1

Consistency of non-cognitive skills and their relation to educational outcomes in a UK cohort.

Running title: Consistency of non-cognitive skills.

Tim T Morris Ph.D.* [1,2], George Davey Smith Ph.D. [1,2], Gerard van Den Berg Ph.D. [3], Neil M Davies Ph.D.[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. Tel: (+44)(0)117 331 4032.

Keywords: Education; cognitive; socioemotional; genetics; soft skills; ALSPAC.

18 Abstract

19 Non-cognitive skills have previously been associated with a range of health and socioeconomic

20 outcomes, though there has been considerable heterogeneity in published research. Many studies

21 have used cross sectional data and therefore the longitudinal consistency of measures designed to

22 capture non-cognitive skills is poorly understood. Using data from a UK cohort, we assess the

23 consistency of non-cognitive skills over a 17-year period throughout childhood and adolescence,

24 their genomic architecture, and their associations with socioeconomic outcomes. We find that

25 longitudinal measurement consistency is high for behavioural and communication skills but low for

26 other non-cognitive skills, implicating a high noise to signal ratio for many non-cognitive skills.

27 Consistent non-zero heritability estimates and genetic correlations applied to cross-sectional

28 measures are observed only for behavioural difficulties. When aggregating across multiple

29 measurements, we find evidence of low heritability (h^2_{SNP}=0.1 − 0.2) for behaviour, communication,

30 self-esteem and locus of control. We find weak correlations between aggregate measures of skills,

31 further supporting cross-sectional measurement error in the non-cognitive measures. Associations

32 between non-cognitive skills and educational outcomes are observed for skills measured in mid to

33 late childhood and these are at most a third of the size of IQ-education associations. These results

34 suggest that individual measures designed to capture non-cognitive skills may be subject to

35 considerable measurement error and provide unreliable indicators of children's skills. However,

36 aggregate measures that leverage longitudinal data may more reliably identify underlying non-

37 cognitive traits.

38

39 # Introduction

40 Non-cognitive (or socioemotional) skills have been posited as an important driver of individual

41 differences for a range of health and socioeconomic outcomes [1-8]. Non-cognitive skills are broadly

42 defined as "personality traits, goals, character, motivations, and preferences" [3] an are considered

43 complementary to cognitive measures such as intelligence or general cognitive ability [8]. A wide

44 range of characteristics have been proposed to fall under the umbrella term of non-cognitive skills,

45 including multifaceted personality and behavioural traits such as persistence, motivation and

46 temperament [2,8].

47 Research into noncognitive skills has demonstrated the role that these may play in outcomes.

48 Borghans and colleagues reported that personality explains 16% of the variance in achievement

49 scores in a US sample [7]. A meta-analysis found modest correlations between conscientiousness and

50 academic achievement (r=0.24) but not for other personality types (r=-0.05 to 0.6) [9]. Schmidt and

51 Hunter showed that integrity and conscientiousness predict job and training performance [10]. Von

52 Stumm and colleagues show that intellectual curiosity and effort are associated with academic

53 attainment [11].

54 The evidence supporting the role of non-cognitive skills on socioeconomic and health outcomes is

55 however diverse and inconsistent. While research has demonstrated that associations between

56 personality types and life outcomes replicate well [12], a recent systematic review showed that many

57 other non-cognitive skills have small and heterogenous effects [13]. These heterogenous effects may

58 be driven by poor measurement of non-cognitive skills. Many studies have been restricted by short

59 follow-up [13] or a lack of longitudinal data on non-cognitive skills to measure measurement

60 consistency over time [14]. Test-retest reliability has been estimated at 0.31 for personality in

61 childhood [15], 0.12 and 0.46 to 0.61 for risk aversion, and 0.49 for locus of control [16,17]. These values

62 are lower than those reported for cognitive skills (c.f. 0.52 for digit span to 0.82 for reading) [16],

63   suggesting that non-cognitive skills may be more variable or have higher measurement error than

64   cognitive skills [3].

65   Few studies have measured a wide battery of non-cognitive skills, but modest correlations have

66   been observed between some non-cognitive skills including grit and conscientiousness [18]; academic

67   effort and academic problems [6]; and internalising and externalising behaviours [19]. While modest [6,18],

68   these correlations support a broad multidimensional definition of non-cognitive skills. Due to data

69   limitations, many studies have also been unable to adjust estimates for cognitive ability despite its

70   strong attenuating effect on non-cognitive associations with outcomes [10,11]. More evidence is

71   required about the relationship between non-cognitive skills in childhood and later outcomes, and

72   the factors that mediate these effects.

73   The measurement consistency of non-cognitive skills and the robustness of their associations with

74   socioeconomic outcomes can also be explored with genetic data. Identifying signal in genetic

75   analyses of a phenotype requires both a genotypic component and reliable measurement of the

76   phenotype. Personality types have been the most studied of the non-cognitive skills with heritability

77   estimated at ~0.5 [20]. Heritability estimates for other non-cognitive skills has varied widely from 0.44

78   to 0.79 for self-control [21,22]; 0.36 for alienation [6]; 0.83 for academic effort [6]; 0.31 to 0.56 for aspects

79   of openness [23]; 0.18 to 0.49 for aspects of conscientiousness [24]; and 0.40 for enjoyment and self-

80   perceived ability [25]. Sibling correlations for non-cognitive skills have also been observed [26]. There is

81   also genetic evidence that personality types are stable over time [27], providing support that

82   personality measures are sufficiently free from measurement error to isolate genetic signal. There

83   has however not yet been a comprehensive analysis investigating the genetics of multiple non-

84   cognitive skills within the same sample of individuals [28].

85   In this study, we contribute to the literature with a comprehensive analysis into the phenotypic and

86   genotypic measurement consistency of non-cognitive skills. Using data from a UK cohort study, we

87   investigate the relationships between 10 non-cognitive skills, educational achievement and labour

4

88   market outcomes. We answer two related research questions: 1) How consistent are non-cognitive

89   skills measured in a large cohort study over time? 2) Do non-cognitive skills associate with

90   socioeconomic outcomes once other factors including cognitive ability have been accounted for?

91

Materials and Methods

Study sample

94   Participants were children from the Avon Longitudinal Study of Parents and Children (ALSPAC).

95   Pregnant women resident in Avon, UK with expected dates of delivery 1st April 1991 to 31st

96   December 1992 were invited to take part in the study. Two phases of recruitment resulted in a total

97   sample of 14,899 children who were alive at one year of age, of whom 7,988 had genetic data. For

98   full details of the cohort profile and study design see [29,30]. The ALSPAC cohort is largely

99   representative of the UK population when compared with 1991 Census data; there is under

100   representation of some ethnic minorities, single parent families, and those living in rented

101   accommodation [29]. Ethical approval for the study was obtained from the ALSPAC Ethics and Law

102   Committee and the Local Research Ethics Committees. We used the largest available samples in each

103   of our analyses to increase precision of estimates, regardless of whether a child has data on other

104   non-cognitive skills (see Supplementary Table 1 for sample sizes).

Genetic data

106   DNA of the ALSPAC children was extracted from blood, cell line and mouthwash samples, then

107   genotyped using references panels and subjected to standard quality control approaches. Briefly,

108   ALSPAC children were genotyped using Illumina HumanHap550 quad chip genotyping platforms and

109   ALSPAC mothers were genotyped using Illumina human660W-quad array platforms. All individuals

110   with non-European ancestry were removed and exclusions were made based on gender mismatches;

111   minimal or excessive heterozygosity; disproportionate levels of individual missingness (>3%);

112   insufficient sample replication (IBD < 0.8); low SNP frequency (<1%), call rate (<95%) and violations

113   of Hardy-Weinberg equilibrium (P < 5E-7). Genotypes in common were combined and imputed to

114   the Haplotype Reference Consortium (HRCr1.1, 2016) panel of approximately 31 000 phased whole

115   genomes, giving 8,237 eligible children with available genotype data after exclusion of related

116   individuals using cryptic relatedness measures. Principal components were generated by extracting

117   unrelated individuals (IBS < 0.05) and independent SNPs with long range LD regions removed. For full

118   details of genotyping see the Supplementary Material.

119   ## Non-cognitive skills

120   We used all non-cognitive skills that were available in ALSPAC except for attention which was

121   omitted due to low responses. The supplementary material contains more detailed information on

122   the measures used, their components and a timeline of when all skills and outcomes were measured

123   (Supplementary Figure 1).

124   ## SDQ

125   Study mothers reported on the Strengths and Difficulties Questionnaire (SDQ) for children on seven

126   occasions at child ages 4, 7, 8, 10, 12, 13 and 16 years, and the children's teachers completed an SDQ

127   questionnaire for each child at ages 7 and 10. The SDQ scale is used to assess child emotional and

128   behavioural difficulties, consisting of 25 items that cover common areas of emotional and

129   behavioural difficulties. Total SDQ score was defined as the count of problems on the first four

130   scales. Sensitivity analyses were run using the internalising and externalising sub-scales. All SDQ

131   scores are reverse coded so that high values refer to fewer problems.

132   ## Denver

133   The Denver Developmental Screening Test was used to identify developmental problems in young

134   children at ages 6, 18, 30 and 42 months. ALSPAC mothers reported on their child's development in

135   response to 42 questions across four different categories: social and communication skills, fine

136   motor skills, hearing and speech, and gross motor skills. Prorated scores combining all four scales

137   were used with missing values assigned the mean score of that child's responses, provided that

138   three or less items had missing scores.

Social skills

140   Social skills at age 13 were determined using a battery of 10 questions reported by the study

141   mother. Responses were reported on a five-point scale and then summed to provide a total overall

142   social skills score.

Communication

144   Communication at 6 months was calculated from mother-reported responses to a battery of eight

145   questions asking about the development of their child's communication skills. At age 1

146   communication was calculated from mother-reported responses using response to 82 questions on

147   the MacArthur Infant Communication questionnaire. At 18 months communication was calculated

148   using mother-reported responses to a battery of 14 questions asking about the development of their

149   child's communication skills. At age 3 communication was calculated from mother-reported

150   responses to a battery of 123 questions forming a vocabulary score. At age 10 communication was

151   calculated from mother-reported responses as the sum of five domains of communication from a

152   total battery of 39 questions.

Self-esteem

154   Self-esteem at age eight was measured using self-report responses to the 12-item shortened form of

155   Harter's Self Perception Profile for Children comprising the global self-worth and scholastic

156   competence subscales. Self-esteem at age 18 was measured using self-report responses to 10

157   questions of the Bachman revision of the Rosenberg Self-Esteem Scale.

Persistence

159   Persistence at age 6 months was measured as a weighted score from mother-reported responses to

160   seven questions relating to child temperament. At age 2 persistence was measured as a weighted

161   score from nine mother-reported responses. At age 7 persistence was recorded as the participants

162   persistence when completing tasks under direct assessment.

163   Locus of control

164   Locus of control, the strength of connection between actions and consequences, was measured at

165   age eight using responses to 12 questions from the shortened version of the Nowicki-Strickland

166   Internal-External (NSIE) scale. At age 16 it was measured using the 12 item Nowicki-Strickland Locus

167   of Control Scale.

168   Empathy

169   Empathy was measured at age seven using mother reported responses to five questions about the

170   child's attitudes towards sharing and caring.

171   Impulsivity

172   Impulsivity was measured during two sessions at the age 8 direct assessment using a behaviour

173   checklist. Testers rated whether the children demonstrated restlessness, impulsivity, fleeting

174   attention, and lacking persistence. At age 11 the children were asked a battery of 10 questions

175   designed to capture impulsive behaviour.

176   Personality

177   Personality was measured at age 13 using the five-factor model of personality to capture five broad

178   and independent dimensions of personality; the "Big Five" (extraversion, neuroticism,

179   agreeableness, conscientiousness, and intellect)[31]. These were measured using self-report responses

180   to 50 items of the International Personality Item Pool.

181   Cognitive skills and outcomes

182   IQ

183   Intelligence was measured during the direct assessments at ages eight and 15 using the short form

184   Wechsler Intelligence Scale for Children (WISC) from verbal, performance, and digit span tests and

8

185   the Wechsler Abbreviated Scale of Intelligence (WASI) from vocabulary and matrix reasoning tests

186   respectively. These assessments were administered by members of the ALSPAC psychology team

187   overseen by an expert in psychometric testing. The short form tests have high reliability and the

188   ALSPAC measures utilise subtests with reliability ranging from 0.70 to 0.96.

189   ## Educational achievement

190   We used four measures of educational achievement. The first three were average fine-graded point

191   scores from three end of 'Key Stage' assessments during compulsory education at ages 11, 14 and

192   16. The fourth measure was a ranking of grades attained in post-compulsory A-levels at age 18,

193   which are required for progression to university education. We used a measure of the three highest

194   A-level grades grouped into ordered categories (see [32] for a detailed description). At the time the

195   cohort were studying, A-levels were non-compulsory and therefore all participants who did not

196   continue into further education were set to missing. All measures were obtained through data

197   linkage to the UK National Pupil Database (NPD) which represents the most accurate record of

198   educational achievement available in the UK. All education data were extracted from the NPD Key

199   Stage 4 (age 16) and Key Stage 5 databases (for further information see

200   https://www.gov.uk/government/collections/national-pupil-database).

201   ## Employment

202   At age 23 participants were asked to report whether they were in full time paid employment of

203   more than 30 hours per week, with responses coded as binary.

204   ## Not in education employment or training (NEET)

205   Because some participants may not be employed because they are still in full-time education or

206   training, we used a measure of not in education employment or training (NEET) at age 23 with

207   responses coded in binary. The use of a NEET measure ensures that employment results are not

208   biased by participation of the cohort in education or training.

209 Income

210 At age 23 participants were asked to report their take-home pay each month if they reported being

211 in paid employment, with responses banded into the following categories: £1-£499; £500-£999;

212 £1000-£1499; £1500-£1999; £2000-£2499; £2500-£3000; £3000+.

213 Non-response

214 We also include a binary measure of questionnaire non-response at ages 18 and 24 to allow us to

215 investigate correlations between non-cognitive skills and cohort participation.

216 Statistical analysis

217 We estimated phenotypic correlations between each measurement-pair (45 measurements [34 non-

218 cognitive skills; 2 cognitive skills; 7 socioeconomic outcomes; 2 non-response measures]; 990 unique

219 measurement-pairs). Heritability of each occasion-specific non-cognitive skill was estimated using

220 genomic-relatedness-based restricted maximum likelihood (GREML) in the software package GCTA

221 (see [33] for a detailed description of this method). GCTA uses measured SNP level variation across all

222 SNPs to estimate the genetic similarity between each pair of unrelated individuals in the sample.

223 Univariate analyses are specified as:

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

224 where y is the inverse normally rank transformed sex and age of measurement standardised

225 measure of phenotype, X is a series of covariates indicating the first 20 principal components of

226 inferred population structure to control for systematic differences in allele frequencies due to

227 ancestral differences between different subpopulations (population stratification), g is a normally

228 distributed random effect with variance σ_g^2 denoting the contribution of SNPs, and ϵ is residual error

229 with variance σ_ϵ^2. Heritability is then defined as the proportion of phenotypic variance that can be

230 statistically explained by common genetic variation while holding inferred population structure

231 constant:

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

232   To estimate the extent to which non-cognitive traits share underlying genetic architecture we

233   estimate genetic correlations between each phenotype-pair. Genetic correlations provide an

234   estimate of the proportion of variance that two phenotypes share due to genetic variation (the

235   overlap of genetic associations between two phenotypes). Genetic correlations are estimated as:

$$r_g = \frac{\sigma_g^2(A, B)}{\sqrt{\sigma_g^2(A)\sigma_g^2(B)}} \#(3)$$

236   Where r_g is the genetic correlation between phenotypes A and B, $\sigma_g^2(A)$ is the genetic variance of

237   phenotype A and $\sigma_g^2(A, B)$ is the genetic covariance between phenotypes A and B. All analyses are

238   adjusted for false discovery rate using the Benjamini-Hochberg procedure [34] and include the 20

239   principal components of population structure.

Results

How consistent are non-cognitive skills over time?

242   Behavioural skills measured on the SDQ scale were consistent over time (parent reported SDQ

243   r=0.36 to 0.73; teacher reported SDQ r=0.28 to 0.53), though consistency decreased with greater

244   elapsed time between measures increased (Figure 1). Correlations between parent reported and

245   teacher reported SDQ measures were weak even when measured contemporaneously. Correlations

246   for other non-cognitive measures over time were generally low (Figure 1), suggesting within-trait

247   inconsistency (highly variable skills) or lack of reliability (high measurement error). Correlations

248   amongst the Big 5 personality types were low except between the intellect/imagination and

249   agreeableness subscales (r=0.46). While the phenotypic correlations for repeat measures of many

250   non-cognitive skills were low, they were generally positive. The weak temporal phenotypic

251   correlations of non-cognitive measures contrast sharply to those for cognitive measures of IQ

252   (r=0.60) and measures of educational achievement (r>0.78 for compulsory education).

253

254   < Figure 1 here >

255

256   How correlated are different non-cognitive skills?

257   Phenotypic correlations between measures of different non-cognitive skills were generally low

258   (Figure 1). Parent-reported and teacher-reported SDQ was the only measure that correlated

259   consistently with other non-cognitive skills at r>0.2. This between-trait correlation was strongest for

260   social skills (r =0.24 to 0.49), communication at age 10 (r =0.29 to 0.56) and empathy (r =0.17 to

261   0.38). Between-skill phenotypic correlations were almost exclusively positive, suggesting that where

262   patterns were observed, children who scored high on one non-cognitive skill generally also scored

263   highly on another. Teacher reported measures of SDQ, locus of control at age 8 and the

264   intellect/imagination personality type correlated modestly with cognitive skills (r>0.3 for IQ at age

265   8). To further explore if measurement imprecision may be driving these low correlations, we took

266   the mean value of non-cognitive skills that had been measured more than once (Figure 2). Weak

267   correlations were observed between SDQ, Teacher SDQ, communication, self-esteem, locus of

268   control and impulsivity. Positive correlations between these mean variables and educational

269   achievement were observe for all variables other than persistence.

270

271   < Figure 2 here >

272

Do non-cognitive skills associate with education and labour market outcomes?

273

274 Phenotypic correlations with educational achievement were highest for the SDQ scale (r=0.13 to

275 0.42) (Figure 1). The teachers' SDQ assessments were more strongly associated than the

276 contemporaneous parent reports. For example, the correlation between educational achievement at

277 age 11 and teacher reported SDQ at age 10 was r=0.40 while the correlation with parent reported

278 SDQ at age 10 was r=0.29. Correlations with educational outcomes were modest for social skills

279 (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;

280 r=0.19 to 0.28 at age 16), the agreeableness and the intellect/imagination subscales of the Big 5

281 (r=0.16 to 0.28 and r=0.27 to 0.39 respectively). Correlations between educational achievement and

282 cognitive ability were considerably higher (r=0.43 to 0.73). Phenotypic correlations for non-cognitive

283 skills with employment, NEET and income were low (r<=0.08 for employment; r<=0.07 for NEET;

284 r<=0.19 for income). Correlations between labour market outcomes and cognitive skills were also

285 very low (r<=0.09). Non-cognitive and cognitive skills were generally weakly negatively correlated

286 with non-response at ages 18 and 24 (r=0.05 to 0.15). Only the extraversion subscale of the Big 5

287 was positively correlated with non-response (r=0.06).

288 To further investigate the potential impact of skills we ran a series of multivariable regressions of

289 age 16 educational achievement on non-cognitive and cognitive skills (Figure 3). Each skill was

290 standardised and analysed independently controlling for sex, month of birth and IQ at age 8. A one

291 standard deviation (SD) increase in non-cognitive skill measures was associated with a 0.04 SD

292 decrease to 0.21 SD increase in age 16 achievement. There was considerable heterogeneity for

293 estimates between skills and measurement occasions for many non-cognitive measures.

294 Associations were generally larger the closer in time non-cognitive skills were measured to

295 educational achievement. By comparison, a one SD increase in cognitive skills was consistently

296 associated with a 0.5 SD increase in achievement. These patterns were similar for educational

297 achievement at all ages (Supplementary Figure 2).

298

299   < Figure 2 here >

300

Are non-cognitive skills heritable?

302   Figure 4 displays the heritability estimates for all our measures. The parent reported SDQ scale was

303   the only non-cognitive measure for which there was reasonable evidence of non-zero heritability at

304   multiple occasions (h^2_{SNP}=0.10 to 0.23). Results were broadly comparable when using the

305   internalising and externalising subscales of the SDQ (Supplementary Figure 3). Non-zero heritability

306   was also observed for communication at age 18 months (h^2_{SNP}=0.17), self-esteem at age 18

307   (h^2_{SNP}=0.25), locus of control at age 8 (h^2_{SNP}=0.21), and the intellect/imagination subscale of

308   personality type (h^2_{SNP}=0.21). There was imprecision in the heritability point estimates, but the

309   upper limit of the 95% confidence intervals were estimated below 0.3 for most non-cognitive skills.

310   Heritability of the mean value non-cognitive skills was estimated with greater precision, but there

311   was evidence of heritability for the mean responses of SDQ, (h^2_{SNP}=0.20), communication

312   (h^2_{SNP}=0.13), self-esteem and locus of control (h^2_{SNP}=0.22) (Figure 5). The heritability of cognitive

313   skills was higher at 0.43 at age 8 and 0.47 at age 15. Educational outcomes were highly heritable

314   (h^2_{SNP}>0.4) but there was little evidence of heritability for the labour market outcomes. Heritability

315   of questionnaire non-response was estimated higher and with less uncertainty than the non-

316   cognitive measures (h^2_{SNP}=0.34 and h^2_{SNP}=0.21 at ages 18 and 24 respectively).

317

318   < Figure 3 here >

319   < Figure 4 here >

320

321 ## Do non-cognitive skills have a shared genetic architecture?

322 There was strong evidence for within trait genetic correlations over time for only the parent

323 reported SDQ measures (rG= 0.62 to 1.00) and IQ, which had near-unity genetic overlap (rG= 0.97).

324 There was very limited evidence for genetic correlations across different non-cognitive skills (Figure

325 1, above the diagonal), though estimation precision was low for most skills given the low SNP

326 heritabilities. Genetic correlations between different non-cognitive measures were only observed

327 between the parent reported SDQ measures and communication at age 10 (rG=0.68 to 0.91). This

328 remained the case when using mean values of the non-cognitive skills (Figure 3); genetic correlations

329 were only observed between SDQ and locus of control (r=0.50), SDQ and impulsivity (r=0.76), and

330 self-esteem and locus of control (r=0.50). The SDQ measures, communication at age 10, self-esteem

331 at age 18, the agreeableness and the intellect/imagination subscales of the Big 5 personality types

332 had strong genetic correlations with educational achievement. These were higher for teacher report

333 at age 7 than any of the parents and teacher reported at age 16 SDQ measures. There was strong

334 evidence for non-zero genetic correlations with IQ and education for the mean values of parent SDQ,

335 teacher SDQ and locus of control. Genetic correlations between cognitive measures and educational

336 achievement are all estimated near unity (rG>0.96). There was little evidence of genetic correlations

337 between labour market outcomes and any of the non-cognitive or cognitive skills.

338

339 < Figure 5 here >

340

341 ## Discussion

342 Our results provide longitudinal evidence into the consistency of a large range of non-cognitive skill

343 measures throughout childhood and adolescence. Measures of behavioural and communication

344 skills were correlated phenotypically and genotypically over time. Correlations of other measured

345   non-cognitive skills over time were low, contrasting with previous research that has evidenced

346   temporal stability of non-cognitive skills [14,15]. This heterogeneity may be due to the diversity of non-

347   cognitive skills measured. For example, the SDQ scale used responses to a large battery of validated

348   questions which are likely to be more reliable than other non-cognitive skill measures used here. The

349   increased estimation precision that we observed when using aggregated values of non-cognitive

350   skills supports this. By incorporating information from multiple occasions, aggregate measures are

351   expected to contain less measurement error than individual measures. Future research proposing

352   more detailed measurement of non-cognitive skills may therefore improve the ability of measures to

353   capture underlying skills. The lack of consistency may have also reflected genuine temporal intra-

354   individual variation in the expression of skills or differences over time (e.g. due to schooling). It is

355   important to note there is some variation in the measurement of skills that we used. This is

356   somewhat unavoidable as skill measurements vary depending on the age at which a child is

357   assessed, but measurement artefact may nevertheless reflect some of the variation in

358   measurement. Non-cognitive skills have been argued as promising targets for interventions to

359   improve outcomes because they are potentially more modifiable than cognitive skills [1,3,15]. However,

360   ideally to be suitable for policy interventions, non-cognitive measures must reliably pick up a

361   consistent signal of an underlying skill. While changes to non-cognitive skills are required (and

362   indeed anticipated) during interventions, they should remain relatively stable in individuals that

363   experience no intervention. These findings suggest that many measures of non-cognitive skills may

364   be too inaccurate for accurately measuring the impact of interventions.

365   Correlations between different non-cognitive skills were weak, supporting the notion that these

366   constructs capture empirically different psychosocial phenomena rather than a single underlying or

367   latent non-cognitive factor. The SDQ scale was the only measure to consistently correlate with other

368   non-cognitive skills, conforming to a previous study that found low phenotypic correlations between

369   different non-cognitive skills [6]. Stronger between-trait correlations were observed using aggregated

370   values of the non-cognitive skills, which again may reflect reduced measurement error. The cognitive

16

371   measures we used were more strongly correlated over the same period, though this may reflect that

372   they are more reliably measured than the non-cognitive measures. Future studies that combine

373   longitudinal data on non-cognitive skills with multi-source multi-method approaches may benefit

374   from greater measurement precision.

375   We found limited evidence of genetic correlations within or between non-cognitive skills, with only

376   the parent reported SDQ demonstrating consistent genetic architecture over time. This may reflect

377   the influence of shared parent-offspring genetics on reporting, or parental genetics influencing

378   offspring non-cognitive skills indirectly through dynastic effects [35] (there would be no such shared

379   teacher-student genetics). Previous twin studies have found strong genetic correlations between

380   non-cognitive and cognitive skills [23,24], but our results did not support this. The differences between

381   these results and previous studies may have arisen due to differences in the quality of non-cognitive

382   skill measurements; differences due to cross-sectional and longitudinal study designs; reporting bias

383   by the study mothers (for example where their child performed poorly); differences in the study

384   populations; selective reporting and the use of samples or measures of convenience in previous

385   studies [13]. While we reported a range of non-cognitive skills in ALSPAC, many previous studies have

386   reported only one or a small number of non-cognitive skills.

387   Our results also showed that associations between non-cognitive skills and socioeconomic outcomes

388   were generally weak, contradicting findings from previous individual studies [1–3] but supporting a

389   recent systematic review [13]. Behavioural problems as captured by the SDQ scale, social skills, and

390   locus of control were the only non-cognitive measures to phenotypically associate with educational

391   outcomes strongly and consistently. Associations were stronger between educational achievement

392   and the teacher than the parent reported measures of the SDQ scale. This may suggest that teachers

393   more objectively identify education related problematic behaviour than parents, or that teachers'

394   response to their pupils directly influences their educational outcomes. Non-cognitive skills capture

395   a diverse range of characteristics and may have varying relevance to educational and labour market

396   outcomes; further longitudinal research is required to better elucidate these relationships. Neither

397   non-cognitive nor cognitive skills associated strongly with the labour market outcomes measured

398   here. While this contrasts with previous studies [1,3], our labour market outcomes were observed soon

399   after entry to the labour market and therefore may have more closely resembled institutional

400   effects than non-cognitive skills. Some have argued that non-cognitive skills and personality traits

401   are as important as cognitive skills for many dimensions of behaviour and socioeconomic outcomes

402   [5], but our results do not support this. The standardized effect sizes of non-cognitive skills were at

403   most a third that of cognitive skills.

404   Many non-cognitive and cognitive skills were weakly negatively correlated with non-response

405   phenotypically, suggesting that individuals who scored low on these skills were more likely to later

406   drop out of the ALSPAC study. This may have important implications for participant

407   representativeness and generalisability in cohort studies. Cognitive ability and achievement were

408   negatively genotypically correlated with non-response, adding to the growing body of evidence that

409   non-response is genetically patterned [36].

410   The SDQ scale was the only non-cognitive measure for which we consistently found strong evidence

411   of heritability (~0.15), far lower than the estimated heritabilties for cognitive measures (~0.45) and

412   educational achievement (~0.50). Heritability and stability depend on the developmental period

413   being examined and so comparison of estimates is difficult. Previous research has demonstrated that

414   the heritability of cognitive ability rises with age while the heritability of personality traits decreases

415   with age [37], which our data contradicts. Many of the genetic correlations we observed were

416   imprecise due to the low heritability estimates, but our estimates suggested an upper bound

417   heritability of 0.3 for most non-cognitive skills.

418   This study has several limitations. First, it is possible that measurement error was unusually high in

419   the non-cognitive measures used in the ALSPAC study. However, the measures used in ALSPAC have

420   been widely validated are consistent with those used in previous studies [19]. Furthermore,

18

421   measurement error would need to have been high across all measures used from birth to age 18 so

422   to explain these results. Our findings were consistent when using aggregated measures of non-

423   cognitive skills where they had been measured at least twice, suggesting that our results were not

424   driven by differential measurement error across occasions. Future studies into test-retest reliability

425   of non-cognitive skills based on different longitudinal samples could help resolve these questions.

426   Second, many of the genetic correlations were estimated with extremely low precision, often being

427   constrained at the values of -1 or 1 (see supplementary Table 2). This is likely due to the low

428   estimated heritability of the non-cognitive skills. Low heritability implies a small genotypic

429   contribution (either in the number of individual variants associated or the strength of associations)

430   and therefore lower power to detect genetic correlations between these skills. Our sample sizes are

431   fairly low to detect such small univariate and bivariate genetic associations. Future studies

432   conducted on larger samples are required to more accurately the estimate heritability of, and

433   genetic correlations between non-cognitive skills and other phenotypes. Third, our genetic

434   associations may have been biased by uneven linkage disequilibrium, residual population structure,

435   or assortative mating [35,38]. We controlled for the first twenty principle components of population

436   structure to account for population structure but this may not have accounted for all differences [39].

437   Assortative mating is thought to be low for non-cognitive traits [20], but will have inflated our genetic

438   associations if present [40]. It is possible that assortment on non-cognitive skills may be negative and

439   future work is required to determine this. Fourth, the use of labour market outcomes at age 23 may

440   mean that some of our study participants have not yet transitioned into their stable career

441   employment. While our use of a NEET outcome variable reduces the problem that those still in

442   education or unemployed have not yet entered the labour market, it does not provide any indication

443   that employed study participants have entered desired or long-term employment. Finally, it is

444   possible that some of the measures could reflect parental rather than child genes where they are

445   parent rather than self-reported.

446   In conclusion, our results highlight that non-cognitive skills are likely to be highly heterogenous.

447   Measures of noncognitive skills were varied both over time and across different measures in the

448   same individuals, but some individual measures such as the SDQ demonstrated strong internal

449   consistency throughout childhood. Furthermore, many non-cognitive measures associated weakly

450   with educational and employment outcomes at entry to the labour market, particularly when

451   measured early in childhood.

452

478   http://www.bris.ac.uk/acrc/ and the Research Data Storage Facility of the University of Bristol -

479   http://www.bris.ac.uk/acrc/storage/.

Conflict of interest
480

481   The authors declare no conflict of interest.

References
482

483   1.   Brunello G, Schlotter M. Non Cognitive Skills and Personality Traits⊡: Labour Market

484   Relevance and their Development in Education & Training Systems. *IZA Discuss Pap*. 2011.

485   2.   Heckman JJ, Kautz T. Hard evidence on soft skills. *Labour Econ*. 2012.

486   doi:10.1016/j.labeco.2012.05.014

487   3.   Kautz T, Heckman JJ, Diris R, et al. *Fostering and Measuring Skills: Improving Cognitive and*

488   *Non-Cognitive Skills to Promote Lifetime Success*.; 2014. doi:10.1787/19939019

489   4.   Carneiro P, Crawford C, Goodman A. The Impact of Early Cognitive and Non-Cognitive Skills

490   on Later Outcomes. *Cent Econ Educ*. 2007.

491   5.   Heckman JJ, Stixrud J, Urzua S. The Effects of Cognitive and Noncognitive Abilities on Labor

492   Market Outcomes and Social Behavior. *J Labor Econ*. 2006;24(3):411-482.

493   doi:10.1086/504455

494   6.   McGue M, Rustichini A, Iacono WG. Cognitive, Noncognitive, and Family Background

495   Contributions to College Attainment: A Behavioral Genetic Perspective. *J Pers*. 2015;85(1):65-

496   78. doi:10.1111/jopy.12230

497   7.   Borghans L, Golsteyn BHH, Heckman J, Humphries JE. Identification problems in personality

498   psychology. *Pers Individ Dif*. 2011. doi:10.1016/j.paid.2011.03.029

499   8.   Bowles S, Gintis H. *Schooling in Capitalist America: Educational Reform and the*

500   *Contradictions of Economic Life*.; 1976. doi:10.2307/3090251

501   9.   O'Connor MC, Paunonen S V. Big Five personality predictors of post-secondary academic

502   performance. *Pers Individ Dif*. 2007;43(5):971-990.

503   10.   Schmidt FL, Hunter JE. The validity and utility of selection methods in personnel psychology:

504   Practical and theoretical implications of 85 years of research findings. *Psychol Bull*.

505   1998;124(2):262.

506   11.   Von Stumm S, Hell B, Chamorro-Premuzic T. The hungry mind: Intellectual curiosity is the

507   third pillar of academic performance. *Perspect Psychol Sci*. 2011;6(6):574-588.

508   12.   Soto CJ. How replicable are links between personality traits and consequential life outcomes?

509   The Life Outcomes Of Personality Replication Project. *Psychol Sci*. 2019.

510   13.   Smithers LG, Sawyer ACP, Chittleborough CR, Davies NM, Davey Smith G, Lynch JW. A

511   systematic review and meta-analysis of effects of early life non-cognitive skills on academic,

512   psychosocial, cognitive and health outcomes. *Nat Hum Behav*. 2018;2(11):867-880.

513   doi:10.1038/s41562-018-0461-x

514   14.   Cobb-Clark DA, Schurer S. Two economists' musings on the stability of locus of control. *Econ*

515   *J*. 2013. doi:10.1111/ecoj.12069

516   15.   Roberts BW, DelVecchio WF. The rank-order consistency of personality traits from childhood

517   to old age: A quantitative review of longitudinal studies. *Psychol Bull*. 2000.

518   doi:10.1037/0033-2909.126.1.3

519   16.   Laajaj R, Macours K. Measuring Skills in Developing Countries. *Policy Res Work Pap*. 2017.

520   doi:doi:10.1596/1813-9450-8000

521   17.   Dohmen T, Falk A, Huffman D, Sunde U, Schupp J, Wagner GG. Individual risk attitudes:

522   Measurement, determinants, and behavioral consequences. *J Eur Econ Assoc*. 2011;9(3):522-

523   550. doi:10.1111/j.1542-4774.2011.01015.x

524   18.   Duckworth AL, Quinn PD. Development and validation of the short Grit Scale (Grit-S). *J Pers*

525   *Assess*. 2009. doi:10.1080/00223890802634290

526   19.   Fleche S. Teacher Quality, Test Scores and Non-Cognitive Skills: Evidence from Primary School

527   Teachers in the UK. 2017. http://eprints.lse.ac.uk/83602/.

528   20.   Bouchard TJ, Loehlin JC. Genes, evolution, and personality. *Behav Genet*. 2001.

529   doi:10.1023/A:1012294324713

530   21.   Willems YE, Dolan C V, van Beijsterveldt CEM, et al. Genetic and Environmental Influences on

531   Self-Control: Assessing Self-Control with the ASEBA Self-Control Scale. *Behav Genet*.

532   2018;48(2):135-146. doi:10.1007/s10519-018-9887-1

533   22.   Lemery-Chalfant K, Doelger L, Goldsmith HH. Genetic relations between effortful and

534   attentional control and symptoms of psychopathology in middle childhood. *Infant Child Dev*

535   *An Int J Res Pract*. 2008;17(4):365-385.

536   23.   Wainwright M a, Wright MJ, Luciano M, Geffen GM, Martin NG. Genetic covariation among

537   facets of openness to experience and general cognitive ability. *Twin Res Hum Genet*. 2008.

538   doi:10.1375/twin.11.3.275

539   24.   Luciano M, Wainwright MA, Wright MJ, Martin NG. The heritability of conscientiousness

540   facets and their relationship to IQ and academic achievement. *Pers Individ Dif*. 2006.

541   doi:10.1016/j.paid.2005.10.013

542   25.   Kovas Y, Garon-Carrier G, Boivin M, et al. Why children differ in motivation to learn: Insights

543   from over 13,000 twins from 6 countries. *Pers Individ Dif*. 2015.

544   doi:10.1016/j.paid.2015.02.006

545   26.   Anger S, Schnitzlein DD. Cognitive skills, non-cognitive skills, and family background: evidence

546   from sibling correlations. *J Popul Econ*. 2017. doi:10.1007/s00148-016-0625-9

547   27.   Briley DA, Tucker-Drob EM. Genetic and environmental continuity in personality

548   development: A meta-analysis. *Psychol Bull*. 2014. doi:10.1037/a0037091

549   28.   Tucker-Drob EM, Harden KP. A behavioral genetic perspective on noncognitive factors and

550   academic achievement. 2015.

551   29.   Boyd A, Golding J, Macleod J, et al. Cohort Profile: the 'children of the 90s'--the index

552   offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*.

553   2013;42(1):111-127. doi:10.1093/ije/dys064

554   30.   Fraser A, Macdonald-wallis C, Tilling K, et al. Cohort profile: The avon longitudinal study of

555   parents and children: ALSPAC mothers cohort. *Int J Epidemiol*. 2013;42:97-110.

556   doi:10.1093/ije/dys066

557   31.   BARRICK MR, MOUNT MK. THE BIG FIVE PERSONALITY DIMENSIONS AND JOB

558   PERFORMANCE: A META-ANALYSIS. *Pers Psychol*. 1991. doi:10.1111/j.1744-

559   6570.1991.tb00688.x

560   32.   Morris TT, Dorling D, Davey Smith G. How well can we predict educational outcomes?

561   Examining the roles of cognitive ability and social position in educational attainment.

562   *Contemp Soc Sci*. 2016;11(2-3):1-15. doi:10.1080/21582041.2016.1138502

563   33.   Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: A tool for genome-wide complex trait

564   analysis. *Am J Hum Genet*. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011

565   34.   Benjamini Y, Hochberg Y. Controlling the false discovery rate: a practical and powerful

566   approach to multiple testing. *J R Stat Soc Ser B*. 1995. doi:10.2307/2346101

567   35.   Kong A, Thorleifsson G, Frigge ML, et al. The nature of nurture: Effects of parental genotypes.

568   *Science (80-)*. 2018;359(6374):424-428. doi:10.1126/science.aan6877

569   36.   Taylor A, Jones H, Sallis H, et al. Exploring the association of genetic factors with participation

570   in the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2018;In press.

571   37.   Briley DA, Tucker-Drob EM. Comparing the Developmental Genetics of Cognition and

572   Personality over the Life Span. *J Pers*. 2017;85(1):51-64. doi:doi:10.1111/jopy.12186

573   38.   Young AI, Frigge ML, Gudbjartsson DF, et al. Relatedness disequilibrium regression estimates

574   heritability without environmental bias. *Nat Genet*. 2018. doi:10.1038/s41588-018-0178-9

575   39.   Haworth S, Mitchell R, Corbin L, et al. Apparent latent structure within the UK Biobank

576   sample has implications for epidemiological analysis. *Nat Commun*. 2019;10(1):333.

577   doi:10.1038/s41467-018-08219-1

578   40.   Redden DT, Allison DB. The effect of assortative mating upon genetic association studies:

579   Spurious associations and population substructure in the absence of admixture. *Behav Genet*.

580   2006. doi:10.1007/s10519-006-9060-0

581

582

583   # Figure legends

584   **Figure 1: Heritability, phenotypic correlations and genetic correlations of skills and outcomes**. Age

585   in parentheses. SDQ, strengths and difficulties; NEET, not in education, employed or training. Values

586   on the diagonal represent univariate heritabilities; values below the diagonals represent phenotypic

587   correlations; values above the diagonal represent genetic correlations. Multiple testing was handled

588   using an FDR threshold of 5%. Empty cells display correlations that fell below FDR threshold. Black

589   outline boxes indicate the same skills measured at different occasions. See Supplementary Tables 2

590   & 3 for full estimates.

591

592   **Figure 2: Heritability, phenotypic correlations and genetic correlations of aggregate skills and**

593   **outcomes**. Age in parentheses. SDQ, strengths and difficulties; NEET, not in education, employed or

594   training. Values on the diagonal represent univariate heritabilities; values below the diagonals

26

595   represent phenotypic correlations; values above the diagonal represent genetic correlations.

596   Multiple testing was handled using an FDR threshold of 5%. Empty cells display correlations that fell

597   below FDR threshold. Black outline boxes indicate the same skills measured at different occasions.

598   See Supplementary Tables 4 & 5 for full estimates.

599

600   **Figure 3: Associations between skills and educational attainment adjusted for sex, month of birth**

601   **and IQ at age 8**. Educational attainment measured as exam point score at age 16. Age (years) in

602   parentheses. SDQ, strengths and difficulties.

603

604   **Figure 4: Heritability of skills and outcomes.** Age in parentheses. SDQ, strengths and difficulties;

605   NEET, not in education, employed or training. Dark shaded bars represent estimates below FDR

606   threshold. Heritability estimated using GCTA-GREML on the full sample available for each

607   phenotype. See Supplementary Table 1 for full estimates.

608

609   **Figure 5: Heritability of aggregate skills and outcomes**. Age in parentheses. SDQ, strengths and

610   difficulties; NEET, not in education, employed or training. Dark shaded bars represent estimates

611   below FDR threshold. Heritability estimated using GCTA-GREML on the full sample available for each

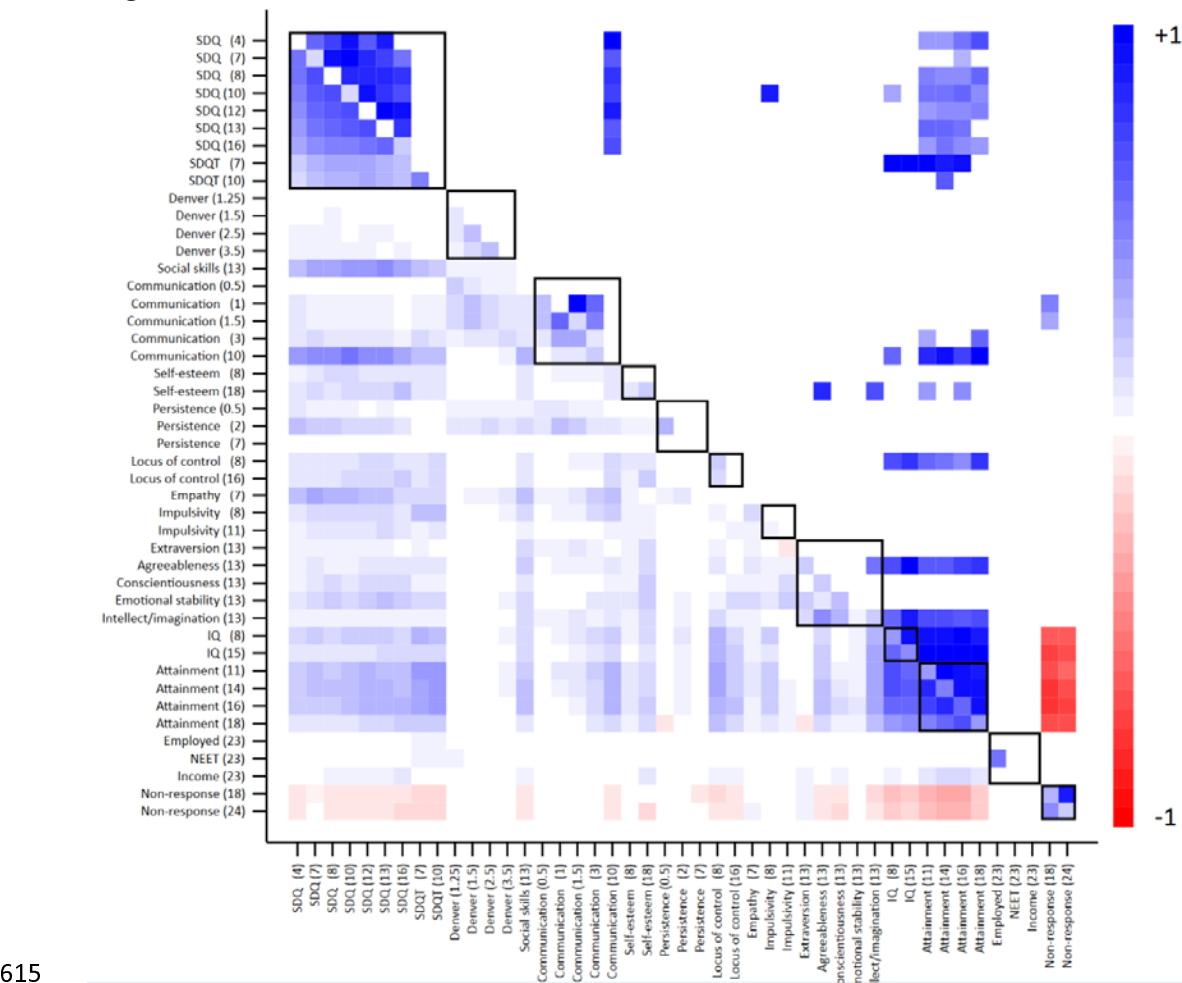612   phenotype. See Supplementary Table 1 for full estimates.

613

Figures

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. See Supplementary Tables 2 & 3 for full estimates.
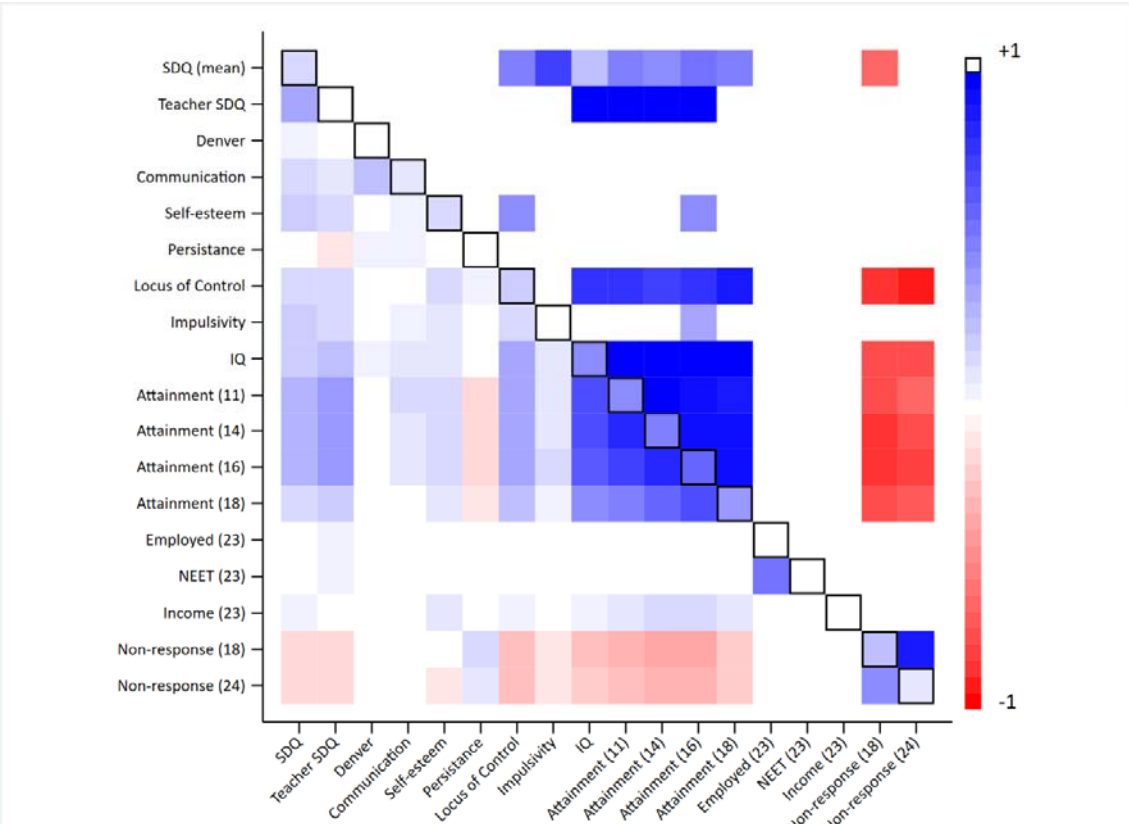
Figure 2: **Heritability, phenotypic correlations and genetic correlations of aggregate skills and outcomes. Age in parentheses.** SDQ, strengths and difficulties; NEET, not in education, employed or training. Values on the diagonal represent univariate heritabilities; values below the diagonals represent 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. See Supplementary Tables 4 & 5 for full estimates.
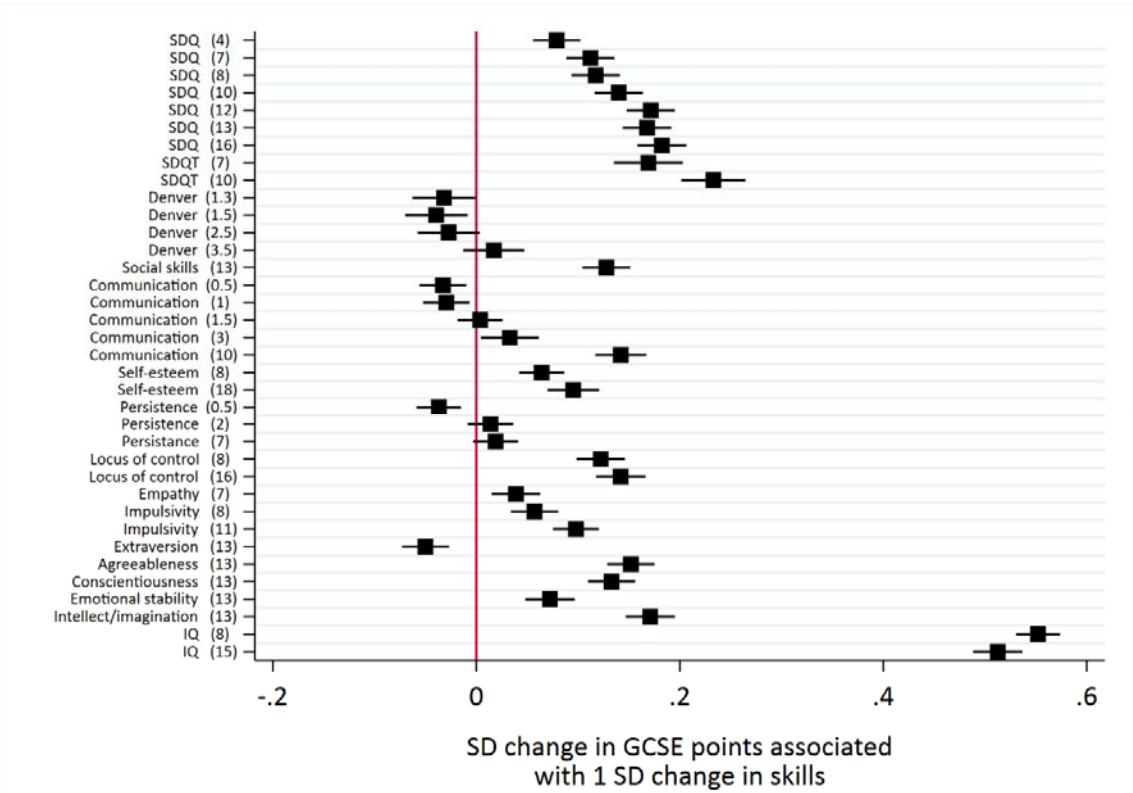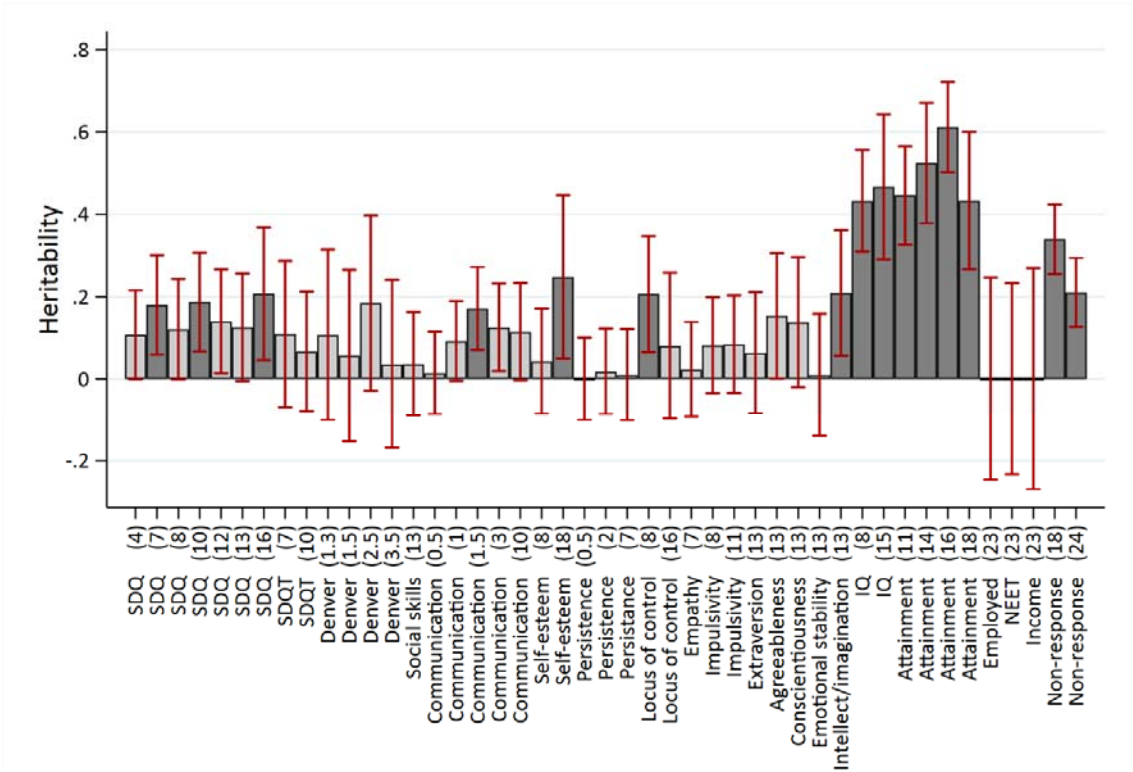
Figure 3: **Associations between skills and educational attainment adjusted for sex, month of birth and IQ at age 8.** Educational attainment measured as exam point score at age 16. Age (years) in parentheses. SDQ, strengths and difficulties.

637
638   Figure 4: **Heritability of skills and outcomes.** Age in parentheses. SDQ, strengths and difficulties;
639   NEET, not in education, employed or training. Dark shaded bars represent estimates below FDR
640   threshold. Heritability estimated using GCTA-GREML on the full sample available for each
641   phenotype. See Supplementary Table 1 for full estimates.
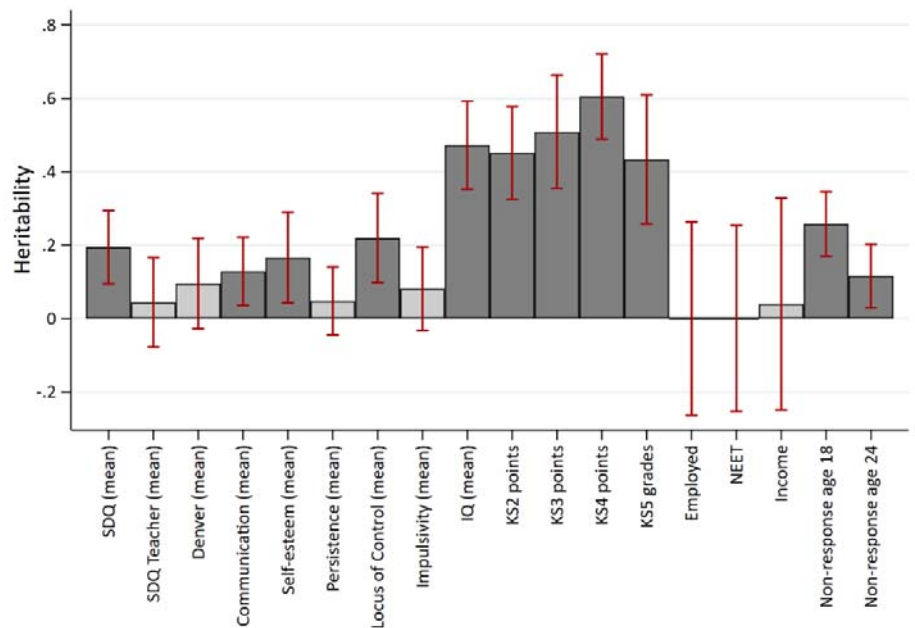
642

643

644

Figure 5: **Heritability of aggregate skills and outcomes.** Age in parentheses. SDQ, strengths and difficulties; NEET, not in education, employed or training. Dark shaded bars represent estimates below FDR threshold. Heritability estimated using GCTA-GREML on the full sample available for each phenotype. See Supplementary Table 1 for full estimates.

645
646
647
648

649