ARTICLE   OPEN

The stability of educational achievement across school years is largely explained by genetic factors

Kaili Rimfeld [1], Margherita Malanchini[1,2], Eva Krapohl[1], Laurie J. Hannigan [1], Philip S. Dale [3] and Robert Plomin[1]

Little is known about the etiology of developmental change and continuity in educational achievement. Here, we study achievement from primary school to the end of compulsory education for 6000 twin pairs in the UK-representative Twins Early Development Study sample. Results showed that educational achievement is highly heritable across school years and across subjects studied at school (twin heritability ~60%; SNP heritability ~30%); achievement is highly stable (phenotypic correlations ~0.70 from ages 7 to 16). Twin analyses, applying simplex and common pathway models, showed that genetic factors accounted for most of this stability (70%), even after controlling for intelligence (60%). Shared environmental factors also contributed to the stability, while change was mostly accounted for by individual-specific environmental factors. Polygenic scores, derived from a genome-wide association analysis of adult years of education, also showed stable effects on school achievement. We conclude that the remarkable stability of achievement is largely driven genetically even after accounting for intelligence.

INTRODUCTION

Educational achievement is important to society and to children as individuals. In fact, educational achievement has been shown to be a good predictor of many life outcomes, such as occupational status, happiness, health, and even life expectancy.[1–5] Influences on educational achievement, including genetic and environmental etiologies, can best be studied during the period of compulsory education when the full range of family characteristics is represented. Compulsory education in the UK culminates with standardized nation-wide exams, the General Certificate of Secondary Education (GCSE). GCSE grades are a gateway to further education, university acceptance, and even later employment, shaping individuals' life-long educational and professional trajectories. Previous twin research has shown that GCSE performance is highly heritable, and to a lesser extent explained by environmental factors.[6] However, little is known about whether the same or different genetic and environmental effects contribute to individual differences in achievement over the course of compulsory education. In the present paper, quantitative (twin) and molecular genetic (DNA) methods are used to examine the etiology and developmental course of educational achievement during the primary and secondary education period, culminating in GCSE grades.

There is now converging evidence for the heritability of educational achievement across school years using family designs, such as twin and adoption studies, and DNA-based methods. Twin studies have shown that around 60% of individual differences in school achievement are explained by inherited differences in children's DNA sequence.[6–16] This holds when considering overall achievement scores as well as separate school subjects, from Sciences to Humanities.[11,12] It is also possible to estimate heritability using DNA of unrelated individuals, where small DNA

differences between individuals (single-nucleotide polymorphisms (SNPs)) are associated with the individuals' scores in a trait of interest. Rather than estimating the association between each SNP and the trait, this method estimates the association between the trait and all the SNPs combined.[17,18] This so-called SNP heritability for educational achievement has been shown to be around 20–30%.[12,19–21] The SNP heritability is less than that estimated by twin studies partly because SNP heritability is limited to additive effects of common SNPs that are included in current arrays used to genotype SNPs. Because genome-wide association (GWA) studies have the same limitations as SNP heritability, SNP heritability is the current ceiling for the phenotypic variance that GWA studies can explain.

These univariate genetic analyses have shown that the heritability of educational achievement is substantial and consistent across school years, from primary to secondary education and culminating in the GCSEs.[6,9] However, that conclusion is agnostic about the extent to which the same or different genetic factors contribute to individual differences in educational achievement longitudinally from age to age, that is, to stability and change. Understanding the developmental etiology of educational achievement in this way has considerable potential for illuminating the mechanisms that trigger differences in GCSE performance and, consequently, in educational and professional outcomes.

Multivariate genetic methods can be used to address this question of the etiology of age-to-age stability and change. Using a multivariate twin design we have previously demonstrated that, during the primary school years, genetic and shared environmental factors show substantial stability in English, Mathematics, and Science, while non-shared environmental factors contribute to change.[9] However, the genetic and environmental etiology of

[1]Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, Psychology and Neuroscience, King's College London, London, UK; [2]Department of Psychology, University of Texas at Austin, Austin, USA and [3]Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, USA
Correspondence: Kaili Rimfeld (Kaili.rimfeld@kcl.ac.uk)
These authors jointly supervised this work: Kaili Rimfeld, Margherita Malanchini

np̄j

The stability of educational achievement across school…
K Rimfeld et al.

2

stability and change of educational achievement across the longer span of school years, from primary school to secondary education and beyond, remains unexplored. Only a few longitudinal studies of reading ability have been reported. In one study, the stability of reading, measured as word recognition, was explained largely by genetic factors (around 70%) from primary through secondary school.[22] Another study found that the etiology of reading fluency across the first five years of schooling, an important developmental time when students transition from 'learning to read' to 'reading to learn', was characterized by stable genetic and shared environmental influences.[23] Two additional longitudinal analyses of reading comprehension in two different samples from the UK[24] and US[25] also showed substantial genetic stability. However, school achievement involves much more than reading.

To our knowledge, no longitudinal analysis has been conducted to assess the genetic and environmental etiology of continuity and change of educational achievement throughout compulsory education, for specific subjects as well as for general educational achievement. This is the purpose of the current study, which uses longitudinal data from age 7 to 16 on educational achievement from a UK-representative sample of 6000 twin pairs participating in the Twins Early Development Study (TEDS).[26]

We also addressed the issue of stability and change in school achievement, for the first time using DNA-based analyses. In addition to SNP heritability, which was described earlier, another recently developed method predicts academic achievement directly from DNA using specific SNPs that have been shown to be associated with the trait in GWA analyses. This method aggregates thousands of SNP associations, which individually have very small effects, into a genome-wide polygenic score (GPS)[27] with effects weighted by results from the GWA discovery sample. A GPS can be used to predict variance in a trait for unrelated individuals in a sample independent of the GWA discovery sample. We will refer to this estimate as GPS heritability. It explains less variance than SNP heritability or twin study heritability because GPS heritability predicts educational achievement from specific SNPs.

Our *EduYears* GPS was derived from a GWA study of years of education for 300,000 individuals.[28] We used the GWA summary data to create an *EduYears* GPS for each of 6000 unrelated individuals (one member of a twin pair) in our TEDS sample [26] in the UK. We correlated *EduYears* GPS with achievement measures at ages 7, 9, 12, and 16. We have previously shown that *EduYears* GPS predicts up to 9% of the variance in GCSE scores;[29] here we extend this analysis and investigate results for specific subjects in addition to general achievement. The focus of our present analyses is the extent to which the *EduYears* GPS contributes to stability of educational achievement.

Genetic stability of school achievement might be explained fully or in part by general cognitive ability (*g*), which has also been shown to be substantially heritable[10,30,31] and developmentally stable,[32] and is one of the strongest predictors of school achievement.[33–36] Moreover, the links between achievement and *g* have shown to be explained by genetic factors.[7,33,37] Because *g* is a likely candidate to explain stability of school achievement across compulsory education, we also investigate the role of *g* in the stability of educational achievement, using both the twin design and DNA-based methods.

In summary, in this study we use twin analyses and GPS analyses of longitudinal data from TEDS from age 7 to age 16, including GCSE scores, to investigate three issues—the stability of general educational achievement, the stability of achievement in specific subjects, and the contribution of *g* to the stability of educational achievement.

RESULTS

Phenotypic analyses
Means and standard deviations were calculated for school achievement across compulsory education for the whole sample, males and females separately, and for all five sex and zygosity groups: monozygotic (MZ) males, dizygotic (DZ) males, MZ females, DZ females, and DZ opposite-sex twin pairs. One twin per pair was randomly selected for phenotypic analyses to maintain independence of data. Analyses of variance (ANOVA) were used to test the significance of these group differences. ANOVA results showed some significant sex differences, however, sex and zygosity together explain only 1% of variance in achievement on average (Supplementary Table 1). For subsequent analyses, the data were corrected for mean sex differences, as described in the Methods section.
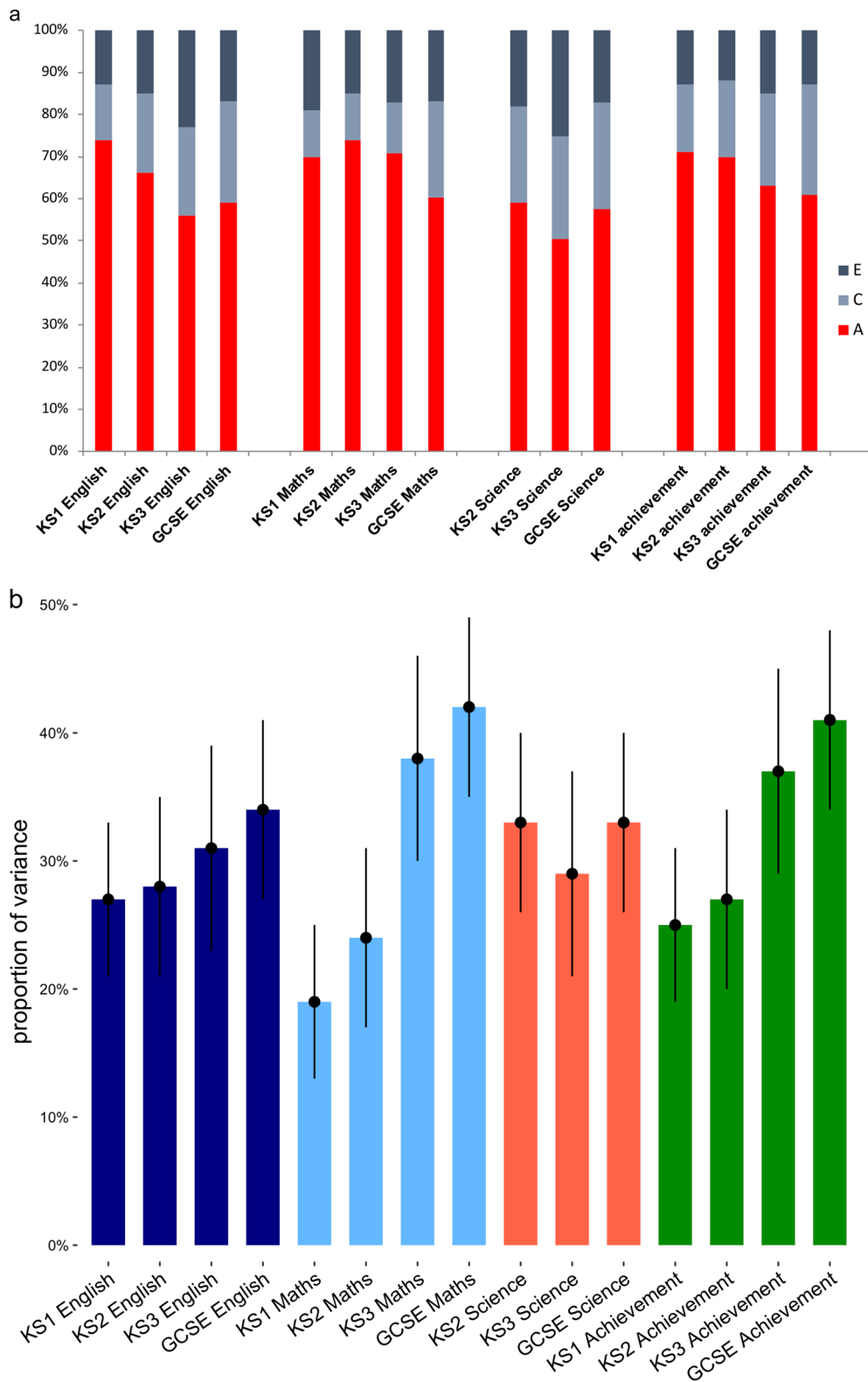
Genetic analyses
Univariate genetic analyses. Figure 1a presents the twin ACE estimates for achievement across development. All achievement measures show substantial heritability (A ~60%). Shared (C) and non-shared (E) environmental factors both explained about 20% of the variance. Estimates did not vary systematically across subjects or school years. Twin intra-class correlations and parameter estimates with confidence intervals are presented in Supplementary Table 2, which shows that parameter estimates were also similar for teacher ratings, exam performance, and achievement scores that combined teacher ratings and exam performance.

SNP heritabilities were calculated for the same achievement measures using the GCTA package (see Methods). Figure 1b shows that SNP heritabilities were substantial (~30%) but, as expected, only about half as large as the twin estimates, although there was a trend towards increasing SNP heritability with age. For example, the SNP heritability of Mathematics achievement (composite of test scores and teacher ratings) was 19% (SE = 0.06) in KS1 and 38% (SE = 0.08) in KS3 and 42% (SE = 0.07) for GCSE. Twin heritabilities and SNP heritabilities did not differ much across age after the variance accounted for by general cognitive ability (*g*) was controlled for by means of linear regression (Supplementary Figure 1). The trend towards increasing SNP heritability with age seen in Fig. 1 disappeared when controlling for *g* (Supplementary Figure 1(b)), which shows increasing heritability with age.[38]

Multivariate genetic analyses of age-to-age stability. Academic achievement (measured as the mean of English and Mathematics) was highly stable, with age-to-age correlations ranging from 0.66 to 0.85 (Fig. 2a). In bivariate twin analyses comparing each pair of ages, genetic factors accounted for a substantial proportion of the covariance between ages, explaining from 63 to 79% of the phenotypic correlations (Fig. 2a). Controlling for *g* only slightly reduced the phenotypic stability (range = 0.50–0.78) and genetic stability (range = 0.53–0.82) of the correlations (Fig. 2b). The phenotypic stability from age to age was still mostly accounted for by genetic factors, even after controlling for *g* (52–72%; Fig. 2b). Supplementary Table 3 presents the phenotypic and genetic correlations with 95% confidence intervals for the overall achievement and for separate subjects.

Etiological contributions to stability and change were assessed using multivariate models encompassing all ages of assessment. The first was a simplex longitudinal model[39] (see Methods and Supplementary Figure 2 for details). The results, presented in Fig. 3, indicate that the stability of core academic achievement was largely explained by additive genetic (A) factors—the genetic paths from age to age are 0.86, 0.84, and 0.86. C was also stable from age to age, accounting for a smaller proportion of variance in academic achievement, amounting to around one-third of the proportion of variance explained by A. E contributed variance that

The stability of educational achievement across school…
K Rimfeld et al.

npj

3

Fig. 1 a Twin model-fitting results for univariate analyses of educational achievement. A = additive genetic, C = shared environmental, E = non-shared environmental proportions of the variance. **b** SNP heritability estimates of the proportion of variance explained by the additive effects of common SNPs (standard errors as error bars) for the same measures of educational achievement. SNP heritabilities were calculated following adjustment for sex and population stratification. Note: KS1 age 7; KS2 age 11; KS3 age 14; GCSE age 16; Note: Achievement is a composite score of English and Mathematics

npj
The stability of educational achievement across school…
K Rimfeld et al.
4
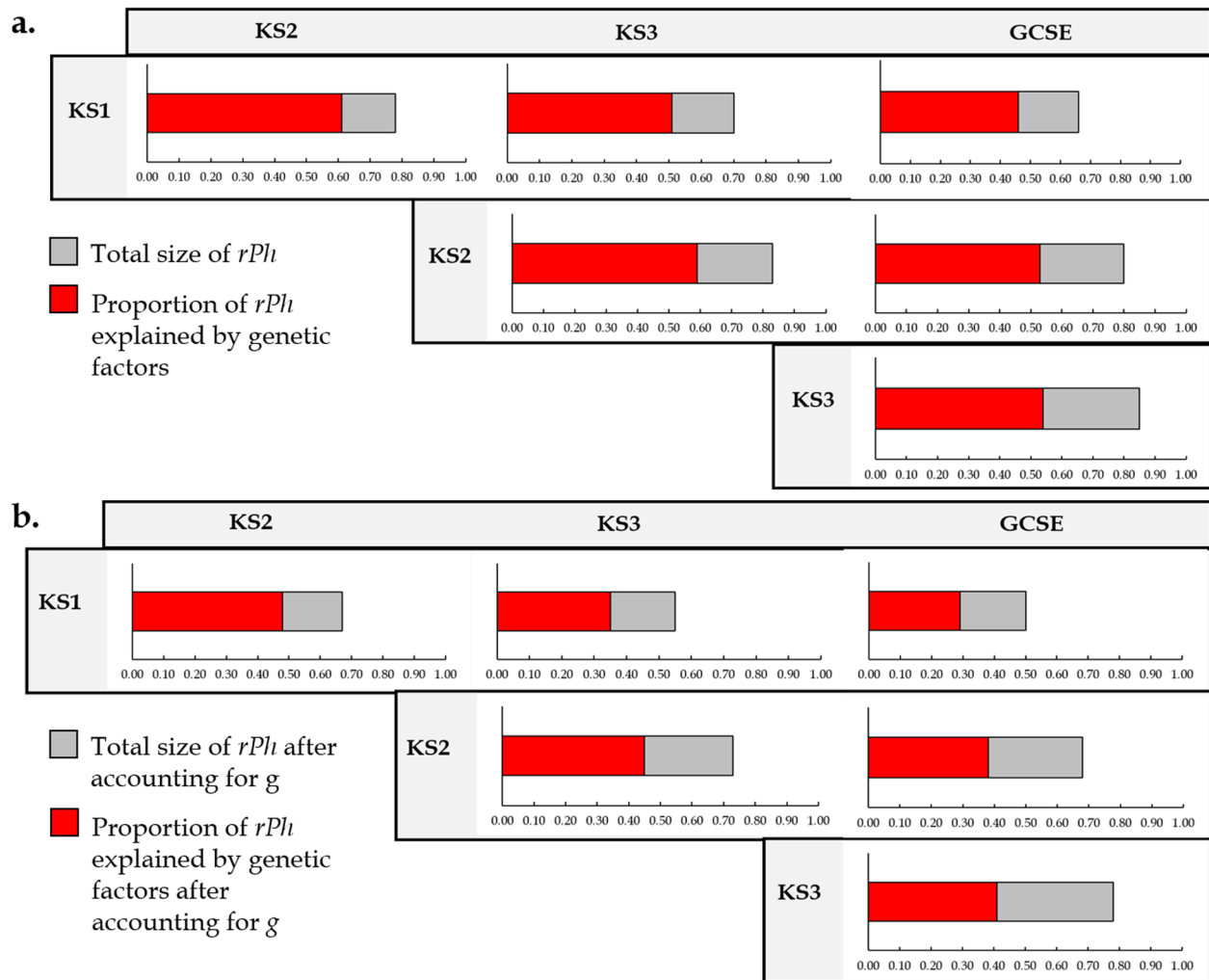
Fig. 2 **a** Proportion of the phenotypic correlation (rPh) across ages accounted for by genetic factors. **b** Proportion of the phenotypic correlation across ages accounted for by genetic factors after controlling for g. Note: KS1 age 7; KS2 age 11; KS3 age 14; GCSE age 16; Note: Achievement is a composite score of English and Mathematics

was unique to the measurement occasion, and did not influence subsequent academic achievement across school years, as indicated by the residuals (age-specific effects; E_{s2}, E_{s3}, and E_{i4}). (See Supplementary Figure 3 for the results of simplex model for English, Mathematics and Science separately.)

The proportion of heritability at each age that is accounted for by genetic effects different from those operating at the previous age can be calculated by dividing the sum of the innovation path squared (A_i) and the age-specific genetic path squared (A_s) by the overall heritability. For example, for GCSE in Fig. 3a, 17% (i.e., $0.31^2/0.58$) of the heritability of core GCSE performance is innovation (there is no age-specific genetic path); the rest of the heritability (83%) is transmitted from previous achievement ages. For KS3 core achievement, 78% (i.e., 0.70 (heritability of KS2) $\times 0.84^2$ (genetic transmission)/0.63 (heritability of KS3)) of the genetic variance was transmitted from KS2, and for KS2 77% ($0.73 \times 0.86^2/0.70$) of the genetic variance was transmitted from KS1. There was substantial innovative genetic influence at each age (A_i)—24%, 15%, and 17% at ages 12, 14, and 16, respectively. To investigate whether the new genetic influence was due to increasing use of test assessments and decreasing use of teacher assessments across the four ages, we repeated the analyses using only standardized test scores across the school years (Supplementary Figure S4), but the results were highly similar. The remaining genetic variance (0% at age 12 and 3% at age 14) was

age specific (path A_s), in other words, not operating at the previous age and not transmitted to the next age. These paths were not significant as indicated by their 95% confidence intervals.

We also repeated the simplex models statistically controlling for g (Fig. 3b). The heritability of core school achievement was somewhat lower after controlling for g, comparable to the bivariate genetic results shown in Fig. 2. Nonetheless, educational achievement continued to be highly stable and its stability was still largely accounted for by genetic factors; genetic paths from age to age are 0.75, 0.76, and 0.79.

In order to assess how much variance in the stability of core educational achievement is explained by a single genetic factor, a genetic common pathway model was used (See Methods and Supplementary Figure 5). The results of the common pathway model are presented in Fig. 4. Seventy percent of the overall stability of core educational achievement across compulsory education (heritability of the latent factor) was explained by genetic factors; 24% of the stability of educational achievement was explained by shared environmental factors (Fig. 4a). The results were similar when we controlled for g—genetic factors explained 59% of the stability in core educational achievement after controlling for intelligence, 21% of the stability was explained by shared environmental factors (Fig. 4b).

The stability of educational achievement across school…
K Rimfeld et al.

npj

5
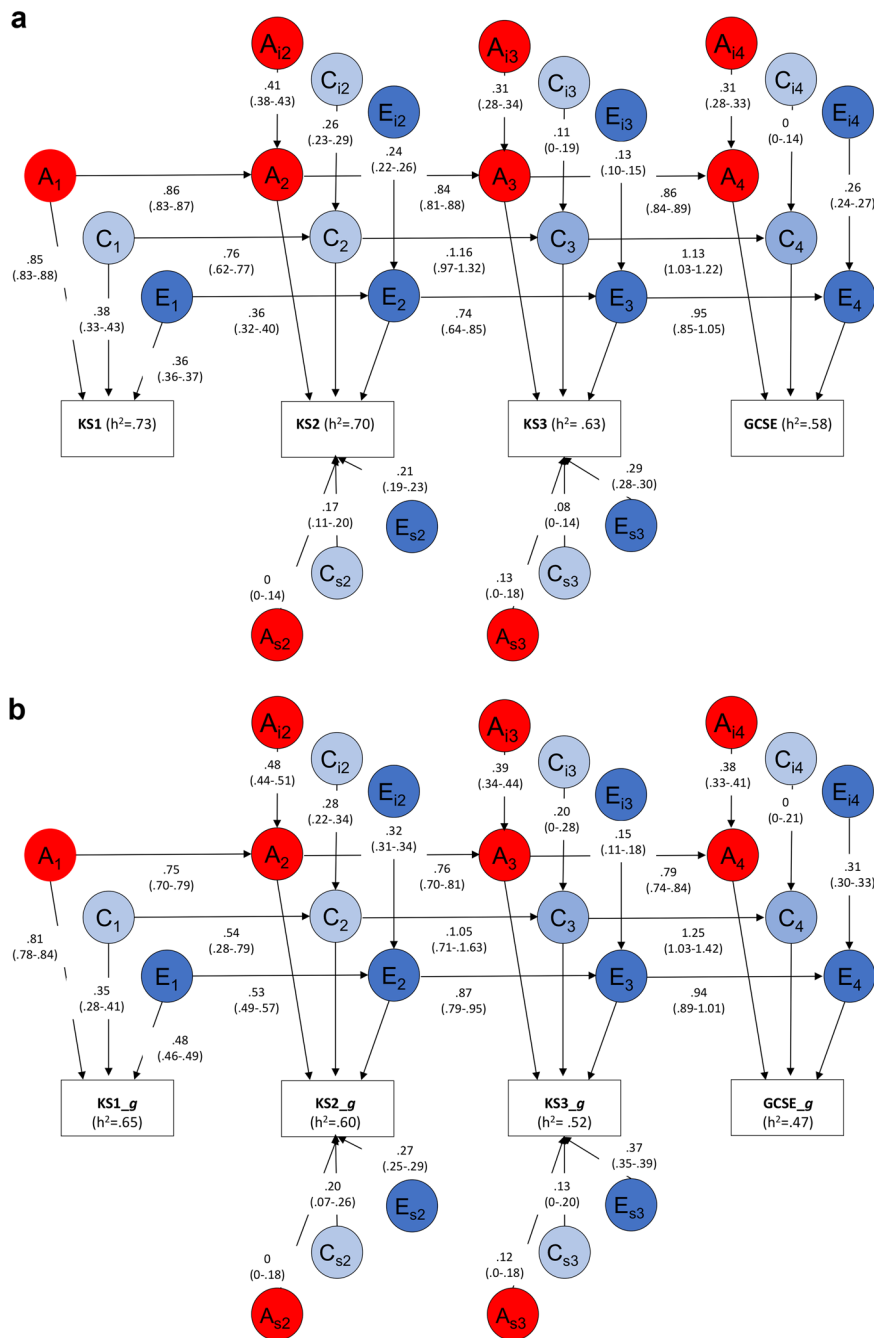
Fig. 3 **a** Simplex model-fitting results for stability and change of overall achievement across compulsory education. **b** Simplex model-fitting results for stability and change of overall achievement across compulsory education after controlling for *g*. Note: KS1 age 7; KS2 age 11; KS3 age 14; GCSE age 16; Note: Achievement is a composite score of English and Mathematics, Note: The path estimates are reported rather than standardized variance components

GPS analyses. As a complement to our twin results, we investigated genetic stability for core educational achievement using a different methodological approach: *EduYears* GPS. *EduYears* GPS increasingly predicted core educational achievement—about 4% for KS1, 6% for KS2, 8% for KS3, and 10% for GCSE. In order to address the question of genetic stability and innovation, we explored the age-specificity of the *EduYears* GPS prediction, after accounting for the variance explained at all preceding ages. In line with the multivariate twin analyses, *EduYears* GPS' prediction of educational achievement was largely stable from age to age (Fig. 5). That is, our regression analyses indicated little

(<1%) age-specific genetic prediction once the stable prediction of *EduYears* GPS from all previous ages was taken into account. Details of these analyses for core achievement, for subjects separately, and controlling for *g* and previous achievement are presented in Supplementary Table 4. In summary, results were similar for separate subjects and after controlling for *g* and previous achievement. However, *EduYears* GPS still predicts educational achievement when only controlling for *g*, explaining around 4% in GCSE performance, as illustrated in Supplementary Figure 6.

npj

The stability of educational achievement across school…
K Rimfeld et al.

6
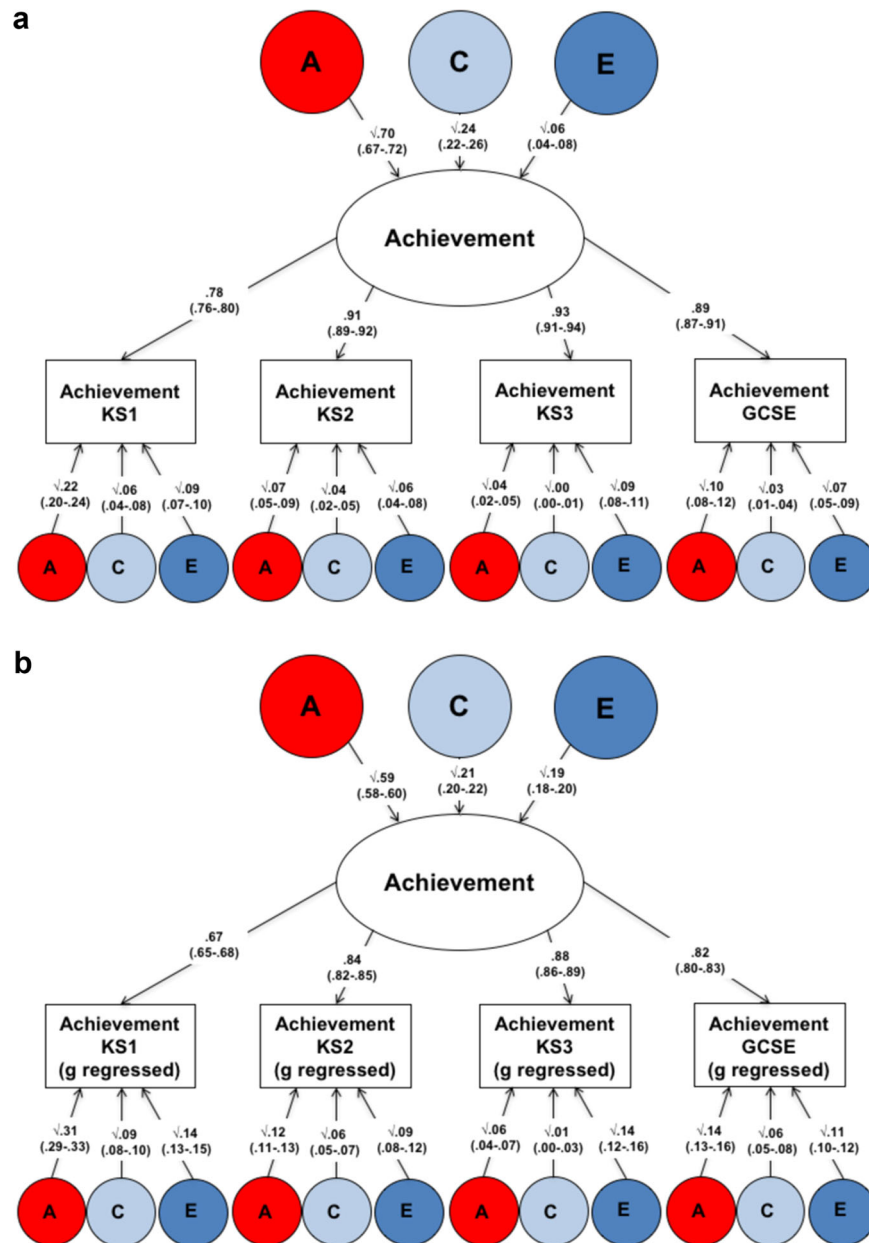
Fig. 4 **a** Common pathway model presenting the standardized squared path estimates for overall achievement. **b** and for overall achievement when controlling for *g*. A = additive genetic, C = shared environmental, and E = non-shared environmental components of variance. Note: KS1 age 7; KS2 age 11; KS3 age 14; and GCSE age 16; Note: Achievement is a composite score of English and Mathematics

DISCUSSION

The present study shows that individual differences in educational achievement are highly stable across the years of compulsory schooling from primary through secondary school. Children who do well at the beginning of primary school also tend to do well at the end of compulsory education for much the same reasons. The very high stability of academic achievement across compulsory school years is an interesting finding, particularly when considering that children go through major cognitive and emotional changes from childhood to adolescence, as well as experiencing changes in teachers, friends, and schools.

In addition, the nature of educational achievement also changes during the school years as children are exposed to more subjects and more complex subjects. For reading, children move from learning to read to using reading to learn. The present twin analyses address, for the first time, the etiology of the stability of

academic outcomes over compulsory education, showing that genetic factors are largely responsible for this stability. In other words, the same genetic factors largely shape individual differences in achievement from primary through secondary school. Shared environmental factors were also largely stable, although they explained a smaller proportion of overall variance in achievement. However, it has been suggested that shared environmental effects might actually be driven genetically.[40,41] We show that age-to-age change in achievement scores was to a large extent explained by non-shared environmental factors. This is another example of the general rubric of behavioral genetics, "genetic stability, environmental change".[22,30,42–44] We also noted some genetic innovation (change) at each stage of assessment, but, consistent with an overall pattern of stability, all of these new genetic influences were transmitted to the next achievement stage.

The stability of educational achievement across school…
K Rimfeld et al.

npj

7
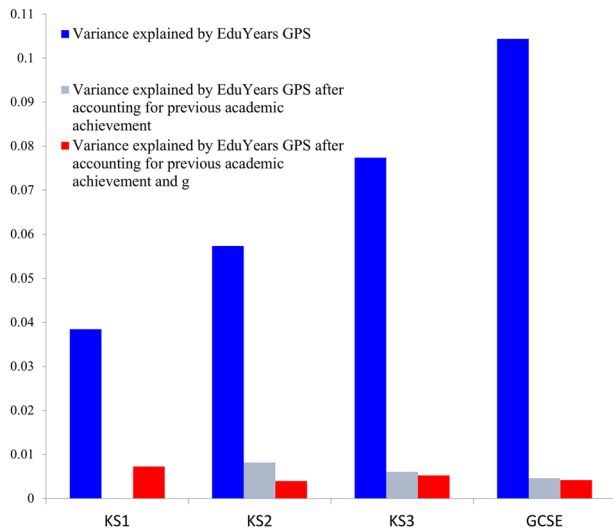
Fig. 5 Variance explained by GPS (*EduYears*) using Gaussian mixture weights of 1.0 for overall educational achievement across compulsory education. Note: KS1 age 7; KS2 age 11; KS3 age 14; GCSE age 16; Note: Achievement is a composite score of English and Mathematics

A reasonable assumption is that the substantial genetic stability observed here is explained by general cognitive ability (*g*, intelligence). Importantly, however, we showed that the heritability of educational achievement over school years and its stability is not explained by *g* alone. The results of our twin analyses showed that when *g* was controlled for, educational achievement remained highly heritable and stable and the stability of educational achievement independent of *g* was still explained by genetic factors. Although there was evidence for some specific (new) genetic influence at each age, again these new genetic influences were not age specific but were transmitted to the next assessment stage. This is in line with our earlier reports in which we showed that educational achievement at age 16 is not explained by intelligence alone.[10,12] The *EduYears* GPS regression analysis yielded similar results showing genetic stability, even after controlling for *g*. This GPS result is not exactly analogous to the twin study results, as we tested the effect of the same genetic variants over time. Nevertheless, our multi-method approach yielded similar results indicating that the substantial stability of educational achievement is to a large extent explained by genetic factors.

As new, more powerful GWA studies are conducted, the predictive power of the *EduYears* GPS prediction is likely to increase. The GPS calculated using the 2013 *EduYears* GWA summary statistics with a sample size of 126,000[45] predicted around 3% of variance in educational achievement in TEDS,[20] compared to 10% of variance explained in the current study based on the 2016 *EduYears* GWAS with a sample size of 330,000. Another more powerful GWA of educational attainment was recently published, involving over one million participants, this is likely to be a game changer in terms of predictive power.[46]

It should be noted that *EduYears* GPS predicts only about 4% of the variance in adult years of education (educational attainment)[28] in independent samples, but it predicts more than twice as much variance in GCSE scores at age 16. We are not aware of any other example in which a GPS predicts less variance in the GWA target trait (educational attainment) than in another trait (GCSE scores). We suggest that the reason for this unusual finding is that educational attainment is a much coarser measure than GCSE scores, which are the result of hours of standardized assessment. In support of this hypothesis, we find that *EduYears* GPS also predicts 4% of the variance when we analyzed a similarly coarse

dichotomous item about whether or not TEDS participants planned to go to university. Furthermore, *EduYears* GPS also predicts 4% of the variance in a cruder measure of GCSE achievement—5 passes at grades A* to C, which is often used in government statistics, and used for selection purposes by many employers and educational institutions (Supplementary Table 4).

The limitations of this study include the usual assumptions of the twin design, which are described in detail elsewhere.[43,47] One of these limitations involves assortative mating, in which mate selection is not at random but is instead based on trait similarity. Assortative mating on cognitive abilities and educational achievement has been shown to be substantial (~0.40).[36,43,48] In the twin design, assortative mating increases DZ correlations relative to MZ correlations and could therefore lead to underestimating heritability and overestimating shared environmental influence; in effect this makes the present findings concerning heritability conservative. GCTA and GPS methods also have their limitations. Notably, both of these DNA-based methods rely on the additive effects of common SNPs genotyped on SNP arrays, and do not capture gene-gene or gene-environment interplay or the effects of less common SNPs.[49] However, since the main limitations are different for each method used in the current study, the fact that our multi-method approach yielded similar results is a strength.

Our multi-method analyses corroborated previous findings showing that individual differences in educational achievement are largely explained by inherited differences in DNA sequence. The novel contribution of our study is to show that the substantial stability of educational achievement across compulsory education is to a large extent explained by genetic factors, even after controlling for *g*.

Our finding of genetically driven stability of educational achievement should provide additional motivation to identify children in need of interventions as early as possible, as the problems are likely to remain throughout the school years. GPS prediction, specifically, might in the future provide a tool to identify children with educational problems very early in life and aid in providing both individualized prevention and individualized learning programs. We hope that with GPS, we can move towards precision education, just as medicine is moving towards precision medicine.[50,51] For example, GPS could be used to identify children at birth at genetic risk for developing reading problems, thus enabling early intervention. As preventive interventions have greater chances of succeeding early in life, a great strength of GPS is that they can predict at birth just as well as later in life, which enables early intervention, particularly for those children who are likely to struggle the most.

METHODS

Participants

The present study used the TEDS sample. TEDS is a large twin study that recruited over 16,000 twin pairs born between 1994 and 1996 in England and Wales. More than 10,000 twin pairs are still actively involved in the study. Rich cognitive and behavioral data, including educational achievement, have been collected from the twins, their parents and teachers, over compulsory education and beyond. Importantly, TEDS was a representative sample of the UK population at first contact, and remains representative in terms of family socioeconomic status and ethinicity.[26,52] Ethical approval for this study was received from King's College London Ethics Committee.

The sample for the present study included all twins with available academic achievement measures over the school years. Participants who had major medical or psychiatric conditions, or those with severe perinatal complications, were removed from the analyses. Zygosity was assessed by the parent-reported questionnaire of physical similarity. This measure has been shown to be highly reliable.[53] Nevertheless, DNA testing was conducted when zygosity was unclear from the questionnaire. The sample size per academic achievement measure is shown in Supplementary Table S1.

npj

The stability of educational achievement across school…
K Rimfeld et al.

8

DNA has been genotyped for a subsample of unrelated individuals from TEDS (one twin per pair). We processed genotypes for 6710 individuals using the standard quality control procedure followed by imputation of genetic variants to the Haplotype Reference Consortium[54] (see Supplementary Methods). We then matched the individuals with genotyped data to those participants with available academic achievement data.

Measures

Measures of educational achievement obtained by TEDS. TEDS has obtained assessments of academic achievement directly from the twins' teachers who reported grades following the UK National Curriculum guidelines, a standardized core academic curriculum formulated by the National Foundation for Educational Research (NFER) and the Qualifications and Curriculum Authority (QCA) (NFER: http://www.nfer.ac.uk/index.cfm; QCA: http://www.qca.org.uk). Data were obtained directly from teachers. At age 7 data are available for English and Mathematics; at ages 12 and 14 data are available for English, Mathematics, and Science. The teacher rating of English used a combined rating of students' reading, writing, and speaking and listening; Mathematics used a combined score of knowledge in numbers, shapes, space, using and applying mathematics, and measures; and Science used a score combining life process, scientific enquiry, and physical process. These teacher ratings were found to be highly reliable when compared to the achievement measures collected by the UK National Pupil Database (NPD), as described later.

GCSE exam results were obtained from twins themselves or from their parents via questionnaires sent over mail or via telephone. GCSEs are UK-wide standardized examinations taken at age of 16 at the end of compulsory education. Children choose from a variety of different subjects, while English, Mathematics, and Science are compulsory. We used exam grades from English, Mathematics, and Science for the current analyses. Composite measures were created for English (mean of English language and English literature grades), Science (mean of single or double-weighted Science or, when taken separately Chemistry, Physics, and Biology grade), and Mathematics.

Measures of educational achievement obtained from the NPD. The TEDS dataset was linked to the NPD for every participant for whom we received written informed consent from either the twin or the parent. NPD is a rich UK database collecting data about students' academic achievement across the school years (https://www.gov.uk/government/collections/national-pupil-database). Data are available for each Key Stage (KS) completed in the UK following the National Curriculum (NC). Teachers provide NC ratings for every student at the end of each KS (similarly to data collected at TEDS for the NC ratings in English, Mathematics, and Science). Exam scores as well as teacher ratings are available from KS1–KS3; and exam scores only are available for KS4 and KS5. Children's ages for KS1, KS2, and KS3 are about 7, 11, and 14 years. KS4 marks the end of compulsory education with GCSE testing at about age 16. Sample size and descriptive characteristics for each measure are provided in Supplementary Table S1.

Composite scores of educational achievement. Composite scores were calculated at each KS combining the teacher ratings (both TEDS and NPD) with the exam scores for English, Mathematics, and Science separately by taking a mean of the three scores. The average correlation between NPD and TEDS teacher ratings was 0.70 (see Supplementary Table S5), and the average correlation between teacher ratings and exam scores was 0.80 (see Supplementary Table S6). For GCSE performance at the end of compulsory education, GCSE grades collected by TEDS and by NPD correlated 0.98 for English, 0.99 for Mathematics, and >0.95 for all Sciences. A mean score for NPD and TEDS was created to increase the sample size; when fewer measures were available we used any available data to calculate the composite score of educational achievement.

The overall achievement measure (core achievement) was calculated at each KS by taking a mean of English NC teacher ratings, Mathematics NC teacher rating (for both NPD and TEDS), English exam score, and Mathematics exam score. We did not include Science grades in overall achievement scores to make a more direct comparison across age because Science is not part of the National Curriculum at KS1.

Measures of general cognitive ability (g). General cognitive ability (g; intelligence) was assessed in TEDS at ages 7, 9, 10, 12, 14, and 16. For the present analyses we created a longitudinal composite measure of g as a mean of these six assessments. See Supplementary Methods for a more detailed description of g measures.

Analyses

Phenotypic analyses. The measures were described in terms of means and variance, comparing males and females and identical and non-identical twins; mean differences for age and sex and their interaction were tested using univariate ANOVA. Phenotypic correlations were calculated between academic achievement measures across development. The academic achievement measures were corrected for the small mean effects of age and sex (Supplementary Table S1) by rescoring the variable as a standardized residual correcting for age and sex, because in the analysis of twin data members of a twin pair are identical in age and MZ twins are identical for sex, and this would otherwise inflate twin estimates of shared environment.[55] Full sex limitation genetic modeling has previously been reported for academic achievement and found only very minor sex differences in genetic and environmental estimates.[6,9,12] For these reasons, and to increase power in the present analyses, the full sample was used, combining males and females and including opposite-sex pairs.

Finally, before conducting twin analyses, the achievement measures were corrected for skew because they were slightly negatively skewed. The achievement measures were corrected for skew by mapping it on to a standard normal distribution using the rank-based van der Waerden's transformation.[56]

Twin design. The twin design was used for univariate and multivariate genetic analyses. The twin method offers a natural experiment capitalizing on the known genetic relatedness of MZ and DZ twin pairs. MZ twins are genetically identical and share 100% of their genes, while DZ twins share on average 50% of their segregating genes. Both MZ and DZ twins are assumed to share 100% of their shared environmental influences growing up in the same family. Non-shared environmental influences are unique to individuals, not contributing to similarity between twins. Using these known family relatedness coefficients, it is possible to estimate the relative contribution of additive genetic (A), shared environmental (C), and non-shared environmental (E) effects on the variance and covariance of the phenotypes, by comparing MZ correlations to DZ correlations. Heritability can be roughly calculated by doubling the difference between MZ and DZ correlations, C can be calculated by deducting heritability from MZ correlation and E can be estimated by deducting MZ correlation from unity (following Falconer's formula).[47] These parameters can be estimated more accurately using structural equation modeling, which also provides 95% confidence intervals and estimates of model fit. The structural equation modeling program OpenMx was used for all model-fitting analyses.[57]

These univariate analyses can be extended to multivariate analyses to study the etiology of covariance between multiple traits. Multivariate genetic method decomposes the covariance between traits into additive genetic (A), shared environmental (C), and non-shared environmental (E) components by comparing the cross-trait cross-twin correlations between MZ and DZ twin pairs. This method also enables estimation of the genetic correlation (rG), which is an index of pleiotropy, indicating the extent to which the same genetic variants influence two traits or measures of the same trait at two times. The shared environmental correlation (rC) and non-shared environmental correlation (rE) are estimated in a similar manner.[43,47]

We used two longitudinal models to study the issue of age-to-age stability of educational achievement.

The simplex model is a multivariate genetic model that estimates the extent to which the genetic and environmental influences on a trait are transmitted from age to age, and the extent to which innovative and age-specific influences emerge.[58] The covariance or correlation matrix for such data is called simplex because the strength of the associations tends to correspond to differences between ages, that is, they are often highest along the diagonal and fall systematically as the difference between ages increases.[58] The simplex model is illustrated in Supplementary Figure S2.

The common pathway model is a multivariate genetic model in which the variance common to all measures included in the analysis can be reduced to a common latent factor, for which the A, C, and E components are estimated. As well as estimating the etiology of the common latent factor, the model allows for the estimation of the A, C, and E components of the residual variance in each measure that is not captured by the latent construct.[59] The common pathway model estimates the extent to which the stable variance in educational achievement across compulsory education (the latent factor of achievement) is explained by A, C, and E. The common pathway model is illustrated in Supplementary Figure S6.

SNP heritability. The genome-wide complex trait analysis (GCTA) software package enables estimates of the proportion of phenotypic variance or

The stability of educational achievement across school…
K Rimfeld et al.

9

covariance that is explained by all SNPs that are available on genotype arrays, without testing the association of any single SNP individually.[17,49,60] This estimate is often called SNP heritability. This method does not use known genetic relatedness coefficients but estimates heritability from DNA using only unrelated individuals. SNP heritability is calculated using restricted maximum likelihood and the variance and covariance is decomposed using mixed linear models.

First, the genetic relatedness matrix is calculated by weighting genetic similarities between all possible pairs of individuals with the allele frequencies across all SNPs on the DNA array. Individuals who are found to be even remotely related (greater than fifth cousins) are removed from the analyses as they would otherwise bias the results, which rely on chance genetic similarity between pairs of individuals.[17,18,61] The matrix of pair-by-pair genetic similarity is compared to the matrix of pair-by-pair phenotypic similarity using the residual maximum likelihood estimation. SNP heritabilities were calculated for overall achievement across compulsory education, as well as for specific subjects.

Genome-wide polygenic scores. GPSs aggregate the effects of individual SNPs shown to be associated with the trait in a GWA study.[62] GPSs were calculated for 6710 participants using summary statistics from Okbay et al.[28] GWA analysis of years of education (*EduYears*).[28] Of the 293,723 participants in the *EduYears* GWA discovery sample, the summary statistics excluded 23andMe participants, for legal reasons. Polygenic scores were constructed as the weighted sums of each individual's genotype across all SNPs using the *LDpred* method[63] (see Supplementary Methods for details). Delta R^2 is reported as the estimate of variance explained by the GPS. These delta R^2 estimates were obtained by comparing the incremental increase in the model R^2 after adding the GPS to the regression model, and comparing this to the model that included 10 principal components in order to control for population stratification. See Supplementary Methods for genetic quality control and further information about GPS calculation.

We correlated *EduYears* with general educational achievement composites, as well as with performance in specific subjects at each age to estimate *EduYears* GPS heritability. Delta R^2 are reported as the estimates of variance explained by adding the GPS to the regression model that included the academic achievement from all earlier ages to assess the extent to which *EduYears* contributes to age-to-age stability.

DATA AVAILABILITY

For information on data availability, please see the TEDS data access policy. This can be found at: http://www.teds.ac.uk/research/collaborators-and-data/teds-data-access-policy. All relevant data are available from the authors according to the TEDS data access policy.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: K.R., M.M., R.P. Analyzed the data: K.R., M.M., L.J.H., E.K. Wrote the paper: K.R., M.M., P.S.D., R.P. All authors approved the final draft of the paper.

ADDITIONAL INFORMATION

Supplementary information accompanies the paper on the *npj Science of Learning* website (https://doi.org/10.1038/s41539-018-0030-0).

Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. Arendt, J. N. Does education cause better health? A panel data analysis using school reforms for identification. *Econ. Educ. Rev.* **24**, 149–160 (2005).
2. Cutler, D. M. & Lleras-Muney, A. Education and health: insights from international comparisons. NBER working paper no. 17738. (The National Bureau of Economic Research, Cambridge, MA, 2012).
3. Gottfredson, L. S. & Deary, I. J. Intelligence predicts health and longevity, but why? *Curr. Dir. Psychol. Sci.* **13**, 1–4 (2004).
4. Oreopoulos, P. & Salvanes, K. G. Priceless: the nonpecuniary benefits of schooling. *J. Econ. Perspect.* **25**, 159–184 (2011).
5. Furnham, A. & Cheng, H. Childhood cognitive ability predicts adult financial well-being. *J. Intell.* **5**, 3 (2016).
6. Shakeshaft, N. G. et al. Strong genetic influence on a UK nationwide test of educational achievement at the end of compulsory education at age 16. *PLoS ONE* **8**, e80341 (2013).
7. Bartels, M., Rietveld, M. J. H., Van Baal, G. C. M. & Boomsma, D. I. Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. *Twin Res.* **5**, 544–553 (2002).
8. Coventry, W. et al. The etiology of individual differences in second language acquisition in Australian school students: a behavior-genetic study. *Lang. Learn.* **62**, 880–901 (2012).
9. Kovas, Y., Haworth, C. M. A., Dale, P. S. & Plomin, R. The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monogr. Soc. Res. Child Dev.* **72**, 1–144 (2007).
10. Krapohl, E. et al. The high heritability of educational achievement reflects many genetically influenced traits, not just intelligence. *Proc. Natl Acad. Sci. USA* **111**, 15273–15278 (2014).
11. Rimfeld, K., Ayorech, Z., Dale, P. S., Kovas, Y. & Plomin, R. Genetics affects choice of academic subjects as well as achievement. *Sci. Rep.* **6**, 26373 (2016).
12. Rimfeld, K., Kovas, Y., Dale, P. S. & Plomin, R. Pleiotropy across academic subjects at the end of compulsory education. *Sci. Rep.* **5**, 11713 (2015).
13. Petrill, S. A. et al. Genetic and environmental influences on the growth of early reading skills. *J. Child Psychol. Psychiatry Allied Discip.* **51**, 660–667 (2010).
14. Wadsworth, S. J., DeFries, J. C., Fulker, D. W. & Plomin, R. Cognitive ability and academic achievement in the Colorado adoption project: A multivariate genetic analysis of parent-offspring and sibling data. *Behav. Genet.* **25**, 1–15 (1995).
15. Wainwright, M. A., Wright, M. J., Geffen, G. M., Luciano, M. & Martin, N. G. The genetic basis of academic achievement on the Queensland Core Skills Test and its shared genetic variance with IQ. *Behav. Genet.* **35**, 133–145 (2005).
16. Martin, N. G. & Martin, P. G. The inheritance of scholastic abilities in a sample of twins. I. Ascertainments of the sample and diagnosis of zygosity. *Ann. Hum. Genet.* **39**, 213–218 (1975).
17. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. GCTA: a tool for genome-wide complex trait analysis. *Am. J. Hum. Genet.* **88**, 76–82 (2011).
18. Yang, J., Lee, S. H., Goddard, M. E. & Visscher, P. M. Genome-wide complex trait analysis (GCTA): Methods, data analyses, and interpretations. *Methods Mol. Biol.* **1019**, 215–236 (2013).
19. Trzaskowski, M. et al. DNA evidence for strong genome-wide pleiotropy of cognitive and learning abilities. *Behav. Genet.* **43**, 267–273 (2013).
20. Krapohl, E. & Plomin, R. Genetic link between family socioeconomic status and children's educational achievement estimated from genome-wide SNPs. *Mol. Psychiatry* **21**, 437–443 (2016).
21. Davis, O. S. P. et al. The correlation between reading and mathematics ability at age twelve has a substantial genetic component. *Nat. Commun.* **5**, 4204 (2014).
22. Wadsworth, S. J., Corley, R. P., Hewitt, J. K. & DeFries, J. C. Stability of genetic and environmental influences on reading performance at 7, 12, and 16 years of age in the Colorado Adoption Project. *Behav. Genet.* **31**, 353–359 (2001).
23. Hart, S. A. et al. Exploring how nature and nurture affect the development of reading: an analysis of the Florida twin project on reading. *Dev. Psychol.* **49**, 1971–1981 (2013).
24. Malanchini, M. et al. Reading self-perceived ability, enjoyment and achievement: a genetically informative study of their reciprocal links over time. *Dev. Psychol.* **53**, 698–712 (2017).
25. Soden, B. et al. Longitudinal stability in reading comprehension is largely heritable from grades 1 to 6. *PLoS ONE* **10**, e0113807 (2015).
26. Haworth, C. M. A., Davis, O. S. P. & Plomin, R. Twins Early Development Study (TEDS): a genetically sensitive investigation of cognitive and behavioral development from childhood to young adulthood. *Twin Res. Hum. Genet.* **16**, 117–125 (2013).

npj

The stability of educational achievement across school…
K Rimfeld et al.

10

27. Palla, L. & Dudbridge, F. A fast method that uses polygenic scores to estimate the variance explained by genome-wide marker panels and the proportion of variants affecting a trait. *Am. J. Hum. Genet.* **97**, 250–259 (2015).

28. Okbay, A. et al. Genome-wide association study identifies 74 loci associated with educational attainment. *Nature* **533**, 539–542 (2016).

29. Selzam, S. et al. Predicting educational achievement from DNA. *Mol. Psychiatry* **22**, 267–272 (2017).

30. Deary, I. J. et al. Genetic contributions to stability and change in intelligence from childhood to old age. *Nature* **482**, 212 (2012).

31. Deary, I. J., Spinath, F. M. & Bates, T. C. Genetics of intelligence. *Eur. J. Hum. Genet.* **14**, 690–700 (2006).

32. Gow, A. J. et al. Stability and change in intelligence from age 11 to ages 70, 79, and 87: the Lothian Birth Cohorts of 1921 and 1936. *Psychol. Aging* **26**, 232–240 (2011).

33. Calvin, C. M. et al. Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behav. Genet.* **42**, 699–710 (2012).

34. Deary, I. J., Strand, S., Smith, P. & Fernandes, C. Intelligence and educational achievement. *Intelligence* **35**, 13–21 (2007).

35. Deary, I. J. Intelligence. *Annu. Rev. Psychol.* **63**, 453–482 (2012).

36. Plomin, R. & Deary, I. J. Genetics and intelligence differences: five special findings. *Mol. Psychiatry* **20**, 98–108 (2015).

37. Petrill, S. A. & Wilkerson, B. Intelligence and achievement: a behavioral genetic perspective. *Educ. Psychol. Rev.* **12**, 185–199 (2000).

38. Haworth, C. M. A. et al. The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Mol. Psychiatry* **15**, 1112–1120 (2010).

39. Eaves, L. J., Long, J. & Heath, A. C. A theory of developmental change in quantitative phenotypes applied to cognitive development. *Behav. Genet.* **16**, 143–162 (1986).

40. Kong, A. et al. The nature of nurture: effects of parental genotypes. *Science* **359**, 424–428 (2018).

41. Bates, T. C. et al. The nature of nurture: using a virtual-parent design to test parenting effects on children's educational attainment in genotyped families. *Twin Res. Hum. Genet.* **21**, 73–83 (2018).

42. DeFries, J. C., Plomin, R. & LaBuda, M. C. Genetic stability of cognitive development from childhood to adulthood. *Dev. Psychol.* **23**, 4–12 (1987).

43. Knopik, V. S., Neiderhiser, J. M., DeFries, J. C. & Plomin, R. *Behavioral Genetics.* (Worth Publishers, New York, 2017).

44. Plomin, R., Pedersen, N. L., Lichtenstein, P. & McClearn, G. E. Variability and stability in cognitive abilities are largely genetic later in life. *Behav. Genet.* **24**, 207–215 (1994).

45. Rietveld, C. A. et al. GWAS of 126,559 individuals identifies genetic variants associated with educational attainment. *Sci. (80-.).* **340**, 1467–1471 (2013).

46. Lee, J. J. et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat. Genet* **50**, 1112–1121 (2018).

47. Rijsdijk, F. V. & Sham, P. C. Analytic approaches to twin data using structural equation models. *Brief. Bioinform.* **3**, 119–133 (2002).

48. Hugh-Jones, D., Verweij, K. J. H. St, Pourcain, B. & Abdellaoui, A. Assortative mating on educational attainment leads to genetic spousal resemblance for polygenic scores. *Intelligence* **59**, 103–108 (2016).

49. Yang, J. et al. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* **42**, 565–569 (2010).

50. Plomin, R. & Von Stumm, S. The new genetics of intelligence. *Nat. Rev. Genet.* **19**, 148–159 (2018).

51. Aronson, S. J. & Rehm, H. L. Building the foundation for genomics in precision medicine. *Nature* **526**, 336–342 (2015).

52. Oliver, B. R. & Plomin, R. Twins' Early Development Study (TEDS): a multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Res. Hum. Genet.* **10**, 96–105 (2007).

53. Price, T. S. et al. Infant zygosity can be assigned by parental report questionnaire data. *Twin Res.* **3**, 129–133 (2000).

54. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet* **48**, 1279–83 (2016).

55. McGue, M. & Bouchard, T. J. Adjustment of twin data for the effects of age and sex. *Behav. Genet.* **14**, 325–343 (1984).

56. Van Der Waerden, B. L. On the sources of my book Moderne Algebra. *Hist. Math.* 2, (31–40) (1975).

57. Boker, S. et al. OpenMx: an open source extended structural equation modeling framework. *Psychometrika* **76**, 306–317 (2011).

58. Boomsma, D. I. & Molenaar, P. C. M. The genetic analysis of repeated measures. I. Simplex models. *Behav. Genet.* **17**, 111–123 (1987).

59. Rijsdijk, F. V. *Encyclopedia of Statistics in Behavioral Science* (eds Everitt, B. S. & Howell, D. C.) 1, 330–331 (John Wiley & Sons Ltd., New Jersey, 2005).

60. Lee, S. H., Yang, J., Goddard, M. E., Visscher, P. M. & Wray, N. R. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. *Bioinformatics* **28**, 2540–2542 (2012).

61. Trzaskowski, M. et al. Genetic influence on family socioeconomic status and children's intelligence. *Intelligence* **42**, 83–88 (2014).

62. Dudbridge, F. Power and predictive accuracy of polygenic risk scores. *PLoS Genet.* **9**, e1003348 (2013).

63. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).