=0.33 for IQ at age 8 and r=0.35 for IQ at age 15).

Do noncognitive skills associate with education and labour market outcomes?

Phenotypic correlations with educational attainment were low for most noncognitive skills. Correlations were highest for the SDQ scale (r=0.13 to 0.42) and the teachers' assessments were more strongly associated than the contemporaneous parent reports. For example, the correlation between educational attainment at age 11 and teacher reported SDQ at age 10 was r=0.40 while the correlation with parent reported SDQ at age 10 was r=0.29. Correlations with educational outcomes were modest for social skills (r=0.13 to 0.27); communication at age 10 (r=0.18 to 0.34), locus of control (r=0.26 to 0.36 at age 8; r=0.19 to 0.28 at age 16), the agreeableness and the intellect/imagination subscales of the Big 5 (r=0.16 to 0.28 and r=0.27 to 0.39 respectively). These contrasted strongly with the higher correlations between cognitive ability and educational attainment (r=0.43 to 0.73). Phenotypic correlations for noncognitive skills with employment and NEET were very low (r<=0.08 for employment; r<=0.07 for NEET) but were slightly higher between some measures and income (r= -0.01 to 0.19). Correlations between labour market outcomes and cognitive skills were also very low (r<=0.09). Many noncognitive skills were weakly negatively correlated with non-response at ages 18 and 24 (r=0.05 to 0.15). Only the extraversion subscale of the Big 5 was positively correlated with non-response (r=0.06).

To further investigate the potential impact of skills we ran a series of regressions of age 16 educational attainment on noncognitive and cognitive skills (Figure 2). Each skill was standardised and analysed independently controlling for sex and month of birth. A one standard deviation (SD) increase in noncognitive skills was associated with a 0.04 SD decrease to 0.41 SD increase in age 16 attainment. There was considerable heterogeneity for estimates between skills and measurement occasions for all noncognitive measures except the SDQ, which was similarly consistent to IQ. By comparison, a one SD increase in cognitive skills was consistently associated with a 0.5 SD increase. These patterns were similar for educational attainment at all ages (Supplementary Figure 1).

Figure 2: Associations between skills and educational attainment at age 16

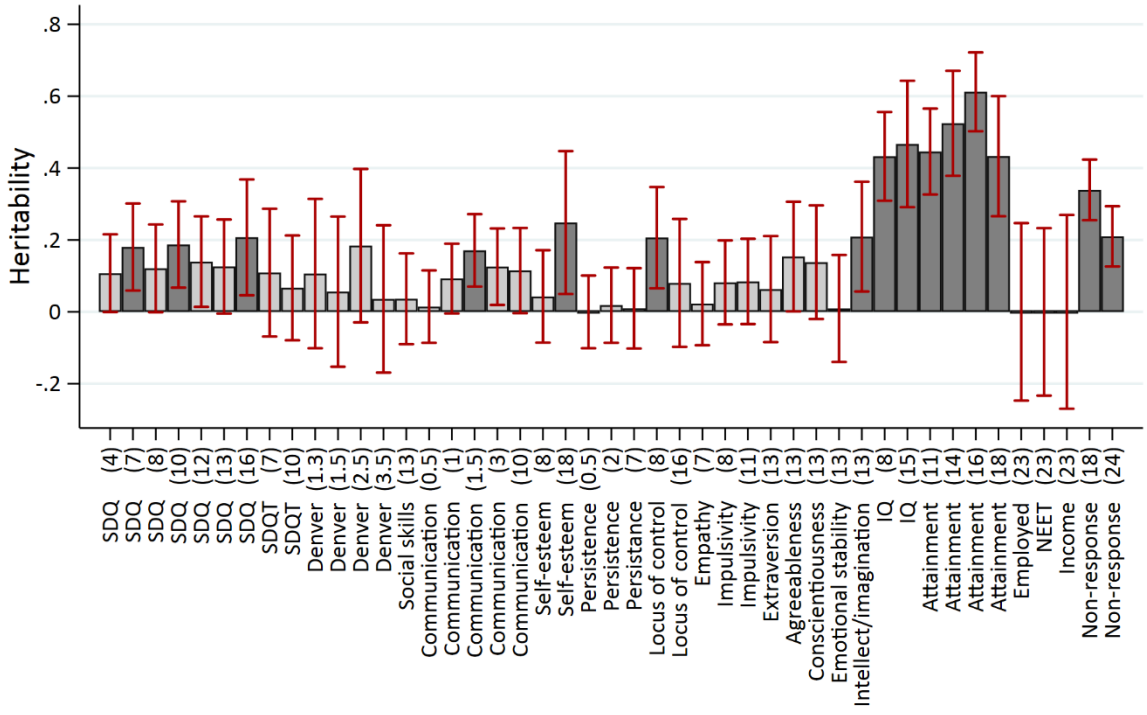


Age (years) in parentheses. SDQ, strengths and difficulties

Are noncognitive skills heritable?

Figure 3 displays the heritability estimates for all our measures (estimates and sample sizes in Supplementary Table 2). The SDQ scale is the only noncognitive measure for which there was non-zero heritability at multiple occasions (h^2_{SNP}=0.18 to 0.23). Results are broadly comparable when using the internalising and externalising subscales of the SDQ (Supplementary Figure 2). Non-zero heritability is observed for communication at age 18 months (h^2_{SNP}=0.17), self-esteem at age 18 (h^2_{SNP}=0.25), locus of control at age 8 (h^2_{SNP}=0.21), and the 'intellect/imagination' subscale of personality type (h^2_{SNP}=0.21). Imprecision of the heritability point estimates means that non-zero heritabilties cannot be ruled out, but the upper bounds are estimated below 0.30 for most noncognitive skills. In contrast, the heritability of cognitive skills is far higher (h^2_{SNP}=0.43 and 0.47 at ages 8 and 15 respectively). Educational outcomes are highly heritable (h^2_{SNP}>0.4) but there is little evidence of heritability for labour market outcomes. Heritability of questionnaire non-response was estimated higher and with less uncertainty than the noncognitive measures (h^2_{SNP}=0.34 and h^2_{SNP}=0.21 at ages 18 and 24 respectively).

Figure 3: Heritability of skills and outcomes



Age in parentheses. SDQ, strengths and difficulties; NEET, not in education, employed or training. Dark coloured bars represent estimates below FDR threshold. See Supplementary Table 2 for full results.

Do noncognitive skills have a shared genetic architecture?

There was very limited evidence for genetic correlations across noncognitive skills (Figure 1, above the diagonal), though estimation precision is low for most skills. Genetic correlations within traits over time are only observed for the parent reported SDQ measures (rG= 0.62 to 1.00) and cognitive

skills which have near-unity genetic overlap (rG= 0.97). Genetic correlations between different noncognitive measures are only observed between the parent, but not teacher, reported SDQ measures and communication at age 10 (rG=0.68 to 0.91). The SDQ measures, communication at age 10, self-esteem at age 18, the agreeableness and the intellect/imagination subscales of the Big 5 personality types had strong genetic correlations with educational attainment. These were higher for teacher report at age 7 than any of the parents and teacher reported at age 16 SDQ measures. Genetic correlations between cognitive measures and educational attainment are all estimated near unity (rG>0.96). There was little evidence of genetic correlations between labour market outcomes and any of the noncognitive or cognitive skills.

Discussion

Our results question the notion that noncognitive skills are consistent, reliable or highly predictive of outcomes[8], and provide two key contributions to the literature. First, except for the SDQ scale, measures of noncognitive skills were weakly correlated (phenotypically and genotypically) over time. This contrasts with previous research that has demonstrated temporal stability of noncognitive skills[15,28], and strongly cautions that many measures of noncognitive skills fail to capture a consistent underlying trait. Correlation of non-cognitive skills across different measures were also weak. The SDQ scale was also the only measure to consistently correlate with other noncognitive skills, namely social skills, communication and empathy. These results conform to a previous study that found low phenotypic correlations between different noncognitive skills[11]. This low between-trait consistency could reflect that different noncognitive measures capture fundamentally different underlying traits, but the lack of within-trait consistency over time suggests that this is unlikely. Our results also question previous suggestions that cognitive skills reflect noncognitive skills[3] as we observed only weak correlations between noncognitive and cognitive measures. By contrast, the cognitive measures of IQ were far more strongly correlated over the same period. There was very limited evidence of genetic correlations within and between noncognitive skills. The only measure for which we observed consistent genetic architecture was the parent reported SDQ scale. This could reflect the influence of shared parent-offspring genetics on reporting or parents genetics influencing offspring noncognitive skills indirectly through dynastic effects[46]; there would be no such shared teacher-student genetics. Previous twin studies have found strong genetic correlations between noncognitive and cognitive skills[40,41], but our results did not support this. The only noncognitive measures that were genetically correlated with cognitive measures were the teacher reported SDQ scale, locus of control, and the agreeableness and intellect subscales of the Big 5. The differences between these results and previous studies may have arisen due to several reasons. First, the noncognitive skill measures in ALSPAC may have been of lower quality than those used in previous studies. However, most of the measures we used have been widely validated and are consistent with those used in previous studies so this is unlikely to account for all the differences. Second, we used longitudinal measures that spanned the whole of childhood whereas previous studies used cross sectional measures or single repeat measurements. However, there was no age at which all noncognitive skills appeared more consistently measured and this is therefore also unlikely to account for all differences. Third, it is possible that the mothers who reported on many of the noncognitive skills were providing biased responses in cases where their child performed poorly. Again, while possible this is unlikely to have accounted for all the differences between this and previous studies. Fourth, it is possible that the ALSPAC cohort are fundamentally different to previously analysed study populations. Finally, it is possible that results from previous studies have

been due to chance or selectively reported. While we reported every noncognitive skill in ALSPAC, many previous studies have reported only one or a small number of noncognitive skills.

Second, associations between noncognitive skills and socioeconomic outcomes were generally weak, contradicting findings from previous studies[3,8,18] and supporting a recent systematic review that highlighted the uncertainty around the impact of these skills on socioeconomic outcomes[9]. That we were able to use multiple measurements of noncognitive skills at different ages and a range of outcomes strengthens our findings. Behavioural problems as captured by the SDQ scale, social skills, and locus of control were the only noncognitive measures to phenotypically associate with educational outcomes strongly and consistently, findings that both support and contradict previous studies[47,48]. Neither noncognitive nor cognitive skills associated strongly with labour market outcomes, which while in contrast to previous studies[3,8]. It has been previously claimed that personality is more important than cognitive ability for predicting socioeconomic outcomes such as education[49], but our results do not support this: the effect sizes of noncognitive skills were at most around half that of cognitive skills. This greater contribution to educational attainment of cognitive than noncognitive skills may be due to several reasons. First and foremost, this may reflect that our study participants had only recently entered the labour market. Second, cognitive skills may fundamentally be a more important driver of educational differences between children than noncognitive skills. UK educational attainment is heavily assessed by performance in tests and exams (which are highly reflective of cognitive test environments). Third, cognitive tests have a long history of development and are therefore likely to be far more accurate and consistent than noncognitive measures. Fourth, cognitive skill assessment is likely to be more objective than noncognitive skill assessments, which can be context dependent. Finally, our cognitive measures were based upon direct assessment of the study children while the noncognitive measures were based upon parental reports. Further work using self-reported noncognitive skills is required to examine the full impact of this. Associations were stronger between educational attainment and the teacher than the parent reported measures of the SDQ scale, suggesting that teachers may more accurately identify education-associated factors related such as problematic behaviour than parents[50]. Many noncognitive and cognitive skills were weakly negatively correlated with non-response phenotypically, suggesting that individuals who score low on these skills are more likely to drop out of studies. Cognitive ability and attainment were strongly negatively correlated genotypically with non-response, adding to the growing body of evidence that non-response is genetically patterned[51,52]. This may have important implications for participant representativeness and generalisability in cohort studies. Each of the SDQ scale, communication, locus of control (age 8), and the agreeableness and intellect/imagination subscales of the Big 5 personality types were strongly genetically correlated with educational attainment. Genetic correlations with educational attainment were higher for teacher than parent reported measures of the SDQ scale. This may reflect that teachers can more accurately identify and report problem behaviours than parents, or that there are different aspects of child behaviour and performance that they focus on. Many of the genetic correlations we observed were imprecise due to the low heritability of noncognitive skills, but our estimates suggested an upper bound heritability of 0.3 for most skills. The SDQ scale was the only noncognitive measure for which we consistently estimated non-zero heritability (~0.15), far lower than the estimated heritabilties for cognitive measures (~0.45) and educational attainment (~0.50).

Our results suggest that while many of these measures of noncognitive skills would be poor intervention targets for policies to improve educational or labour market outcomes, interventions to improve behavioural problems as captured by the SDQ scale may have merit. Most of the noncognitive measures we used were poorly correlated over time suggesting a poor signal to noise

ratio characterised by measurement error. The SDQ scale uses responses to a large battery of questions which may lead to proportionately smaller measurement error compared to the signal of the underlying trait that it captures. If this is the case, then future research requires more detailed measures of non-cognitive skills are required rather than opportunistic single item measures as have often been used. The lack of consistency observed for other skills may also have reflected other mechanisms including i) genuine temporal intra-individual variation amongst study participants in the expression of these skills; ii) differences over time (e.g. due to schooling); or iii) different responders (e.g. parent vs. teacher). Regardless of the mechanism though, the inconsistency suggests that many measures are likely too noisy and variable for intervention as these measures are imprecise indicators of underlying traits. It is of course possible that there are latent underlying noncognitive skills which are not accurately captured by the measures we used, but our findings suggest that noncognitive skills may be more heterogenous than is widely appreciated.

This study has several limitations. First, it is possible that measurement error was unusually high in the noncognitive measures used in the ALSPAC study. However, the measures used in ALSPAC have been widely used across the literature and this has not previously been identified as a problem[12,53]. Furthermore, measurement error would need to have been high across all measures used from birth to age 18 so is unlikely to explain our results. Future studies into test-retest reliability of noncognitive skills based on different longitudinal samples could help resolve these questions. Second, many of the genetic correlations were estimated with extremely low precision, often being constrained at the values of -1 or 1 (see supplementary Table 2), and should therefore be interpreted with caution. This may be due to the low estimated heritability of the noncognitive skills; low heritability implies a small contribution of variants (either in the number of variants associated or the strength of associations) and therefore lower power to detect genetic correlations between these skills. It may also be due to our fairly low sample sizes (n=2,545 to 8,868) and the resulting power to detect univariate and bivariate genetic associations. Future studies conducted on larger samples are required to more accurately estimate heritability of, and genetic correlations between noncognitive skills and other phenotypes. Third, our genetic associations may have been biased by uneven linkage disequilibrium, residual population structure, or assortative mating[46,54]. We controlled for the first twenty principle components of population structure to account for population structure, however we this may not have accounted for all differences[55]. While assortative mating is thought to be low for noncognitive traits[38], there is evidence of assortment on psychiatric traits such as ADHD[56]. This may have biased our estimates, but previous work demonstrates that this should inflate rather than deflate genetic associations where assortment is positive[57]. It is possible that assortment on non-cognitive skills may be negative and future work is required to determine this. Previous work using ALSPAC has demonstrated that the heritability of cognitive variables and educational outcomes are estimated higher than in other samples and we may therefore expect that our heritability of noncognitive skills is unlikely to be underestimated.

In conclusion, many noncognitive skills are poorly correlated over time and associate weakly with education and labour market outcomes. The SDQ scale is an exception, providing a noncognitive measure that is reliable and associates with educational outcomes albeit less strongly than cognitive skills. While the variability of many noncognitive skills over time questions their suitability as interventional targets[8], interventions based on behavioural problems in childhood may have utility for improving educational outcomes.

Methods

Study sample

Participants were children from the Avon Longitudinal Study of Parents and Children (ALSPAC). Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st December 1992 were invited to take part in the study. The initial number of pregnancies enrolled was 14,541. When the oldest children were approximately 7 years of age, an attempt was made to bolster the initial sample with eligible cases who had failed to join the study originally. This additional recruitment resulted in a total sample of 15,247 pregnancies, resulting in 14,899 children who were alive at one year of age. From this sample genetic data is available for 7,988 after quality control and removal of related individuals (see *Genetic data* below). For full details of the cohort profile and study design see [58,59]. The study website contains details of all the data that is available through a fully searchable data dictionary and variable search tool at http://www.bristol.ac.uk/alspac/researchers/our-data/. The ALSPAC cohort is largely representative of the UK population when compared with 1991 Census data; there is under representation of some ethnic minorities, single parent families, and those living in rented accommodation[58]. Ethical approval for the study was obtained from the ALSPAC Ethics and Law Committee and the Local Research Ethics Committees. We used the largest available samples in each of our analyses to increase precision of estimates, regardless of whether a child has data on other noncognitive skills. Sample sizes range from 2,545 to 8,868 (Supplementary Table 1).

Genetic data

DNA of the ALSPAC children was extracted from blood, cell line and mouthwash samples, then genotyped using references panels and subjected to standard quality control approaches. ALSPAC children were genotyped using the Illumina HumanHap550 quad chip genotyping platforms by 23andme subcontracting the Wellcome Trust Sanger Institute, Cambridge, UK and the Laboratory Corporation of America, Burlington, NC, US. The resulting raw genome-wide data were subjected to standard quality control methods. Individuals were excluded on the basis of gender mismatches; minimal or excessive heterozygosity; disproportionate levels of individual missingness (>3%) and insufficient sample replication (IBD < 0.8). Population stratification was assessed by multidimensional scaling analysis and compared with Hapmap II (release 22) European descent (CEU), Han Chinese, Japanese and Yoruba reference populations; all individuals with non-European ancestry were removed. SNPs with a minor allele frequency of < 1%, a call rate of < 95% or evidence for violations of Hardy-Weinberg equilibrium (P < 5E-7) were removed. Cryptic relatedness was measured as proportion of identity by descent (IBD > 0.1). Related subjects that passed all other quality control thresholds were retained during subsequent phasing and imputation. 9,115 subjects and 500,527 SNPs passed these quality control filters.

ALSPAC mothers were genotyped using the Illumina human660W-quad array at Centre National de Génotypage (CNG) and genotypes were called with Illumina GenomeStudio. PLINK (v1.07) was used to carry out quality control measures on an initial set of 10,015 subjects and 557,124 directly genotyped SNPs. SNPs were removed if they displayed more than 5% missingness or a Hardy-Weinberg equilibrium P value of less than 1.0e-06. Additionally, SNPs with a minor allele frequency of less than 1% were removed. Samples were excluded if they displayed more than 5% missingness, had indeterminate X chromosome heterozygosity or extreme autosomal heterozygosity. Samples showing evidence of population stratification were identified by multidimensional scaling of

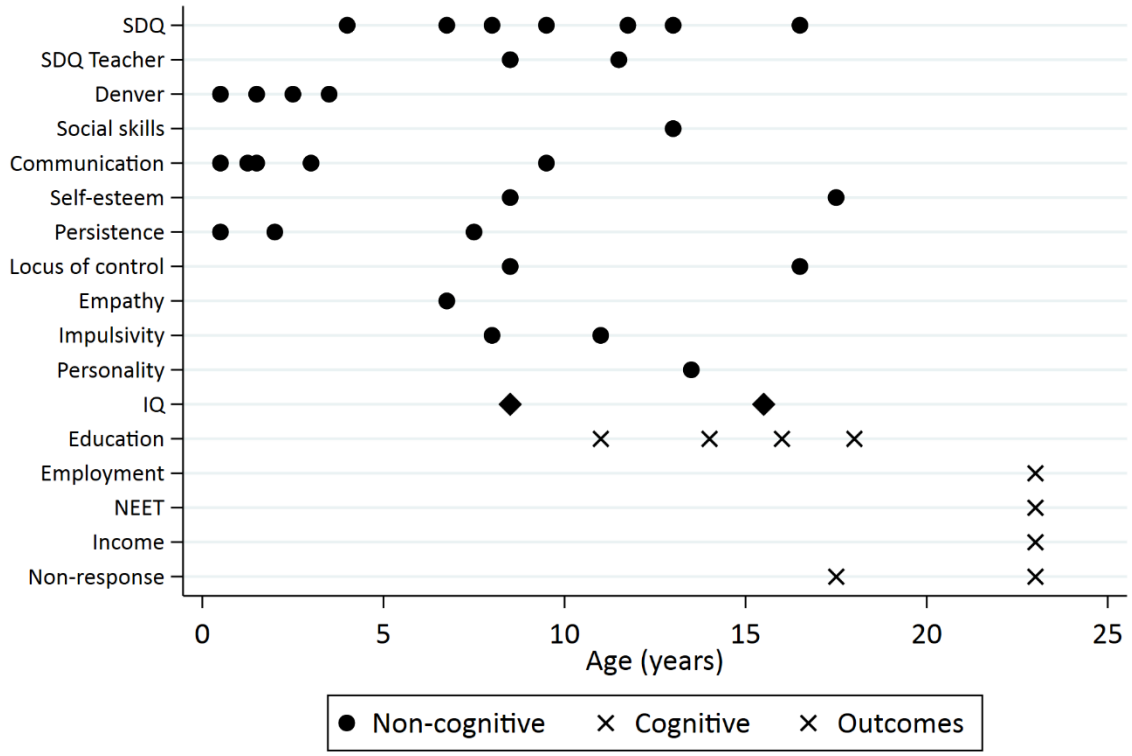
genome-wide identity by state pairwise distances using the four HapMap populations as a reference, and then excluded. Cryptic relatedness was assessed using a IBD estimate of more than 0.125 which is expected to correspond to roughly 12.5% alleles shared IBD or a relatedness at the first cousin level. Related subjects that passed all other quality control thresholds were retained during subsequent phasing and imputation. 9,048 subjects and 526,688 SNPs passed these quality control filters.

We combined 477,482 SNP genotypes in common between the sample of mothers and sample of children. We removed SNPs with genotype missingness above 1% due to poor quality (11,396 SNPs removed) and removed a further 321 subjects due to potential ID mismatches. This resulted in a dataset of 17,842 subjects containing 6,305 duos and 465,740 SNPs (112 were removed during liftover and 234 were out of HWE after combination). We estimated haplotypes using ShapeIT (v2.r644) which utilises relatedness during phasing. The phased haplotypes were then imputed to the Haplotype Reference Consortium (HRCr1.1, 2016) panel of approximately 31,000 phased whole genomes. The HRC panel was phased using ShapeIt v2, and the imputation was performed using the Michigan imputation server. This gave 8,237 eligible children and 8,196 eligible mothers with available genotype data after exclusion of related subjects using cryptic relatedness measures described previously. Principal components were generated by extracting unrelated individuals (IBS < 0.05) and independent SNPs with long range LD regions removed, and then calculating using the `--pca` command in plink1.90.

Noncognitive skills

We used all noncognitive skills outlined in Box 1 that are available in ALSPAC, except for attention which was omitted due to low frequency of events (<5%). Figure 4 displays a timeline of when all skills and outcomes were measured in ALSPAC. The supplementary material contains more detailed information on the measures used and their constituent components.

Figure 4: Timeline of measures in ALSPAC



SDQ, strengths and difficulties; NEET, not in education, employed or training

SDQ

Study mothers reported on the Strengths and Difficulties Questionnaire (SDQ) for children on seven occasions at child ages 4, 7, 8, 10, 12, 13 and 16 years, and the children's teachers completed an SDQ questionnaire for each child at ages 7 and 10. The SDQ is a scale used to assess child emotional and behavioural difficulties[60] and is one of the most widely used questionnaires for evaluating psychological well-being amongst children. It consists of 25 items that cover common areas of emotional and behavioural difficulties (emotional symptoms; conduct problems; hyperactivity/inattention; peer relationship problems; and prosocial behaviour). Responses to each question are on a three-point scale: not true, somewhat true and certainly true, coded to scores of 0, 1 and 2 respectively. The individual subscales of the SDQ have relatively low-prevalence in ALSPAC so we used total SDQ score which is defined as the count of problems on the first four scales. To ensure that our results are not being driven by differences in the internalising (emotional symptoms and peer relationship problems) or externalising (conduct problems and hyperactivity/inattention symptoms) subscales we also ran sensitivity analyses on these separate sub-scales. All SDQ scores are reverse coded so that high values refer to fewer problems.

Denver

The Denver Developmental Screening Test[61] was used to identify developmental problems in young children at ages 6, 18, 30 and 42 months. ALSPAC mothers were asked to report their child's development in response to 42 questions across four different categories: social and communication skills, fine motor skills, hearing and speech, and gross motor skills. Responses to questions were 'often', 'once or twice' and 'not yet started', and were coded with the values of 2, 1 and 0

respectively. Prorated scores combining all four scales were used to boost sample size with missing values assigned the mean score of that child's responses, provided that three or less items had missing scores.

Social skills

Social skills at age 13 were determined using a battery of 10 questions reported by the study mother, such as being "easy to chat with, even if it isn't on a topic that specially interests her". Responses were reported on a five-point scale indicating the mother's perception of how well her teenager's social skills compared to her perception of the average teenager, and then summed to provide a total overall social skills score.

Communication

Communication at 6 months was calculated from mother-reported responses to a battery of eight questions asking about the development of their child's communication skills such as "S/he turns towards someone when they are speaking". At age 1 communication was calculated from mother-reported responses using the MacArthur Infant Communication questionnaire[62] was derived at age one as the sum of mother reported responses to 82 questions across four domains of communication: understanding; vocabulary; non-verbal communication; and social development. At 18 months communication was calculated using mother-reported responses to a battery of 14 questions asking about the development of their child's communication skills. At age 3 communication was calculated from mother-reported responses to a battery of 123 questions forming a vocabulary score. At age 10 communication was calculated from mother-reported responses as the sum of five domains of communication from a total battery of 39 questions. Responses were coded as 2, 1 and 0 reflecting the level of communication and understanding of communication that the child demonstrated. A final communication score was created as the sum of these responses.

Self-esteem

Self-esteem at age eight was measured using self-report responses to the 12-item shortened form of Harter's Self Perception Profile for Children[63] comprising the global self-worth and scholastic competence subscales. Self-esteem at age 18 was measured using self-report responses to 10 questions of the Bachman revision of the Rosenberg Self-Esteem Scale[64,65]. At each age responses are summed to give an overall score. Responses are coded on a Likert scale with the value 0 assigned to the most negative statements and the value 4 assigned to the most positive statement, so that high scores correspond to high self-esteem.

Persistence

Persistence at age 6 months was measured as a weighted score from mother-reported responses to seven questions relating to child temperament. At age 2 persistence was measured as a weighted score from nine mother-reported responses. Questions included items such as "He perseveres for many minutes when working on a new skill (rolling over, picking up object, etc)" and "He loses interest in a new toy or game within an hour" and were recorded on a five-point Likert scale. At age 7 persistence was recorded by ALSPAC interview testers as the study child's persistence when completing the word reading and decoding session. In this session children were shown a series of pictures accompanied by four words starting with the same letter as the picture and were asked to point to the correct word. Response options were coded into persistent (combining persistent and sometimes persistent) and non-persistent. Testers indicated whether they thought the child was

persistent with tasks during the session. On all measures higher scores correspond to higher levels of persistence.

Locus of control

Locus of control, the strength of connection between actions and consequences, was measured at ages 8 and 16. At age eight it was measured using responses to 12 questions from the shortened version of the Nowicki-Strickland Internal-External (NSIE) scale for preschool and primary children (the non-cartoon format Preschool and Primary Nowicki-Strickland Internal-External scale)[66]. At age 16 locus of control was measured using the 12 item Nowicki-Strickland Locus of Control Scale[67]. These scales use questions such as "do you feel that wishing can make good things happen?" and were reverse coded so that high values denote higher locus of control.

Empathy

Empathy was measured at age seven using mother reported responses to five questions about the child's attitudes towards sharing and caring, with responses were on a four-point Likert scale. Responses were coded and summed so that a higher score corresponds to higher levels of empathy.

Impulsivity

Impulsivity was measured during two sessions at the age 8 direct assessment using a behaviour checklist. Testers rated whether the children demonstrated restlessness, impulsivity, fleeting attention, and lacking persistence. Responses were coded to 0 (behaviour not characteristic of the child), 1 (behaviour somewhat characteristic of the child), or 2 (behaviour characteristic of the child). Responses were summed and the mean value of the two sessions was used for each child. At age 11 the children were asked a battery of 10 questions designed to capture impulsive behaviour such as "have you spent all of your money as soon as you got it?". Responses were binary and were summed to give a total impulsivity score.

Personality

Personality was measured at age 13 using the five-factor model of personality. Five broad and independent dimensions of personality, the "Big Five" (extraversion, neuroticism, agreeableness, conscientiousness, and intellect), explain a major portion of judged interindividual difference in personality. They are treated as dimensions with individuals varying continuously along these and most people falling between the extreme. They were measured using self-report responses to 50 items of the International Personality Item Pool in which participants indicate the extent to which statements describe their personality on a five point scale[68].

Cognitive skills and outcomes

IQ

Intelligence was measured during the direct assessments at ages eight and 15 using the short form Wechsler Intelligence Scale for Children (WISC) from verbal, performance, and digit span tests and the Wechsler Abbreviated Scale of Intelligence (WASI) from vocabulary and matrix reasoning tests respectively. These assessments were administered by members of the ALSPAC psychology team overseen by an expert in psychometric testing. The short form tests have high reliability[69] and the ALSPAC measures utilise subtests with reliability ranging from 0.70 to 0.96. Raw scores were recalculated to be comparable to those that would have been obtained had the full test been administered and then age-scaled to give a total overall score combined from the performance and verbal subscales.

Educational attainment

We used four measures of educational attainment. The first three were average fine-graded point scores from three end of 'Key Stage' assessments during compulsory education at ages 11, 14 and 16. As the final stage of compulsory education when the ALSPAC cohort were in school, performance in the age 16 exams impacted further education and employment opportunities. Point scores were used as a richer measure of attainment than categorical level bandings during compulsory education. The fourth measure as a ranking of grades attained in post-compulsory A-levels at age 18, which are a requirement for progressing to university education. We used a measure of the three highest A-level grades grouped into ordered categories (see [70] for a detailed description). At the time the cohort were studying, A-levels were non-compulsory and therefore all participants who did not continue into further education had a coded value of zero. All measures were obtained through data linkage to the UK National Pupil Database (NPD) which represents the most accurate record of educational attainment available in the UK. We extracted the age 11, 14 and 16 scores from the NPD Key Stage 4 (age 16) database as this provided the largest sample size at earlier ages. Age 18 grades were extracted from the NPD Key Stage 5 database (for further information see https://www.gov.uk/government/collections/national-pupil-database).

Employment

At age 23 participants were asked to report whether they were in full time paid employment of more than 30 hours per week, with responses coded as binary.

NEET

Because some participants may not be employed because they are still in full-time education or training, we used a measure of not in education employment or training (NEET) at age 23 with responses coded in binary. The use of a NEET measure ensures that employment results are not biased by participation of the cohort in education or training.

Income

At age 23 participants were asked to report their take-home pay each month if they reported being in paid employment, with responses banded into the following categories: £1-£499; £500-£999; £1000-£1499; £1500-£1999; £2000-£2499; £2500-£3000; £3000+.

Non-response

We also include a binary measure of questionnaire non-response at ages 18 and 24 to allow us to investigate correlations between noncognitive skills and cohort participation.

Statistical analysis

To determine the consistency of noncognitive skills across measures and over time we estimated phenotypic correlations between each measurement-pair (45 measurements [34 noncognitive skills; 2 cognitive skills; 7 socioeconomic outcomes; 2 non-response measures]; 990 unique measurement-pairs). Heritability of each occasion-specific noncognitive skill was estimated using genomic-relatedness-based restricted maximum likelihood (GREML) in the software package GCTA (see [71] for a detailed description of this method). GCTA uses measured SNP level variation across all SNPs (see Genetic data) to estimate the genetic similarity between each pair of unrelated individuals in the sample. Univariate analyses are specified as:

$$y = X\beta + g + \epsilon \tag{1}$$

where y is the inverse normally rank transformed sex and age of measurement standardised measure of phenotype, X is a series of covariates indicating the first 20 principal components of

inferred population structure to control for systematic differences in allele frequencies due to ancestral differences between different subpopulations (population stratification), g is a normally distributed random effect with variance σ_g^2 denoting the contribution of SNPs, and ϵ is residual error with variance σ_ϵ^2. Heritability is then defined as the proportion of total phenotypic variance (genetic variance plus residual variance) explained by common genetic variation:

$$\frac{\sigma_g^2}{\sigma_g^2 + \sigma_\epsilon^2} \tag{2}$$

Non-zero heritability estimates indicate that genetically similar pairs are more phenotypically similar than genetically dissimilar pairs. To estimate the extent to which noncognitive traits share underlying genetic architecture we estimate genetic correlations between each phenotype-pair. Genetic correlations provide an estimate of the extent to which the same genetic variants associate with two phenotypes, that is, the overlap of genetic associations between two phenotypes. Genetic correlations are estimated as:

$$r_g = \frac{cov_g(A, B)}{\sqrt{var_g(A) var_g(B)}} \tag{3}$$

Where r_g is the genetic correlation between phenotypes A and B, $var_g(A)$ is the genetic variance of phenotype A and $cov_g(A, B)$ is the genetic covariance between phenotypes A and B. All analyses are adjusted for false discovery rate using the Benjamini-Hochberg procedure[72] and include the 20 principal components of population structure.

References

1. Bowles, S. & Gintis, H. *Schooling in capitalist America: Educational reform and the contradictions of economic life*. *Sociology of education* (1976). doi:10.2307/3090251

2. Carneiro, P., Crawford, C. & Goodman, A. The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes. *Cent. Econ. Educ.* (2007).

3. Brunello, G. & Schlotter, M. Non Cognitive Skills and Personality Traits : Labour Market Relevance and their Development in Education & Training Systems. *IZA Discuss. Pap.* (2011).

4. Gutman, L. M. & Schoon, I. The impact of non-cognitive skills on outcomes for young people: Literature review. *Educ. Endow. Found. London Engl.* (2013).

5. Heckman, J. J., Stixrud, J. & Urzua, S. The Effects of Cognitive and Noncognitive Abilities on Labor Market Outcomes and Social Behavior. *J. Labor Econ.* **24,** 411–482 (2006).

6. Cunha, F. & Heckman, J. The technology of skill formation. in *American Economic Review* (2007). doi:10.1257/aer.97.2.31

7. Blanden, J., Gregg, P. & Macmillan, L. Accounting for intergenerational income persistence: Noncognitive skills, ability and education. *Econ. J.* (2007). doi:10.1111/j.1468-0297.2007.02034.x

8. Kautz, T. *et al. Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success*. *OECD Education Working Papers* (2014). doi:10.1787/19939019

9. Smithers, L. G. *et al.* A systematic review and meta-analysis of effects of early life non-cognitive skills on academic, psychosocial, cognitive and health outcomes. *Nat. Hum. Behav.* **2,** 867–880 (2018).

10. Duckworth, A. L. & Quinn, P. D. Development and validation of the short Grit Scale (Grit-S). *J. Pers. Assess.* (2009). doi:10.1080/00223890802634290

11. McGue, M., Rustichini, A. & Iacono, W. G. Cognitive, Noncognitive, and Family Background Contributions to College Attainment: A Behavioral Genetic Perspective. *J. Pers.* **85,** 65–78 (2015).

12. Fleche, S. Teacher Quality, Test Scores and Non-Cognitive Skills: Evidence from Primary School Teachers in the UK. (2017).

13. Humphries, J. E. & Kosse, F. On the interpretation of non-cognitive skills – What is being measured and why it matters. *J. Econ. Behav. Organ.* (2017). doi:10.1016/j.jebo.2017.02.001

14. Richardson, M., Abraham, C. & Bond, R. Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychol. Bull.* (2012). doi:10.1037/a0026838

15. Cobb-Clark, D. A. & Schurer, S. Two economists' musings on the stability of locus of control. *Econ. J.* (2013). doi:10.1111/ecoj.12069

16. Fortin, N. M. Greed, Altruism, and the Gender Wage Gap. (2006).

17. Duncan, G. J. *et al.* School Readiness and Later Achievement. *Dev. Psychol.* (2007). doi:10.1037/0012-1649.43.6.1428

18. Heckman, J. J. & Kautz, T. Hard evidence on soft skills. *Labour Econ.* (2012). doi:10.1016/j.labeco.2012.05.014

19. Heckman, J. J. & Rubinstein, Y. The benefits of skill: The importance of noncognitive skills: Lessons from the GED testing program. *Am. Econ. Rev.* (2001). doi:10.2307/2677749

20. Tyler, C. *The role of non-cognitive and cognitive skills in accounting for the intergenerational transmission of 'top job' status*. (Department of Quantitative Social Science - UCL Institute of Education, University College London, 2016). doi:DOI:

21. Hendry, A., Jones, E. J. H. & Charman, T. Executive function in the first three years of life: Precursors, predictors and patterns. *Developmental Review* (2016). doi:10.1016/j.dr.2016.06.005

22. Krueger, A. B. & Schkade, D. Sorting in the Labor Market Do Gregarious Workers Flock to Interactive Jobs? *J. Hum. Resour.* (2008). doi:10.3368/jhr.43.4.859

23. Heckman, J. J. Schools, skills, and synapses. *Econ. Inq.* (2008). doi:10.1111/j.1465-7295.2008.00163.x

24. Lundberg, S. *Non-Cognitive Skills as Human Capital*. *National Bureau of Economic Research* (2017).

25. Lindqvist, E. & Vestman, R. The labor market returns to cognitive and noncognitive ability: Evidence from the swedish enlistment. *Am. Econ. J. Appl. Econ.* (2011). doi:10.1257/app.3.1.101

26. Deke, J. & Haimson, J. Valuing Student Competencies: Which Ones Predict Postsecondary Educational Attainment and Earnings, and for Whom? Final Report. *Math. Policy Res. Inc.* (2006).

27. Borghans, L., Golsteyn, B. H. H., Heckman, J. & Humphries, J. E. Identification problems in personality psychology. *Pers. Individ. Dif.* (2011). doi:10.1016/j.paid.2011.03.029

28. Roberts, B. W. & DelVecchio, W. F. The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies. *Psychol. Bull.* (2000). doi:10.1037/0033-2909.126.1.3

29. Cobb-Clark, D. A. & Schurer, S. The stability of big-five personality traits. *Econ. Lett.* (2012). doi:10.1016/j.econlet.2011.11.015

30. Almlund, M., Duckworth, A. L., Heckman, J. & Kautz, T. Personality Psychology and Economics. *Handb. Econ. Educ.* (2011). doi:10.1016/B978-0-444-53444-6.00001-8

31. Lundberg, S. in *Education, Skills, and Technical Change, and Future US GDP Growth* (University of Chicago Press, 2017).

32. Laajaj, R. & Macours, K. Measuring Skills in Developing Countries. *Policy Res. Work. Pap.* (2017). doi:doi:10.1596/1813-9450-8000

33. Dohmen, T. *et al.* Individual risk attitudes: Measurement, determinants, and behavioral consequences. *J. Eur. Econ. Assoc.* **9,** 522–550 (2011).

34. Turkheimer, E. Three laws of behavior genetics and what they mean. *Curr. Dir. Psychol. Sci.* (2000). doi:10.1111/1467-8721.00084

35. Plomin, R. & Deary, I. J. Genetics and intelligence differences: Five special findings. *Molecular Psychiatry* (2015). doi:10.1038/mp.2014.105

36. Roberts, B. W., Walton, K. E. & Viechtbauer, W. Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies. *Psychol. Bull.* (2006).

doi:10.1037/0033-2909.132.1.1

37. Bouchard, T. J. & McGue, M. Genetic and environmental influences on human psychological differences. *Journal of Neurobiology* (2003). doi:10.1002/neu.10160

38. Bouchard, T. J. & Loehlin, J. C. Genes, evolution, and personality. *Behavior Genetics* (2001). doi:10.1023/A:1012294324713

39. Briley, D. A. & Tucker-Drob, E. M. Genetic and environmental continuity in personality development: A meta-analysis. *Psychol. Bull.* (2014). doi:10.1037/a0037091

40. Wainwright, M. a, Wright, M. J., Luciano, M., Geffen, G. M. & Martin, N. G. Genetic covariation among facets of openness to experience and general cognitive ability. *Twin Res. Hum. Genet.* (2008). doi:10.1375/twin.11.3.275

41. Luciano, M., Wainwright, M. A., Wright, M. J. & Martin, N. G. The heritability of conscientiousness facets and their relationship to IQ and academic achievement. *Pers. Individ. Dif.* (2006). doi:10.1016/j.paid.2005.10.013

42. Kovas, Y. *et al.* Why children differ in motivation to learn: Insights from over 13,000 twins from 6 countries. *Pers. Individ. Dif.* (2015). doi:10.1016/j.paid.2015.02.006

43. Anger, S. & Schnitzlein, D. D. Cognitive skills, non-cognitive skills, and family background: evidence from sibling correlations. *J. Popul. Econ.* (2017). doi:10.1007/s00148-016-0625-9

44. Mazumder, B. Sibling similarities and economic inequality in the US. *J. Popul. Econ.* (2008). doi:10.1007/s00148-006-0127-2

45. Tucker-Drob, E. M. & Harden, K. P. A behavioral genetic perspective on noncognitive factors and academic achievement. (2015).

46. Kong, A. *et al.* The nature of nurture: Effects of parental genotypes. *Science (80-.).* **359,** 424–428 (2018).

47. Baron, J. D. & Cobb-Clark, D. A. *Are Young People's Educational Outcomes Linked to Their Sense of Control? SSRN* (2010). doi:10.2139/ssrn.1594792

48. Romano, E., Babchishin, L., Pagani, L. S. & Kohen, D. School readiness and later achievement: Replication and extension using a nationwide Canadian Survey. *Dev. Psychol.* (2010). doi:10.1037/a0018880

49. Borghans, L., Golsteyn, B. H. H., Heckman, J. J. & Humphries, J. E. What grades and achievement tests measure. *Proc. Natl. Acad. Sci.* (2016). doi:10.1073/pnas.1601135113

50. Heckman, J., Pinto, R. & Savelyev, P. Understanding the mechanisms through which an influential early childhood program boosted adult outcomes. *Am. Econ. Rev.* (2013). doi:10.1257/aer.103.6.2052

51. Taylor, A. *et al.* Exploring the association of genetic factors with participation in the Avon Longitudinal Study of Parents and Children. *Int. J. Epidemiol.* **In press,** (2018).

52. Martin, J. *et al.* Association of genetic risk for schizophrenia with nonparticipation over time in a population-based cohort study. *Am. J. Epidemiol.* **183,** 1149–1158 (2016).

53. Biroli, P. Health and skill formation in early childhood. *UBS Int. Cent. Econ. Soc.* **Working Pa,** (2016).

54. Young, A. I. *et al.* Relatedness disequilibrium regression estimates heritability without environmental bias. *Nat. Genet.* (2018). doi:10.1038/s41588-018-0178-9

55. Haworth, S. *et al.* Common genetic variants and health outcomes appear geographically structured in the UK Biobank sample: Old concerns returning and their implications. *bioRxiv* (2018).

56. Plomin, R., Krapohl, E. & O'Reilly, P. F. Assortative mating-a missing piece in the Jigsaw of psychiatric genetics. *JAMA Psychiatry* (2016). doi:10.1001/jamapsychiatry.2015.3204

57. Redden, D. T. & Allison, D. B. The effect of assortative mating upon genetic association studies: Spurious associations and population substructure in the absence of admixture. *Behav. Genet.* (2006). doi:10.1007/s10519-006-9060-0

58. Boyd, A. *et al.* Cohort Profile: the 'children of the 90s'--the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol* **42,** 111–127 (2013).

59. Fraser, A. *et al.* Cohort profile: The avon longitudinal study of parents and children: ALSPAC mothers cohort. *Int. J. Epidemiol.* **42,** 97–110 (2013).

60. Goodman, R., Ford, T., Simmons, H., Gatward, R. & Meltzer, H. Using the Strengths and Difficulties Questionnaire (SDQ) to screen for child psychiatric disorders in a community sample. *Br. J. Psychiatry* **177,** 534–539 (2000).

61. Frankenburg, W. K., Dodds, J., Archer, P., Shapiro, H. & Bresnick, B. The Denver II: A Major Revision and Restandardization of the Denver Developmental Screening Test. *Pediatrics* (1992).

62. Fenson, L. *et al. The MacArthur Communicative Development Inventories : User's guide and technical manual*. *Singular Publishing Group, Baltimore, USA* (1993).

63. Harter, S. Self-Perception Profile for Children (SPPC): User's Guide. *Developmental Review* (1985). doi:10.1006/drev.2000.0516

64. Rosenberg, M. Rosenberg Self-Esteem Scale. *New York* (1965). doi:/S0034-98872009000600009

65. Bachman, J. *YOUTH IN TRANSITION, VOLUME 2: THE IMPACT OF FAMILY BACKGROUND AND INTELLIGENCE ON TENTH-GRADE BOYS*. *YOUTH IN TRANSITION, VOLUME 2: THE IMPACT OF FAMILY BACKGROUND AND INTELLIGENCE ON TENTH-GRADE BOYS* (1970).

66. Nowicki, S. & Duke, M. P. A preschool and primary internal-external control scale. *Dev. Psychol.* (1974). doi:10.1037/h0037253

67. Nowicki, S. & Strickland, B. R. A locus of control scale for children. *J. Consult. Clin. Psychol.* (1973). doi:10.1037/h0033978

68. Goldberg, L. R. A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models. *Personality psychology in Europe* (1999). doi:citeulike-article-id:1856566

69. Crawford, J. R., Anderson, V., Rankin, P. M. & MacDonald, J. An index-based short-form of the WISC-IV with accompanying analysis of the reliability and abnormality of differences. *Br. J. Clin. Psychol.* **49,** 235–258 (2010).

70. Morris, T., Dorling, D. & Davey Smith, G. How well can we predict educational outcomes? Examining the roles of cognitive ability and social position in educational attainment. *Contemp. Soc. Sci.* **11,** 1–15 (2016).

71. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: A tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88,** 76–82 (2011).

72. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. Ser. B* (1995). doi:10.2307/2346101

Acknowledgements

We are extremely grateful to all the families who took part in this study, the midwives for their help in recruiting them, and the whole ALSPAC team, which includes interviewers, computer and laboratory technicians, clerical workers, research scientists, volunteers, managers, receptionists and nurses. The Medical Research Council (MRC) and the University of Bristol support the MRC Integrative Epidemiology Unit [MC_UU_12013/1, MC_UU_12013/9, MC_UU_00011/1]. A comprehensive list of grants funding is available on the ALSPAC website (http://www.bristol.ac.uk/alspac/external/documents/grant-acknowledgements.pdf); data used at age 23 was specifically funded by the Wellcome Trust and MRC [102215/2/13/2]. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Bristol. REDCap (Research Electronic Data Capture) is a secure, web-based application designed to support data capture for research studies, providing 1) an intuitive interface for validated data entry; 2) audit trails for tracking data manipulation and export procedures; 3) automated export procedures for seamless data downloads to common statistical packages; and 4) procedures for importing data from external sources. The Economics and Social Research Council (ESRC) support NMD via a Future Research Leaders grant [ES/N000757/1]. The UK Medical Research Council and Wellcome (Grant ref: 102215/2/13/2) and the University of Bristol provide core support for ALSPAC. This publication is the work of the authors and Tim Morris will serve as guarantors for the contents of this paper. GWAS data was generated by Sample Logistics and Genotyping Facilities at Wellcome Sanger Institute and LabCorp (Laboratory Corporation of America) using support from 23andMe. No funding body has influenced data collection, analysis or its interpretations. This work was carried out using the computational facilities of the Advanced Computing Research Centre - http://www.bris.ac.uk/acrc/ and the Research Data Storage Facility of the University of Bristol - http://www.bris.ac.uk/acrc/storage/.