Investigating the longitudinal consistency and genetic architecture of non-cognitive skills, and their relation to educational attainment.

Tim T Morris* [1,2], George Davey Smith [1,2], Gerard van Den Berg [3], 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: non-cognitive; cognitive; education; genetics; ALSPAC.

Abstract

Non-cognitive skills predict a range of socioeconomic outcomes including educational attainment and employment. Many studies have used cross sectional data and therefore the longitudinal consistency of non-cognitive skills are poorly understood. Using data from a UK cohort, we assess (1) the consistency of non-cognitive skills over a 17-year period throughout childhood, (2) their associations with educational outcomes, and (3) their genomic architecture. Consistency over time is high for behavioural and communication skills but low for others, implicating a high noise to signal ratio for many non-cognitive skills. Associations between non-cognitive skills and educational outcomes are observed for skills measured in mid to late childhood and are at most a third of the size of IQ-education associations. Consistent non-zero heritability estimates and genetic correlations with education are observed only for behavioural difficulties. Our results suggest caution in the use of non-cognitive measures as consistent indicators of underlying skills during early childhood.

Introduction

Non-cognitive skills have been argued as an important mechanism through which a range of outcomes are differentiated between individuals (*1–4*). Non-cognitive skills are broadly defined as "personality traits, goals, character, motivations, and preferences" (*5*) which are considered complementary to cognitive measures such as intelligence or general cognitive ability (*1*). A wide range of characteristics have been proposed as non-cognitive skills, including persistence, motivation and temperament (*1*, *6*). Non-cognitive skills are therefore more broadly defined than cognitive skills as they are thought to capture a wider range of multifaceted personality and behavioural traits.

Previous studies have demonstrated that people with higher measurements of non-cognitive skills have better education, health and labour market outcomes (*2–8*). For example, Borghans and colleagues report that personality explains 16% of the variance in achievement scores in a US sample (*8*). A meta-analysis found modest correlations between conscientiousness and academic achievement (r=0.24) but not for other personality types (r=-0.05 to 0.6) (*9*). Schmidt and Hunter showed that integrity and conscientiousness predict job and training performance (*10*). Von Stumm and colleagues show that intellectual curiosity and effort are associated with academic attainment (*11*). However, the evidence supporting the importance of non-cognitive skills on socioeconomic and health outcomes is diverse and inconsistent. Recent research has demonstrated that associations between personality types and life outcomes replicate well (*12*). However, a systematic review revealed that many other non-cognitive skills have small and heterogenous effects on a range of outcomes including academic achievement (*13*). While there's been considerable research into the relationship between non-cog skills and a range of outcomes, most studies had very short follow-up, particularly where skills were measured in childhood (*13*). Many studies have also been unable to adjust estimates for cognitive ability, despite its strong attenuating effect on non-cognitive associations with outcomes (*10*, *11*). We therefore need more evidence about the relationship between non-cognitive skills in childhood and later outcomes, and the factors that mediate these effects.

Non-cognitive skills are thought to be a promising target for interventions in early childhood to improve outcomes because they are potentially more malleable than cognitive skills (*3*, *5*, *14*). To be good candidates for policy interventions, non-cognitive measures must reliably pick up a consistent signal of an underlying skill. While changes to non-cognitive skills are required (and indeed anticipated) during interventions, they should remain relatively stable in populations that experience no intervention. There is however a lack of longitudinal evidence on the consistency of non-cognitive skills over time (*15*). Test-retest reliability (calculated as the correlation coefficient between the same measure at two occasions) has been estimated at 0.31 for personality in childhood (*14*), 0.12 and 0.46 to 0.61 for risk aversion, and 0.49 for locus of control (*16*, *17*). These values are lower than those reported for cognitive skills (c.f. 0.52 for digit span to 0.82 for reading) (*16*), and suggest that measurement error on non-cognitive skills may be relatively high (*5*). Modest correlations between different non-cognitive skills have been observed for a range of traits such as grit and conscientiousness (*18*); academic effort and academic problems (*7*); and internalising and externalising behaviours (*19*). The bulk of reported correlations between non-cognitive skills though have been small (*7*, *18*), supporting the multidimensionality of non-cognitive skills. The relationships between non-cognitive skills is however unknown as few datasets have measured many skills.

< Table 1 here >

The consistency of non-cognitive skills and the robustness of their associations with socioeconomic outcomes can also be explored with genetic data. There has been relatively little research into the genetics of non-cognitive skills compared to cognitive skills (*20*). Personality types are the most studied of the non-cognitive skills and have heritability estimated at around 0.5 (*21*). Furthermore, there is genetic evidence that personality types are stable over time (*22*), providing support that personality measures successfully pick up genuine signal of an underlying trait. Heritability estimates for other non-cognitive skills have varied widely from 0.44 to 0.79 for self-control (*23*, *24*); 0.36 for alienation (*7*); 0.83 for academic effort (*7*); 0.31 to 0.56 for aspects of openness (*25*); 0.18 to 0.49 for aspects of conscientiousness (*26*); and 0.40 for enjoyment and self-perceived ability (*27*). Sibling correlations for non-cognitive skills are lower than for cognitive skills (*28*), consistent with a smaller genetic component. There has not yet been a comprehensive analysis investigating the heritability of multiple non-cognitive skills within the same population sample and the extent of genetic correlation between different non-cognitive skills has not been widely investigated (*29*).

In this study, we contribute to the literature with a comprehensive analysis into the phenotypic and genotypic structure of non-cognitive skills. Using data from a UK cohort study, we investigate the relationship between all non-cognitive measures and educational and labour market outcomes up to age 23 available in the study. We answer two related research questions: 1) How consistent are non-cognitive skills measured in a large cohort study over time? and; 2) Do non-cognitive skills associate with socioeconomic outcomes once other factors including cognitive ability have been accounted for?

Results

How consistent are non-cognitive skills over time?

Behavioural skills measured on the SDQ scale were consistent over time (r=0.18 to 0.73), with higher consistency between measures closer together in time (Figure 1). Correlations between parent and teacher reported SDQ measures were weak even when measured contemporaneously. Correlations for other non-cognitive measures over time were generally low (Figure 1), suggesting trait inconsistency (highly variable skills) or lack of reliability (inconsistent measurement). 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 many non-cognitive skills were low, they were generally positive where distinguishable from zero, and therefore do not appear to capture conflicting information at different measurement occasions. The weak temporal phenotypic correlations of non-cognitive measures contrast sharply to those for cognitive measures of IQ (r=0.60) and measures of educational attainment (r>0.78 for compulsory education).

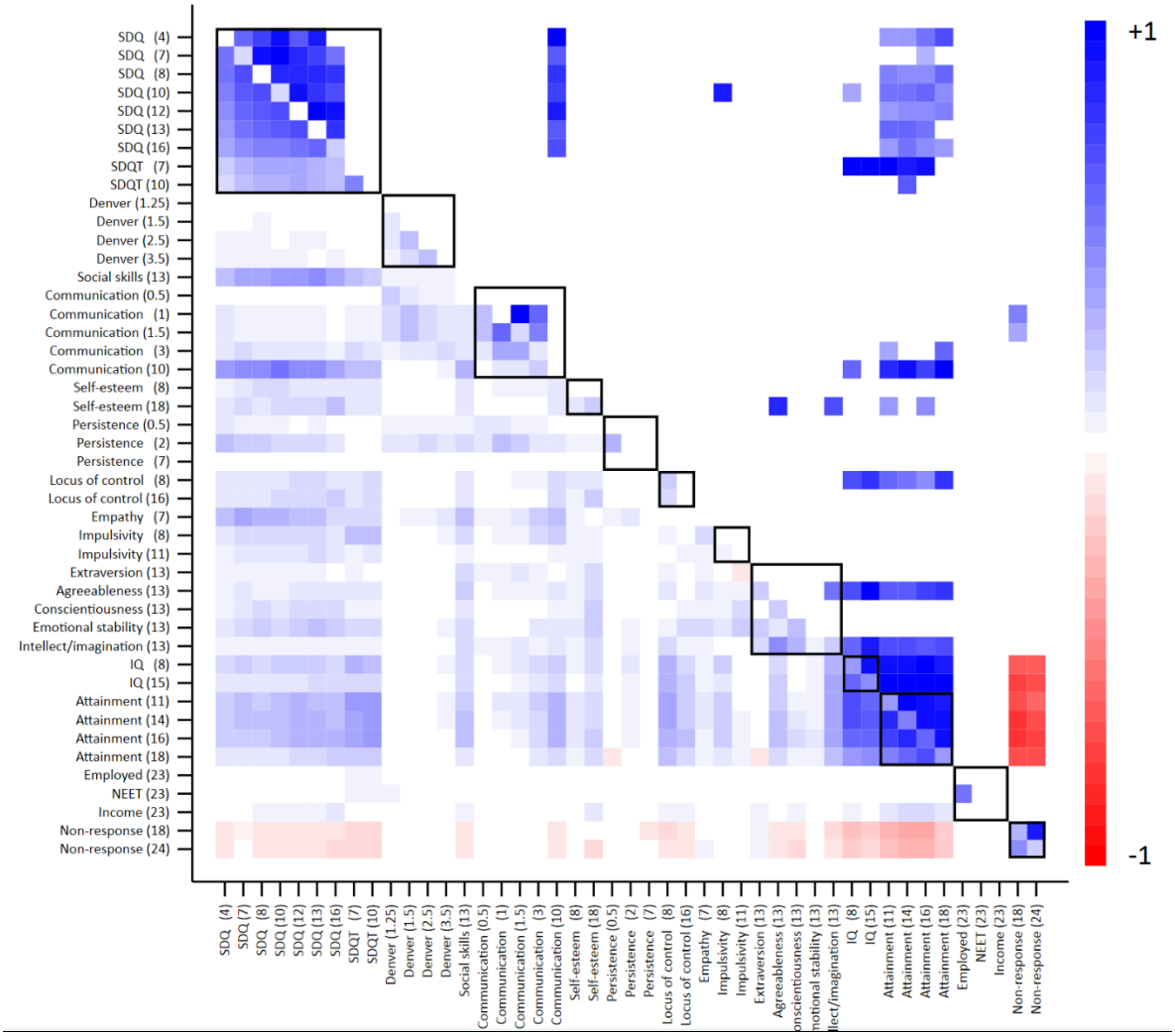

Figure 1: **Heritability, phenotypic correlations and genetic correlations of skills and outcomes. Age in parentheses.** SDQ, strengths and difficulties; NEET, not in education, employed or training. Values on the diagonal represent univariate heritabilities; values below the diagonals represent phenotypic 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 non-cognitive skills?

Phenotypic correlations between different of non-cognitive measures were generally low (Figure 1). This suggests that these different measures capture empirically distinct phenomena rather than a single latent non-cognitive factor that underlies these skills. The SDQ scale was the only measure that correlated consistently with other non-cognitive 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 were almost exclusively positive, suggesting that where patterns were observed, children who scored high on one non-cognitive skill generally also scored highly on another. All non-cognitive skills except for the teacher reported measures of SDQ, locus of control at age 8 and the intellect/imagination personality type correlated 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 was 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 non-cognitive skills associate with education and labour market outcomes?

Phenotypic correlations with educational attainment were highest for the SDQ scale (r=0.13 to 0.42) (Figure 1). The teachers' SDQ 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). Correlations between educational attainment and cognitive ability were considerably higher (r=0.43 to 0.73). Phenotypic correlations for non-cognitive skills with employment, NEET and income were low (r<=0.08 for employment; r<=0.07 for NEET; r<=0.19 for income). Correlations between labour market outcomes and cognitive skills were also very low (r<=0.09). Non-cognitive and cognitive skills were generally 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 multivariable regressions of age 16 educational attainment on non-cognitive and cognitive skills (Figure 2). Each skill was standardised and analysed independently controlling for sex, month of birth and IQ at age 8. A one standard deviation (SD) increase in non-cognitive skills was associated with a 0.04 SD decrease to 0.21 SD increase in age 16 attainment. There was considerable heterogeneity for estimates between skills and measurement occasions for many non-cognitive measures. Associations were generally larger the closer non-cognitive skills were measured to educational attainment. 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).

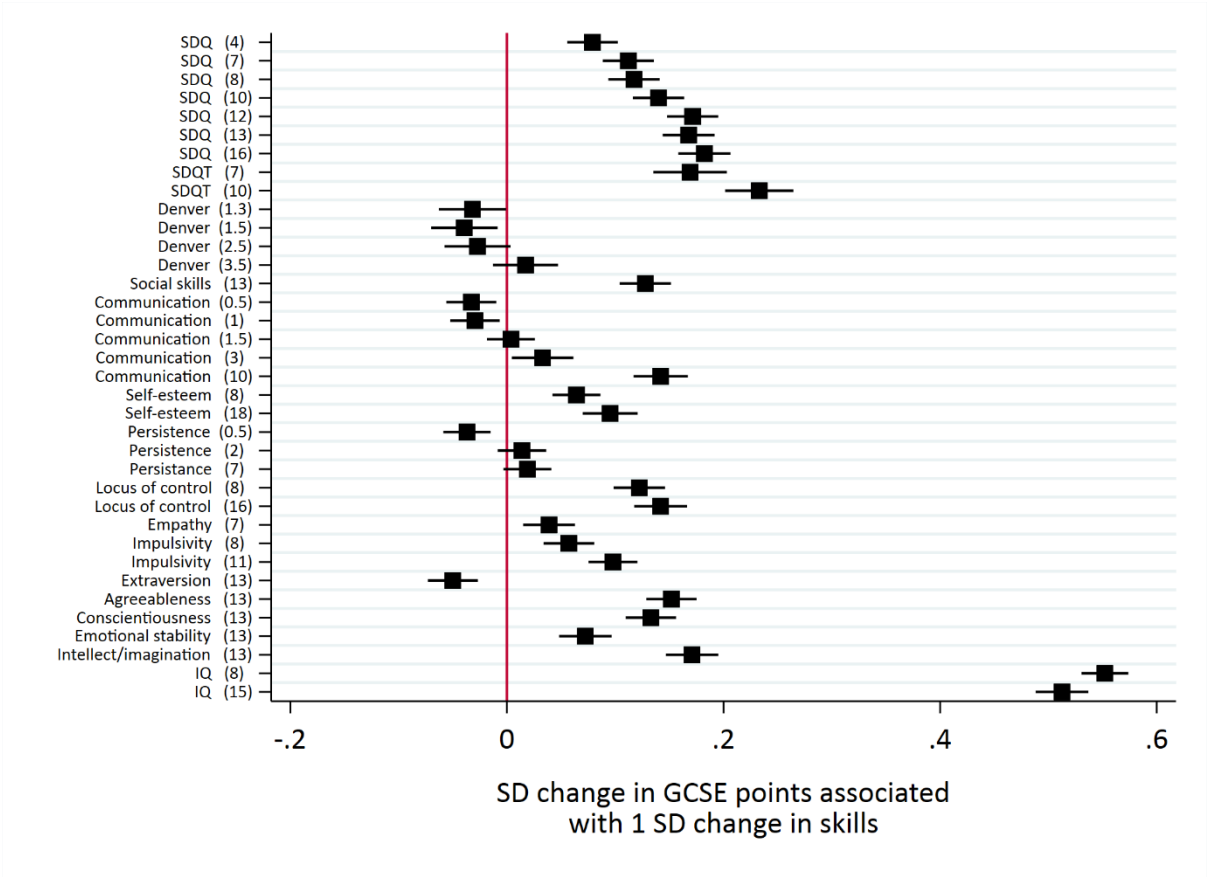

Figure 2: **Associations between skills and educational attainment at age 16 adjusted for sex, month of birth and IQ at age 8.** Age (years) in parentheses. SDQ, strengths and difficulties.

Are non-cognitive skills heritable?

Figure 3 displays the heritability estimates for all our measures. The SDQ scale is the only non-cognitive measure for which there was reasonable evidence of non-zero heritability at multiple occasions (h^2_{SNP}=0.18 to 0.23). Results were broadly comparable when using the internalising and externalising subscales of the SDQ (Supplementary Figure 2). Non-zero heritability was also 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). There was imprecision in the heritability point estimates, but the upper bounds were estimated below 0.30 for most non-cognitive skills. In contrast, the heritability of cognitive skills was higher at 0.43 at age 8 and 0.47 at age 15. Educational outcomes were highly heritable (h^2_{SNP}>0.4) but there was little evidence of heritability for the labour market outcomes. Heritability of questionnaire non-response was estimated higher and with less uncertainty than the non-cognitive measures (h^2_{SNP}=0.34 and h^2_{SNP}=0.21 at ages 18 and 24 respectively).

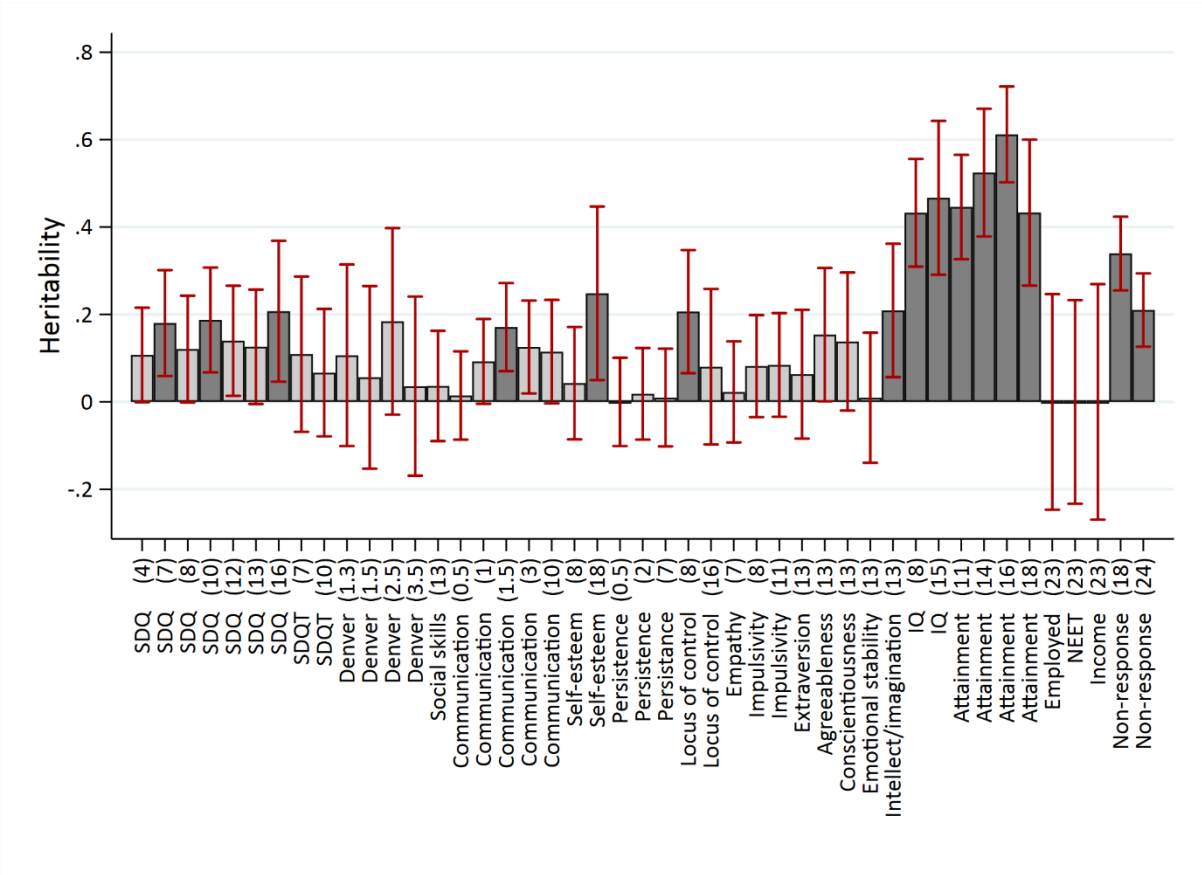

Figure 3: **Heritability of skills and outcomes.** Age in parentheses. SDQ, strengths and difficulties; NEET, not in education, employed or training. Dark coloured bars represent estimates below FDR threshold. See Supplementary Table 2 for full results.

Do non-cognitive skills have a shared genetic architecture?

There was very limited evidence for genetic correlations across non-cognitive skills (Figure 1, above the diagonal), though estimation precision was very low for most skills. There was strong evidence for within trait genetic correlations over time only for the parent reported SDQ measures (rG= 0.62 to 1.00) and IQ, which had near-unity genetic overlap (rG= 0.97). Genetic correlations between different non-cognitive measures were only observed between the parent reported SDQ measures and communication at age 10 (rG=0.68 to 0.91). This may however reflect that not all measures are parent reported by the same person and therefore could reflect child and parent genetic factors. 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 non-cognitive or cognitive skills.

Discussion

Our results provide longitudinal evidence into the consistency of a large range of non-cognitive skills throughout childhood. Behavioural and communication skills were correlated both phenotypically and genotypically from age 4 to 16. However, correlations between other non-cognitive skills were low, contrasting with previous research that has evidenced temporal stability of non-cognitive skills (*14*, *15*). Our results may have reflected that some measures of non-cognitive skills are poor instruments for measuring an underlying trait. For example, the SDQ scale uses responses to a large battery of questions which may have lower measurement error. Future research proposing more detailed measures of non-cognitive skills may therefore help to increase the reliability of non-cognitive skill measurements in capturing underlying skills and the value of these to research. The lack of consistency observed for other skills may also have reflected other mechanisms including genuine temporal intra-individual variation in the expression of skills or differences over time (e.g. due to schooling). It is important to note there is some variation in the measurement of skills that we used. This is somewhat unavoidable as skill measurements vary depending on the age at which a child is assessed, but measurement artefact may nevertheless reflect some of the variation in measurement. Correlations between different non-cognitive skills were weak, highlighting that these constructs are independent and capture different psychosocial phenomena rather than a single latent non-cognitive factor. The SDQ scale was the only measure to consistently correlate with other non-cognitive skills, conforming to a previous study that found low phenotypic correlations between different non-cognitive skills (*7*). The cognitive measures we used were more strongly correlated over the same period, though this may reflect that they are more reliably measured than the non-cognitive measures. One interpretation of these results is that future studies should combine longitudinal data on non-cognitive skills with multi-source multi-method approaches may help to ensure more accurate measurement.

We found limited evidence of genetic correlations within or between non-cognitive skills, with only the parent reported SDQ demonstrating consistent genetic architecture over time. This may reflect the influence of shared parent-offspring genetics on reporting, or parental genetics influencing offspring non-cognitive skills indirectly through dynastic effects (*30*) (there would be no such shared teacher-student genetics). Future work is therefore required to determine the presence of dynastic effects in the development of non-cognitive skills. Previous twin studies have found strong genetic correlations between non-cognitive and cognitive skills (*25*, *26*), but our results did not support this. The differences between these results and previous studies may have arisen due to several reasons, including differences in the quality of non-cognitive skill measurements; differences due to cross-sectional and longitudinal study designs; reporting bias by the study mothers (for example where their child performed poorly); differences in the study populations; selective reporting and the use of samples or measures of convenience in previous studies (*13*). While we reported every non-cognitive skill in ALSPAC, many previous studies have reported only one or a small number of non-cognitive skills.

Our results also showed that associations between non-cognitive skills and socioeconomic outcomes were generally weak, contradicting findings from previous individual studies (*3*, *5*, *6*) but supporting a recent systematic review (*13*). A strength of our findings is that we used multiple measurements of non-cognitive skills at different ages and a range of outcomes. Behavioural problems as captured by the SDQ scale, social skills, and locus of control were the only non-cognitive measures to

phenotypically associate with educational outcomes strongly and consistently. Associations were stronger between educational attainment and the teacher than the parent reported measures of the SDQ scale. This may suggest that teachers more objectively identify education related problematic behaviour than parents, or that teachers' response to their pupils directly influences their educational outcomes. It is important to note though that non-cognitive skills may be expected to have varying relevance to educational and labour market outcomes given the diverse set of characteristics that they are designed to capture. Further longitudinal research into the associations between non-cognitive skills and a wide range of outcomes is required to better elucidate these relationships. Neither non-cognitive nor cognitive skills associated strongly with labour market outcomes. While this contrasts with previous studies (*3*, *5*), our labour market outcomes are observed soon after entry to the labour market and may therefore more closely resemble institutional effects than non-cognitive skills. Some have argued that non-cognitive skills and personality traits are as important as cognitive skills for many dimensions of behaviour and socioeconomic outcomes (*4*), but our results do not support this. The effect sizes of non-cognitive skills were at most a third that of cognitive skills. Many non-cognitive 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. This may have important implications for participant representativeness and generalisability in cohort studies. Furthermore, cognitive ability and attainment were negatively correlated with non-response genotypically, adding to the growing body of evidence that non-response is genetically patterned (*31*).

The SDQ scale was the only non-cognitive 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). Heritability and stability depend on the developmental period being examined and so comparison of estimates is difficult. However, previous research has demonstrated that the heritability of cognitive ability rises with age while the heritability of personality traits decreases with age (*32*), which our data contradicts. 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 non-cognitive skills, but our estimates suggested an upper bound heritability of 0.3 for most skills.

This study has several limitations. First, it is possible that measurement error was unusually high in the non-cognitive measures used in the ALSPAC study. However, the measures used in ALSPAC have been widely validated are consistent with those used in previous studies (*19*). Furthermore, measurement error would need to have been high across all measures used from birth to age 18 so it is unlikely that this explains our results. Future studies into test-retest reliability of non-cognitive 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). This is likely due to the low estimated heritability of the non-cognitive skills. Low heritability implies a small genotypic contribution (either in the number of individual variants associated or the strength of associations) and therefore lower power to detect genetic correlations between these skills. Our sample sizes are fairly low to detect such small

univariate and bivariate genetic associations. Future studies conducted on larger samples are required to more accurately the estimate heritability of, and genetic correlations between non-cognitive skills and other phenotypes. Third, our genetic associations may have been biased by uneven linkage disequilibrium, residual population structure, or assortative mating (*30*, *33*). We controlled for the first twenty principle components of population structure to account for population structure but this may not have accounted for all differences (*34*). Assortative mating is thought to be low for non-cognitive traits (*21*), but will have inflated our genetic associations if present (*35*). It is possible that assortment on non-cognitive skills may be negative and future work is required to determine this. Finally, it is possible that some of the measures could reflect parental rather than child genes where they are parent rather than self-reported.

In conclusion, our results highlight that non-cognitive skills are likely to be highly heterogenous. Measures of noncognitive skills were varied both over time and across different measures in the same individuals, but some individual measures such as the SDQ demonstrated strong internal consistency throughout childhood. Furthermore, many non-cognitive measures associated weakly with educational and employment outcomes at entry to the labour market, particularly when measured early in childhood.

Materials and Methods

Experimental design

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 (*36*, *37*). 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 (*36*). 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 non-cognitive 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.

Non-cognitive skills

We used all non-cognitive skills 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.

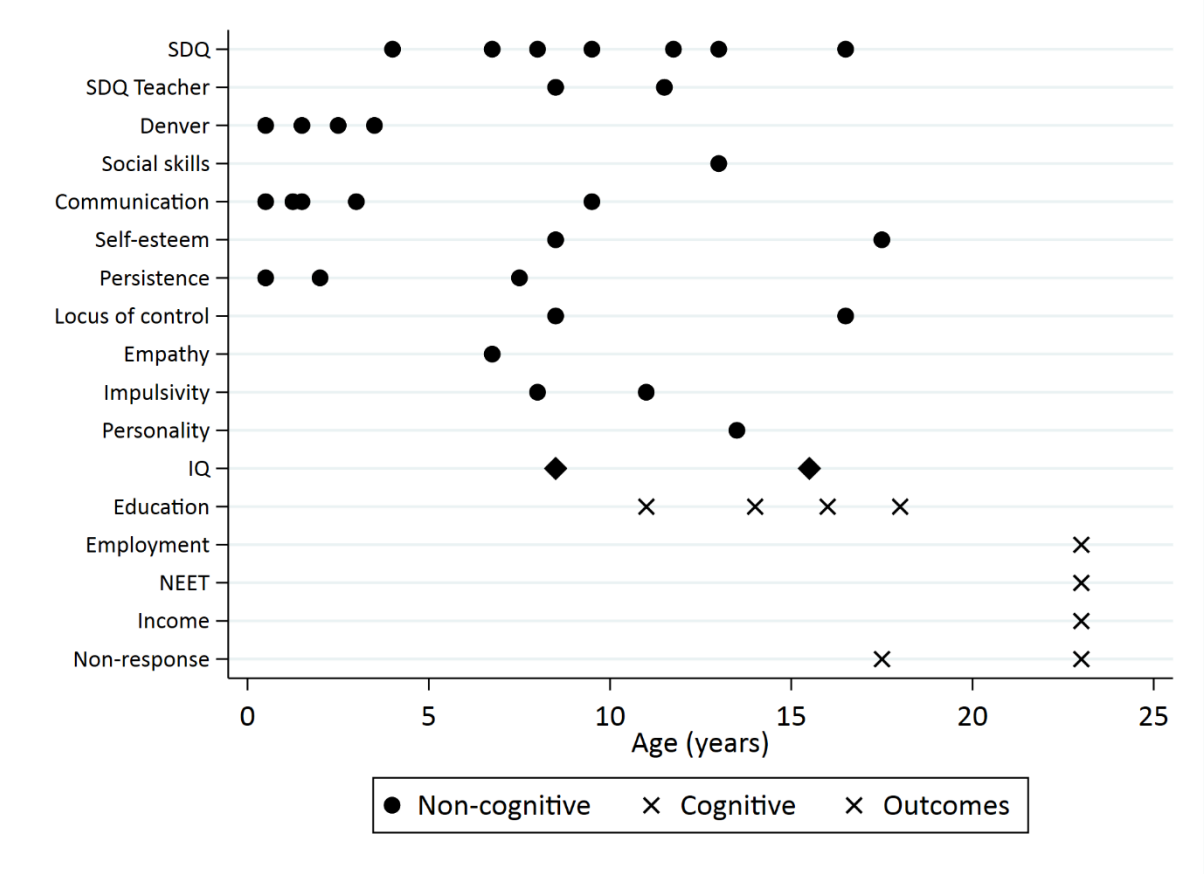

Figure 4: **Timeline of measures in ALSPAC.** SDQ, strengths and difficulties; NEET, not in education, employed or training.

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 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 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 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 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. 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). At age 16 locus of control was measured using the 12 item Nowicki-Strickland Locus of Control Scale. 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.

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 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 (*38*) 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 non-cognitive skills and cohort participation.

Statistical analysis

To determine the consistency of non-cognitive skills across measures and over time we estimated phenotypic correlations between each measurement-pair (45 measurements [34 non-cognitive skills; 2 cognitive skills; 7 socioeconomic outcomes; 2 non-response measures]; 990 unique measurement-pairs). Heritability of each occasion-specific non-cognitive skill was estimated using genomic-relatedness-based restricted maximum likelihood (GREML) in the software package GCTA

510 (see (*39*) for a detailed description of this method). GCTA uses measured SNP level variation across
511 all SNPs (see Genetic data) to estimate the genetic similarity between each pair of unrelated
512 individuals in the sample. Univariate analyses are specified as:

513 $$y = X\beta + g + \epsilon \quad (1)$$

514 where y is the inverse normally rank transformed sex and age of measurement standardised
515 measure of phenotype, X is a series of covariates indicating the first 20 principal components of
516 inferred population structure to control for systematic differences in allele frequencies due to
517 ancestral differences between different subpopulations (population stratification), g is a normally
518 distributed random effect with variance σ_g^2 denoting the contribution of SNPs, and ϵ is residual error
519 with variance σ_ϵ^2. Heritability is then defined as the proportion of total phenotypic variance (genetic
520 variance plus residual variance) explained by common genetic variation:

521 $$\frac{\sigma_g^2}{\sigma_g^2 + \sigma_\epsilon^2} \quad (2)$$

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

528 $$r_g = \frac{cov_g(A, B)}{\sqrt{var_g(A) var_g(B)}} \quad (3)$$

529 Where r_g is the genetic correlation between phenotypes A and B, $var_g(A)$ is the genetic variance of
530 phenotype A and $cov_g(A, B)$ is the genetic covariance between phenotypes A and B. All analyses
531 are adjusted for false discovery rate using the Benjamini-Hochberg procedure (*40*) and include the
532 20 principal components of population structure.

533

References

1. S. Bowles, H. Gintis, *Schooling in capitalist America: Educational reform and the contradictions of economic life* (1976).

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

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

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

5. T. Kautz *et al.*, "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success" (2014), , doi:10.1787/19939019.

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

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

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

9. M. C. O'Connor, S. V Paunonen, Big Five personality predictors of post-secondary academic performance. *Pers. Individ. Dif.* **43**, 971–990 (2007).

10. F. L. Schmidt, J. E. Hunter, The validity and utility of selection methods in personnel psychology: Practical and theoretical implications of 85 years of research findings. *Psychol. Bull.* **124**, 262 (1998).

11. S. Von Stumm, B. Hell, T. Chamorro-Premuzic, The hungry mind: Intellectual curiosity is the third pillar of academic performance. *Perspect. Psychol. Sci.* **6**, 574–588 (2011).

12. C. J. Soto, How replicable are links between personality traits and consequential life outcomes? The Life Outcomes Of Personality Replication Project. *Psychol. Sci.* (2019).

13. L. G. Smithers *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).

14. B. W. Roberts, W. F. DelVecchio, 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.

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

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

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

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

575 19. S. Fleche, Teacher Quality, Test Scores and Non-Cognitive Skills: Evidence from Primary
576 School Teachers in the UK (2017) (available at http://eprints.lse.ac.uk/83602/).

577 20. R. Plomin, I. J. Deary, Genetics and intelligence differences: Five special findings. *Mol.*
578 *Psychiatry* (2015), , doi:10.1038/mp.2014.105.

579 21. T. J. Bouchard, J. C. Loehlin, Genes, evolution, and personality. *Behav. Genet.* (2001), ,
580 doi:10.1023/A:1012294324713.

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

583 23. Y. E. Willems *et al.*, Genetic and Environmental Influences on Self-Control: Assessing Self-
584 Control with the ASEBA Self-Control Scale. *Behav. Genet.* **48**, 135–146 (2018).

585 24. K. Lemery-Chalfant, L. Doelger, H. H. Goldsmith, Genetic relations between effortful and
586 attentional control and symptoms of psychopathology in middle childhood. *Infant Child Dev.*
587 *An Int. J. Res. Pract.* **17**, 365–385 (2008).

588 25. M. a Wainwright, M. J. Wright, M. Luciano, G. M. Geffen, N. G. Martin, Genetic covariation
589 among facets of openness to experience and general cognitive ability. *Twin Res. Hum. Genet.*
590 (2008), doi:10.1375/twin.11.3.275.

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

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

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

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

600 30. A. Kong *et al.*, The nature of nurture: Effects of parental genotypes. *Science (80-.).* **359**, 424–
601 428 (2018).

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

604 32. D. A. Briley, E. M. Tucker-Drob, Comparing the Developmental Genetics of Cognition and
605 Personality over the Life Span. *J. Pers.* **85**, 51–64 (2017).

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

608 34. S. Haworth *et al.*, Common genetic variants and health outcomes appear geographically
609 structured in the UK Biobank sample: Old concerns returning and their implications. *bioRxiv*
610 (2018) (available at http://biorxiv.org/content/early/2018/04/11/294876.abstract).

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

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

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

38. T. Morris, D. Dorling, G. Davey Smith, 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).

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

40. Y. Benjamini, Y. Hochberg, 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] and TTM via a Postdoctoral Research Fellowship [ES/S011021/1]. 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/.

Author Contributions

NMD, GvDB and GDS conceived the research. TTM and NMD developed the research plan. TTM analyzed the data. All authors interpreted and discussed the data and contributed to the writing of the manuscript.

Competing interests

The authors declare no competing interests.

Data and materials availability

The data that support the findings of this study are available from ALSPAC (conditional on approval by the study executive committee) but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data may be available from the authors upon reasonable request and with permission of the ALSPAC executive committee.