This is the accepted manuscript of the following publication:
Calvin CM, Deary IJ, Webbink D, Smith P, Fernandes C, Lee SH, Luciano M & Visscher PM (2012) Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. Behavior Genetics 42: 699-710.
The final publication is available at http://dx.doi.org/10.1007/s10519-012-9549-7

Behavior Genetics

Multivariate genetic analyses of cognition and education from two population samples of 174,000 and 121,000 school children

Abstract: The genetic influence on the association between contemporaneously measured intelligence and academic achievement in childhood was examined in nationally representative cohorts from England and The Netherlands using a whole population indirect twin design, including singleton data. We identified 1,056 same-sex and 495 opposite-sex twin pairs among 174,098 British 11 year-olds with test scores from 2004, and, 785 same-sex and 327 opposite-sex twin pairs among 120,995 Dutch schoolchildren, aged 8, 10 or 12 years, with assessments from 1994 to 2002. The estimate of intelligence heritability was large in both cohorts, consistent with previous studies (h^2 = .70 ±.14, England; h^2 = .43 ±.28 to .67 ±.31, The Netherlands), as was the heritability of academic achievement variables (h^2 = .51 ±.16 to .81 ±.16, England; h^2 = .36 ±.27 to .74 ±.27, The Netherlands). Additive genetic covariance explained the large majority of the phenotypic correlations between intelligence and academic achievement scores in England, when standardised to a bivariate heritability (Biv h^2 = .76 ±.15 to .88 ±.16), and less consistent but often large proportions of the phenotypic correlations in The Netherlands (Biv h^2 = .33 ±.52 to 1.00 ±.43). In the British cohort both nonverbal and verbal reasoning showed very high additive genetic covariance with achievement scores (Biv h^2 = .94 to .98; Biv h^2 = .77 to 1.00 respectively). In The Netherlands, covariance estimates were consistent across age groups. The heritability of intelligence-academic achievement associations in two population cohorts of elementary schoolchildren, using a twin pair extraction method, is at the high end of estimates reported by studies of largely preselected twin samples.

TITLE

Multivariate genetic analyses of cognition and education from two population samples of 174,000 and
166,000 school children

SUGGESTED RUNNING HEAD

Multivariate genetic analyses of cognition and education from two population samples

AUTHORS

Catherine M. Calvina,1, Ian J. Dearya, Dinand Webbinkb, Pauline Smithc, Cres Fernandesc, Sang Hong Leed,
Michelle Lucianoa, and Peter M. Visscherd,

KEY WORDS

academic achievement; heritability; intelligence; genetic covariance; twins.

a Centre for Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of Edinburgh, Edinburgh
EH8 9JZ, UK
b Erasmus School of Economics, Department of Econometrics, Erasmus Centre for Strategic Philanthropy, PO Box 1738,
3000 DR Rotterdam, The Netherlands
c GL Assessment, 9th Floor East, 389 Chiswick High Road, London, W4 4AL, UK
d The University of Queensland, Brisbane, 4072, Queensland, Australia
1 To whom correspondence should be addressed. E-mail: c.m.calvin@sms.ed.ac.uk; Telephone: +44(0)131 650 8275; Fax:
+44(0)131 651 1771

The authors declare no conflict of interest.

Abstract

The genetic influence on the association between contemporaneously measured intelligence and academic achievement in childhood was examined in nationally representative cohorts from England and The Netherlands using a whole population indirect twin design, including singleton data. We identified 1,056 same-sex and 495 opposite-sex twin pairs among 174,098 British 11 year-olds with test scores from 2004, and, 785 same-sex and 327 opposite-sex twin pairs among 120,995 Dutch schoolchildren, aged 8, 10 or 12 years, with assessments from 1994 to 2002. The estimate of intelligence heritability was large in both cohorts, consistent with previous studies ($h2$ = .70 ±.14, England; $h2$ = .43 ±.28 to .67 ±.31, The Netherlands), as was the heritability of academic achievement variables ($h2$ = .51 ±.16 to .81 ±.16, England; $h2$ = .36 ±.27 to .74 ±.27, The Netherlands). Additive genetic covariance explained the large majority of the phenotypic correlations between intelligence and academic achievement scores in England, when standardised to a bivariate heritability (*Biv* $h2$ = .76 ±.15 to .88 ±.16), and less consistent but often large proportions of the phenotypic correlations in The Netherlands (*Biv* $h2$ = .33 ±.52 to 1.00 ±.43). In the British cohort both nonverbal and verbal reasoning showed very high additive genetic covariance with achievement scores (*Biv* $h2$ = .94 to .98; *Biv* $h2$ = .77 to 1.00 respectively). In The Netherlands, covariance estimates were consistent across age groups. The heritability of intelligence-academic achievement associations in two population cohorts of elementary schoolchildren, using a twin pair extraction method, is at the high end of estimates reported by studies of largely preselected twin samples.

Introduction

Intelligence test scores, and academic achievement measured at the same age or older, correlate strongly ($r > .60$) in national samples of schoolchildren (Bartels et al. 2002; Deary et al. 2007). The underlying causes of this association however are a subject of current debate. This is for several reasons. Firstly, understanding the distinction between constructs of intelligence and academic achievement is of theoretical importance (Luo et al. 2003). Secondly, if intelligence is directly causal to educational outcomes, and/or vice versa, this influences how their independent effects on health and social outcomes (Calvin et al. 2011; Lleras-Muney 2005; Strenze 2007) are modelled and interpreted in life course epidemiology (Deary and Johnson 2010; Richards and Sacker 2011). Thirdly, the extent to which their association is genetically or environmentally determined has real-life implications for education and health policy (Petrill and Wilkerson 2000). In the past few decades, family-based research largely on twins (Bartels et al. 2002; Petrill and Wilkerson 2000; Kovas et al. 2007; Thompson et al. 1991; Wadsworth et al. 1995; Wainwright et al. 2005), has provided evidence that a significant proportion of the phenotypic association between these two traits is due to genetic factors—as measured by genetic covariance. However, such studies may risk sampling and measurement bias, questioning the precision of their estimates. In the present study a whole population indirect twin design method is used to quantify additive genetic and environmental effects on the intelligence-education association using two very large contemporary European cohorts with non-self-selected sampling.

General intelligence (or IQ or psychometric general intelligence, which is usually abbreviated as *g*) is a highly stable and heritable trait (Deary et al. 2009). Across all studies, additive genetic effects account for, on average, about half of the variance in cognitive test scores (Devlin et al. 1997; Deary et al. 2006). These estimates become higher when measurement bias is low (Plomin et al. 2009), and with increasing age from childhood to adulthood. For example, in the largest twin study to date, involving 11,000 twin pairs from four countries, heritability accounted for 41% of cognitive test score variance among 9 year-olds, 55% among 12 year-olds, and 66% of the variance observed in 17 year-olds. Shared environmental factors accounted for 33%, 18% and 16% respectively (Haworth et al. 2010). These concurrent trends are also reported by longitudinal childhood studies (Bartels et al.

2002; Kovas et al. 2007), and are evident for both verbal and nonverbal intelligence (Hoekstra et al. 2007). In the present study, a Dutch cohort includes data from three age groups, allowing us to consider this shift in genetic influence on intelligence over time.

Academic achievement is a more recent subject for behaviour genetics among general populations, but it too shows a significant heritable component. In the first study of a representative sample, the variance due to additive genetic effects in age-standardised language and mathematics achievement scores of 6 to 12 year-olds was 40% (Kovas et al. 2007, using data from Thompson et al. 1991). Since then, consistently high estimates have been reported from twin cohorts, with genetic inheritance accounting for in the region of 50 to 70% variance in either: educational years, achievement test scores, or attainment levels (Bartels et al. 2002; Kovas et al. 2007; Baker et al. 1996; Johnson et al. 2006).

Given the high correlation between intelligence and educational performance in national cohorts, and the significant heritability on each trait, it was always likely that, at least some genetic factors driving academic achievements, were shared by those influencing intelligence trait variance (Martin 1975). To test this, multivariate models using twin studies have produced estimates of the proportion of the phenotypic correlation between two traits due to genetic factors and due to the shared and non-shared environment respectively. In such studies genetic covariance estimates have been high, explaining a majority of the phenotypic association. Among 190 Dutch twin pairs, bivariate models applied to educational test scores at age 12 and IQ scores at ages 5, 7, 10 and 12 years (Bartels et al. 2002) observed additive genetic covariance that increased with age (40%, 75% and 83% among 5, 7 and 9 year-olds respectively), but became weaker again at age 12 (41%), with shared environment mainly accounting for the remaining association. An Australian study of 17 year-old students, including 256 monozygotic (MZ) and 326 dizygotic (DZ) twin pairs, reported high additive genetic covariance between the Queensland Core Skills Test and verbal and performance IQ scores; 72 and 75% respectively (Wainwright et al. 2005). However, the educational measure in this study may have been more aligned to conventional intelligence tests of deductive and inductive reasoning than to tests of educational curricula. Finally, the Twins‘ Early Development Study (TEDS) analysed associations between teacher-rated academic attainment (in English, mathematics and science) and intelligence test scores of nearly 5,000 MZ and DZ twin pairs born in

England and Wales between 1994 and 1996, each assessed at ages 7, 9, and 10 years (Kovas et al. 2007). Phenotypic correlations between traits were moderate and relatively stable over time (.37 to .41), and additive genetic covariance accounted for between 63 and 90% of these associations, with shared environment explaining the remaining covariance. Furthermore, these heritable effects were stronger among the oldest (83 to 90%) relative to youngest children (63 to 73%).

It has therefore already been shown that additive genetic factors account for the majority of the phenotypic association between childhood intelligence and educational measures (at least from about 7 years), and that the degree of genetic covariance may increase with age. However, particular issues of sampling and measurement bias may limit the generalisability of these study estimates to general populations including non-twins. Firstly, ratios of MZ to DZ twins are higher among twin study volunteers, relative to the general population of twins, a factor which, if not adjusted for, may significantly influence genetic effect estimates, most likely in the direction of underestimation (Lykken et al. 1987). Secondly, the average cognitive abilities of volunteers tend to be higher than in the general population, including in twin samples (cf. Kovas et al. 2007; Rietveld et al. 2000; Thompson et al. 1991; Wainwright et al. 2005); attrition in longitudinal studies also leads to higher ability and socioeconomic groups. These factors can have the converse effect of inflating heritability estimates (Turkheimer et al. 2003). The TEDS cohort reported attrition rates at 9 and 10 years, and so it cannot be ruled out that the increased heritability estimates with age in this study are in part influenced by of this form of self-selection bias.

Different testing procedures for phenotypic measurement may also influenced genetic covariance estimates. The TEDS study had available various methods to measure intelligence at different ages, including a telephone interview at age 7 years, parent-administered tests at 9, and internet-based tests aged 10. In the other longitudinal study to report genetic covariance estimates, a different intelligence test at ages 5, 7, and 10 years was used (Revisie Amsterdamse Kinder Intelligentie Test) compared to the revised Wechsler Intelligence Scale for Children (WISC-R) at 12 years (Bartels et al. 2002). This may have contributed to the discordant result of an increase in genetic covariance from age 5 to 10 years, and then a lower estimate at age 12. Although statistical adjustment can be made for sampling bias effects or weak inter-method reliability, analysing genetic covariance among non-self-selected twins,

and using consistent methods of intelligence testing for investigating differential age effects, can be a more direct and reliable approach (Luo et al. 2003).

In the present study we use a twin extraction method applied to two national datasets from England and The Netherlands to remove the issue of self-selection of twins. The method was used previously to estimate univariate heritability of intelligence (Benyamin et al. 2005) and validated to a high level of accuracy (Webbink et al. 2006). As this procedure also identifies individuals who are non-twins (―singletons‖), we include their data to account for trait variability in the general population—a method rarely employed in the twin design. In both samples, all assessments of intelligence and academic achievement were group-administered within schools; intelligence tests were well validated, and consistent measures were used in the different Dutch age groups. Educational achievement was measured according to the national curricula of each country. These factors may help to minimise measurement bias in the present study. Given that both cohorts are at an age of studying core school subjects only, variation in educational experience is also minimised. Other factors that the two samples have in common—making possible a valid cross-national comparison of intelligence-education genetic covariance—are: their equivalent age; that they share similar academic achievement tests in language and arithmetic; and, that they are both similarly large.

Additional to the main investigation of estimating genetic covariance between intelligence and academic achievement scores among two large population cohorts, we also use these novel data and methods to explore: (i) genetic components of *specific* cognitive abilities in relation to different educational subject achievement (British cohort only); and (ii) comparisons of these effect sizes in three age groups (Dutch cohort only), given evidence for increasing genetic covariance with age (Petrill and Wilkerson 2000).

Methods

British cohort

This cohort has been characterised in a previous report (Calvin et al. 2009). It originates from a crosssectional linkage of routinely collected academic performance data from the UK Government's Department for Children, Schools and Families, and

cognitive ability test data from 2004, maintained by a national school test provider (*GL Assessment*). The participants were 178,599 11 year-old boys and girls (M = 11 years 2 months; SD = 3.5 months), from 1,531 state and specialist schools in England, representing 93% of local education authorities. Intelligence performance was measured in September to October 2004 using the age-standardised Cognitive Abilities Test — Third Edition (CAT3) of verbal reasoning (VR), nonverbal reasoning (NVR) and quantitative reasoning (QR) abilities (Smith et al. 2001). Using principal components analysis a single unrotated general intelligence factor (*g*) was extracted and retained for analysis. The subtest score loadings on *g* were as follows: QR = .87; NVR = .85; VR = .8l. Academic achievement data were taken from performance levels on Key Stage 2 curriculum tests in English, mathematics, and science, routinely group-administered within schools, using pencil and paper tests, during May 2004. The whole battery lasted five and a half hours (Directgov 2011). These tests were designed to ensure that government targets for educational standards are met in England and Wales. Each individual had a rating from 2 to 5 recorded for each subject, reflecting raw scores in the ranges: <21, 21-26, 27-32, and >32 respectively. These ratings were used as educational achievement scores in the analysis. A single educational factor score was extracted from principal components analysis, accounting for 75% of the total score variance. This score was retained for analysis. The subject score loadings on this general educational factor were as follows: English =
.86; mathematics = .89; science = .89).

Dutch cohort

This cohort has been described in detail in a previous publication (Webbink et al. 2008). It derives from the longitudinal PRIMA Survey, which aims to answer questions about educational strategies and performance in primary education in The Netherlands (Driessen et al. 1994; Driessen et al. 2004). We have incorporated data from the first five waves of the survey, which took place biannually from 1994 to 2002, and included ~60,000 pupils per wave. The PRIMA project targets a panel of ~600 schools at a time, stratified so that pupils from lower socioeconomic backgrounds are over-sampled. Measures of school performance and cognitive ability are administered to all pupils in grades 4, 6 and 8, corresponding to ages 8, 10 and 12

years. A minority of schoolchildren were included in more than one wave, so that in the present study, ~24% of those tested in grade 4 were also tested in grade 6 and, of these, ~7% were again tested in grade 8. The performance scores we have selected from this study are: general intelligence (IQ), as measured by two tests of nonverbal reasoning—specifically selected to minimise bias by socioeconomic status (Webbink et al. 2006); and educational achievement scores in arithmetic (AR) and language (LA). Inter-method reliability among the same tests at different ages are reflected in high within-subject correlations among sub-samples who were tested at more than one wave. Intraclass correlations are: IQ, .63 to 74, language, .62 to .76 and arithmetic, .70 to .83. While in the present study analyses are conducted separately for each age group, this evidence for reliability justifies a comparison of genetic estimates between the groups.

Twin identification

In the British dataset twin status was allocated if two individuals matched on surname, date of birth, and school ID. Eighty-eight individuals were removed because they were identified as triplets or quadruplets (for example, when more then two individuals matched on all three criteria), and the rest were assigned singleton status. Twins constituted 1.8% of the present study‘s cohort of singletons and twins from England—a somewhat smaller proportion than the 2.5% of individuals born as twins among total numbers of twins and singletons born in England and Wales in 1993 (Dunn and Macfarlane 1996), which may be due for example to members of a twin pair attending different schools or a cotwin being absent on the day of testing.

A similar method of twin identification was adopted in the Dutch data (Webbink et al. 2008), with ‘year of survey’ as a fourth matching variable due to the multiple-wave design. In The Netherlands in 1990 the proportion of individuals born as a twin among twin and singleton births was approximately 2.7% (Imaizumi 2003). In our cohort, twins constituted 2.0% of 8 year-olds, 1.8% of 10 year olds, and 1.6% of 12 year olds, again showing a somewhat lower representation compared to national statistics. It is comparable with the lower rate we observed in our English data.

Statistical Analyses

The intelligence and education data were screened for missing data and normality, and assessed for univariate and bivariate outliers. This was done per age group in the Dutch cohort. If a member of a twin pair was excluded due to missing or outlier scores, their cotwin was also removed. As the British educational data showed negative skew, scores of 2 were recoded as 3. Univariate outliers were values beyond 3 SD from the mean, and bivariate outliers were detected if Mahalanobis distance was equal or higher of the critical value of 13.82 ($df = 2$, $p < .001$). The reduced sample sizes appear in Tables 1 and 2.

Descriptive statistics and phenotypic correlations were produced using *PASW Statistics 17.0.3*. Phenotypic correlation coefficients were first calculated for the total cohorts, and then repeated, with the exclusion of one member from every designated twin pair, so that each data point was entirely independent. *ASReml Release 3.0* software (Gilmour et al. 2009) was used to estimate variance components and genetic and environmental effects in univariate and multivariate linear mixed models with the restricted maximum likelihood method (Visscher et al. 2004). Without known zygosity in the present study's data, variance components for opposite-sex (OS) and same-sex (SS) twin pairs were derived instead of those for DZ and MZ twin pairs (see Benyamin et al 2005; Scarr-Salapatek 1971; Visscher et al 2004). This method uses the knowledge that all OS twins are DZ and depends on the assumption that twice the proportion of OS twins is the proportion of DZ in the population[1]. Essentially the estimated proportions of covariance between twin pairs, attributed to additive genetic (A), shared environment (C), and unique environment (E) effects (Neale & Cardon 1992), are weighted by the estimated proportion of MZ twins among SS pairs (*pMZSS*), calculated with the formula[2]:

$$MZSS = OS - OS \cdot p = 1 - 2p / 1 - p$$

where *OS p* is the proportion of OS twins among total twin pairs

In the present study's Dutch and British datasets these parameter estimates were 0.58 (averaged across age groups) and 0.53 respectively. These are based on estimates of

[1] This assumption is based upon the further assumption of a 1:1 sex ratio and equal survival, stemming from Weinberg's differential rule (Weinberg 1901)

[2] Scarr-Salapatek (1971)

the proportion of MZ twins among total twin pairs of 36.2% in England and 40.6% in The Netherlands, which can be compared to national statistics of 35.8% and 34.6% respectively, reported by Imaizumi 2003. In linear models *sex* and *age* were entered as fixed effect variables, as was *year of test* in the models using Dutch data, to take out variation due to systematic effects. Given the inclusion of singleton data in the present study (unusual for a standard ACE twin design), the variables *pair* and *SS pair* were entered as random (latent) effects. The variable *pair* was a unique identifier assigned to each individual in the data unless it was shared by a member of a twin pair. The variable *SS pair* was also a unique identifier given to each individual unless shared by a member of a SS twin pair:

$y = \mu + pair + SS\ pair + residual$
with $var(y) = var(pair) + var(SS\ pair) + var(residual)$

In terms of covariances between twin pairs this implies:

$cov(SS\ pair) = var(pair) + var(SS\ pair)$
$cov(OS\ pair) = var(pair)$

The estimated variance components are therefore linear functions of the covariances of the SS and OS twin pairs, and their expectations are:

$E[var(SS\ pair)] = cov(SS\ pair) - cov(OS\ pair)$
$E[var(pair)] = cov(OS\ pair)$

Models of the underlying variance components in the population (e.g., A, C and E) are linear combinations of the covariances of SS and OS pairs, and therefore linear combinations of the estimated variance components.

In post-analytic procedures, variance components from REML were transformed and weighted using *pMZSS* to estimate variation due to additive genetic (*h*2), shared environment (*c*2) and unique environment (*e*2) effects among individual traits. These estimated parameters and the transformed and weighted cross-trait variance components were then used to estimate genetic covariance between traits (*Biv h*2)—the proportion of their phenotypic correlation due to additive genetic

factors (or the ratio of genetic to phenotypic covariance)—as well as covariance due to shared environment (*Biv c*2) and unique environment (*Biv e*2) effects. Sampling variances were calculated from a first order Taylor series of the likelihood function about the maximum likelihood estimates. Appendix C presents the formulae used in the post-analytic procedures.

Results

Comparison of singletons, opposite-sex (OS) twins and same-sex (SS) twins

Negligible differences were observed in the means and SDs of singletons, opposite-sex twins and same-sex twins on all cognitive and educational variables, in both the British and Dutch datasets (see Tables 1 and 2). These findings are consistent with previous reports from these two cohorts comparing singletons and total twin pairs (Calvin et al. 2009; Webbink et al. 2008).

Univariate models

Tables 1 and 2 include the intra-class correlations for OS and SS twin pairs; that these were higher for SS pairs, which would include all MZ twins and some DZ twins from the total sample, compared to OS twins who were all DZ, indicates additive genetic influences on all academic and intelligence variables. This was confirmed by the univariate estimates of additive genetic and environmental effects (see Figure 1). In the British dataset *g* was strongly heritable at age 11 ($h2$ = .70 ±.14—which herein denotes the estimate and standard error). Verbal and nonverbal reasoning scores showed higher heritability estimates ($h2$ = .79 ±.14, and .70 ±.16, respectively) than quantitative reasoning scores ($h2$ = .50 ±.15). The Dutch data support evidence that additive genetic effects on intelligence test performance increase with age, in our study from 8 years ($h2$ = .43 ±.23) to 12 years ($h2$ = .67 ±.31). The data show an unexpectedly low heritability estimate at age 10 years ($h2$ =.24 ±.29), with a larger influence of shared environmental effects ($c2$ =.29 ±.04) than might be expected given the existing literature.

Academic achievement in England‘s 11 year-olds was strongly influenced by

additive genes, and the heritability estimate for the education factor score ($h2$ = .75 ±.14) was equivalent to that of the intelligence factor, *g*. Among specific academic subjects, English was most strongly influenced by additive genetic effects ($h2$ = .81 ±.16), followed by mathematics ($h2$ = .66 ±.16) and science ($h2$ = .51 ±16). In Dutch schoolchildren, language achievement showed greatest additive genetic influence in the younger cohort ($h2$ = .74 ±.27), compared to ages 10 and 12 ($h2$ = .43 ±.29; .53 ±.31, respectively); of the effect remaining, unique environmental influences were more important than shared environment. Arithmetic performance was largely heritable at 8 and 10 years ($h2$ = .67 ±.27; .73 ±.29 respectively), and unique environment explained the remaining variance. At age 12, additive genetic effects were relatively less important ($h2$ = .36 ±.27), and shared environment was more influential than at younger ages ($c2$ = .35 ±.04).

Multivariate models

Phenotypic correlations between all cognitive and academic performance variables, in both the British and Dutch cohorts, were all positive and statistically significant ($p < .05$), with effect sizes larger in England (r = .52 to .84) compared to The Netherlands (r = .36 to .47) where the magnitude of associations were consistent across different ages. For example, at ages 8, 10 and 12, IQ-language correlations were r = .38, .36 and .36, and IQ-arithmetic correlations were r = .44, .47, and .47, respectively. On the other hand, correlations between the two academic subjects in The Netherlands showed a slight increase with older age groups: r = .43, .50 and .55, respectively.

Intelligence was more strongly associated with arithmetic achievement compared to language achievement in both national cohorts. Bar heights in Figure 2 indicate the phenotypic correlations for each cohort. In England, high genetic correlations with low standard errors were observed between cognitive subtest scores and academic attainments (rG = 0.58 to 0.97) (see Table 3). Correlations of shared environmental effects were markedly high for some cross-trait pairings, for example of verbal reasoning with mathematics and science (rC = 0.86 ±.36; 1.0 ±.29, respectively), and quantitative reasoning with these two subjects (rC = 1.00 ±.32; 1.00 ±.33). By contrast, shared environment correlations were low for nonverbal reasoning with mathematics and science achievements (rC = 0.05 ±.90; 0.24 ±.59, respectively),

although in these latter models the standard errors were high.

A substantial proportion of the intelligence-education phenotypic correlations in the England cohort were due to additive genetic influence, indicated by high genetic covariances (*Biv h*2) in the range 0.53 to 1.00, with standard errors of between 0.13 and 0.21 (see Figure 2). These were consistently high between verbal reasoning and academic achievements (genetic covariance, *Biv h*2 = 0.77 to 1.00), as well as nonverbal reasoning and the achievement scores (*Biv h*2 = 0.94 to 0.98). Where environmental effects influenced phenotypic associations between quantitative reasoning and academic achievement, shared environment was somewhat more important in explaining the relationship (shared environment covariance, *Biv c*2: 0.22 to 0.34), than was the unique environment (*Biv e*2: 0.08 to 0.14).

Table 4 shows the genetic correlations between intelligence and academic performance scores in The Netherlands. Genetic correlations between IQ and language, and IQ and arithmetic, were moderate across all ages (*r*G = 0.31 to 0.58), with the exception of the cross-trait genetic correlation between IQ and language at age 8, which was higher and more consistent with the UK data (*r*G = 0.82). An unexpected finding was that whereas shared environment correlations for IQ and educational attainment were negligible at 8 years and 12 years, among 10 year-olds they were high, albeit with very large standard errors (*r*C = 0.74 ±0.56 for language; 1.79 ±3.02 for arithmetic).

Figure 2 shows the genetic and environmental covariance estimates from the Dutch cohort. To consider the age 8 and 12 year data only for a moment, these show that the language-IQ phenotypic associations were largely explained by additive genetic effects at both ages (*Biv h*2 = 1.00 ±.43 and 0.92 ±.59 respectively). Additive genes also explained the majority of the phenotypic association between arithmetic and IQ within these age groups (*Biv h*2 = 0.63 ±.51 at age 8; 0.64 ±.42 at age 12), and the remainder was explained more by the shared than the unique environment (*Biv c*2 = 0.30 and 0.20; *Biv e*2 = 0.07 and 0.16 respectively). In contrast, at age 10, shared environment was more influential than additive genetic effects on the IQ and education phenotypic association (*Biv h*2 = 0.33 ±.52, *Biv c*2 = 0.49 ±0.38 for language; *Biv h*2 = 0.36 ±.38, *Biv c*2 = 0.45 ±.28 for arithmetic).

Discussion

The main aim of this study was to report on genetic and environmental influences on associations between intelligence test scores and academic achievement scores in two large, nationwide samples of schoolchildren. While this is a relatively well-researched hypothesis in the behaviour genetics literature, the sampling method is novel in this context, providing cross-trait estimates using data on non-self-selected twin pairs and including trait distributions of singletons. Further to eliminating self-selection bias, the testing procedures for our variables were consistent across age groups, showing good reliability. These methods give confidence when generalising estimates to national populations. The substantial cross-trait heritabilities in England and The Netherlands we report are consistent with the two most comparable previous studies of similarly aged children (Bartels et al 2002; Kovas et al 2007) and a third study of older children (Wainwright et al. 2005), in which pre- and/or self-selection of twin pairs may have increased genetic estimates. A strong additive genetic influence on intelligence and academic achievement associations is indicative of a more fundamental cognitive trait that is genetically influenced, and which drives a substantial proportion of the association between the two rather than, for example, individual differences in educational opportunity causing variance in intelligence test performance. Some evidence suggests that processing speed might be a reliable indicator of this low-level antecedent trait (Luo et al. 2003; Luciano et al. 2001), although in one study this variable shared a lower genetic correlation with academic achievement relative to other cognitive ability tasks, including memory and spatial reasoning (Thompson et al. 1991). We also found that shared environment explained most of the remaining variance in England and The Netherlands, that is, for the associations between *g* (or IQ) and academic subject achievements, replicating previous studies‘ results (Bartels et al. 2002; Kovas et al. 2007; Wainwright et al. 2005).

We found some lower genetic covariances, albeit with larger standard errors, in the Dutch relative to British cohorts. There may be several explanations for this. One influence might be that the CAT-3 intelligence test used on the British cohort was designed to minimise demand on verbal skills (Smith et al. 2001), making it a less culturally-biased intelligence indicator, and thus perhaps optimising the heritability estimates. An additional explanation for the relatively lower estimates in the Dutch cohort may be in its stratified sampling that increased representation of low

socioeconomic groups in the PRIMA Survey—particularly as lower heritability estimates for intelligence have been reported in these groups (Turkheimer et al. 2003). Nevertheless, there was notable consistency across the two cohorts‘ equivalent age groups; for example, in the additive genetic covariance estimates of intelligence with language and arithmetic/mathematics respectively, among 12 year-olds in The Netherlands and 11 year-olds in England. Another consistency was the greater additive genetic influence on intelligence-language associations, relative to intelligence-arithmetic associations, which is reported in the existing literature (Luo et al. 2003; Kovas et al. 2007).

The different age groups of the Dutch schoolchildren provided the opportunity to consider changes in the magnitude of additive genetic influence on intelligence-education associations from age 8 to age 12, albeit using cross-sectional inference. Despite a general trend for genetic covariance to increase with age (Bartels et al. 2002; Kovas et al. 2007; Petrill and Wilkerson 2000), one study reported a considerable decrease in the genetic covariance from 10 to 12 years (Bartels et al. 2002). Our findings showed no significant change in genetic covariance estimates between the youngest and oldest age groups, however, and again this may be because their standard errors were typically very large. It has been suggested that an initial increase in additive genetic influence at elementary school age may reflect emergent functionality of higher level cognition or a transition from rote learning to more complex academic engagement (Wainwright et al. 2005). Longitudinal twin studies running into secondary education would help to demarcate the ages at which additive genetic influences on intelligence-academic achievement associations shift in their relative importance.

Among twin studies of the association between cognitive performance and academic achievement, fewer have looked beyond a general intelligence score to using more specific functional tests of cognition. In one study, performance on verbal, spatial and memory tasks showed consistently high genetic correlations with language and mathematics achievement scores, and a processing speed task showed a relatively lower genetic correlation with academic achievement (Thompson et al. 1991). In another, a higher genetic correlation between verbal IQ and achievement scores was reported, relative to a performance IQ measure, although additive genetic effects explained an equally high proportion of the phenotypic association in both cases (Wainwright et al. 2005). In the present study‘s British cohort, very high genetic

correlations were observed between verbal reasoning and English language achievement, and nonverbal reasoning with mathematics and science achievements, which were greater than for other combinations of cognitive and academic variables. However, consistent with Wainwright et al. (2005) we found equivalent and consistently very high genetic covariances between all three subject-specific achievement scores and nonverbal and verbal reasoning respectively. This may further validate the proposition that additive genetic influences on a general cognitive processing trait drive associations with educational achievements, and that the effect of schooling in explaining variance in psychometric *g*, at least in primary school education, is a weakened proposition (Luo et al. 2003). Although we were able to look at performance on three distinct academic subjects in relation to specific cognitive subtest scores, we were not on the other hand able to consider different measures within academic subjects, which may differentially affect genetic covariance estimates. For example, in a U.S. twin study of elementary schoolchildren mathematical fluency and problem solving but not computational mathematics shared genetic covariance with cognitive ability (Hart et al 2009). Studies such as this one, in an expanding area of behaviour genetics research with increasing complexity of multivariate models, include related performance measures such as reading, thereby deriving more reliable heritability estimates for the associations between cognitive ability and specific academic abilities once these covariates are controlled for. With the population-level cohort data available to us in the present study however, specific educational performance measures, and reading, could not be explored.

We minimised several key sources of potential error in our study, including measurement bias (using well-standardised tests administered in the stable classroom environment) and self-selection bias (random sampling from national cohorts). However, our twin extraction method could have incurred error, although a validation study reported a low risk of statistical bias using this method (Webbink et al. 2006). Error that did occur was due to false negative rather than false positive detection. Twins may have been under-detected in the present study (if for example, a cotwin was absent on the day of testing, had missing data, or attended a different school), particularly as we found a somewhat lower proportion of total twins in our samples compared to national twin birth rates. This may compromise how twin population-representative our two cohorts might be. However, false negatives would have had to affect MZ twin pairs to a greater or lesser extent than DZ twins, in order to have

influenced our effect sizes, and there is little to indicate that this might have been the case. Furthermore, our estimates of the proportions of MZ twins among total twin pairs were similar to national birth statistics for the appropriate years (Imaizumi 2003), albeit there was a slight overestimation of the MZ twin proportion in the Dutch cohort. A second potential source of error is the statistical approach, which is based on further assumptions than traditional approaches where zygosity is known. In particular, our heritability estimates were derived from variance components of OS and SS twin pairs, weighted on the basis of statistical assumption. This was an estimation of the proportion of MZ twins among SS pairs, based upon the approximation that DZ twins represent two times the proportion of OS twins in the population, and assumptions of a 1:1 sex ratio among DZ twin births and an equivalent survival of male-male and female-female DZ twins. Furthermore, if there are sex-specific covariance estimates then our results risk bias.

Without known zygosity we could not directly validate our twin extraction method within the present study. An indirect method however is to look for consistency between our univariate estimates of intelligence and education heritabilities, with those from the existing literature, which we did find. In the present study, Dutch heritability estimates for intelligence were consistent with previous studies of similarly aged children (Bartels et al. 2002; Haworth et al. 2010; Kovas et al. 2007). Our relatively low heritability estimate of intelligence in 10 year-old schoolchildren, and a greater shared environment estimate compared to other age groups, may be explained by large overlapping standard errors of these estimates. In England, intelligence heritability was in the higher range of the existing literature, and this may be due to the measure‘s effective removal of cultural bias. Conversely, relatively lower heritability estimates in The Netherlands may have been affected by the over-sampling of children from lower socioeconomic backgrounds in this study cohort. Furthermore, different educational policies of The Netherlands and England may also influence disparities between these two countries‘ heritability estimates. For example, the relatively lower heritability of intelligence test performance in The Netherlands may be influenced by a greater effort by that country‘s government to create national equality in educational opportunities; Dutch schools receive additional personnel and resources for every additional pupil they teach who is from a disadvantaged background (Lindahl et al. 2007), which is likely to reduce genetic effects on the performance of tests taken in the school setting. On academic

achievements, additive genetic effects on Dutch language were consistent with a previous estimate from the CITO test (Bartels et al. 2002), and heritability estimates of mathematics and science attainments in the British sample were compatible with equivalent attainment scores in the UK TEDS study (Kovas et al. 2007). However, we found additive genetic effects accounting for 81% of the variance in English attainment compared to 60% in TEDS. The difference here may be that the present study‘s education data were based on a nationally standardised test score, compared to teacher ratings of scholastic achievement in the TEDS sample, which may have incurred within-classroom effects.

We overcame some key caveats of interpreting additive genetic effect estimates from ACE models (Neale 2009), including our use of large, non-preselected and nationally selected datasets, and by incorporating singleton data in our models to account for trait variability in the general population. However, there are limitations of a standard twin design that the present study shares. Firstly, the analytic method implicitly assumes that twin similarity due to common environmental effects is the same for MZ and DZ twins. A second assumption is that the genetic correlation between males and females is 1.0. Nevertheless, intelligence heritability estimates reported by twin studies depending on these first two assumptions, have recently been validated by the first genome-wide association study using biological samples of unrelated adults, which found that genetic variation explained ~50% of individual differences in fluid intelligence (Davies et al. 2011). Thirdly, and specific to an analysis of twins where zygosity is unknown, is the assumption that among DZ twins the effect of shared environment is the same for SS and OS twin pairs. In future, biological samples collected from national twin registries will enable a more thorough investigation of familial factors in the covariance between intelligence and academic achievements (Kaprio 2011).Preliminary studies have so far suggested similar chromosomal regions relating intelligence to reading (Luciano et al. 2006; Posthuma et al. 2005) and educational performance (Wainwright et al. 2006). Such findings would require replication however if they were to be validated, and a persistent challenge to molecular genetics research is the need for massive databases to achieve adequate statistical power in detecting small effects, of what are likely to be several implicated genes (Fisch 2009).

Our observation of strong additive genetic covariance between intelligence test scores and academic achievements in elementary schoolchildren, consistent with

estimates from self- and/or preselected twin samples, does not refute the position that psychometric scores may be improved in response to specific educational interventions (Petrill and Wilkerson 2000). However, in England and The Netherlands where the school curricula are assessed by standardised tests, this scenario is not evident. We found negligible change in the strong effects of additive genes on intelligence-education associations with age, at a period when children move from elementary to secondary schooling.

Acknowledgements

The authors thank the UK Government‘s Department for Education for granting permission to use the Key Stage 2 educational data and GL Assessment for providing both data for the British cohort. The Dutch data are already available to the public. The work was undertaken by The University of Edinburgh Centre for Cognitive Ageing and Cognitive Epidemiology, part of the cross council Lifelong Health and Wellbeing Initiative (G0700704/84698). Funding from the Biotechnology and Biological Sciences Research Council (BBSRC); Engineering and Physical Sciences Research Council (EPSRC); Economic and Social Research Council (ESRC); and Medical Research Council (MRC) is gratefully acknowledged. ML is a Royal Society of Edinburgh/Lloyds TSB Foundation for Scotland Personal Research Fellow.

References

Baker LS, Treloar SA, Reynolds CA, Heath AC, Martin NG (1996) Genetics of educational attainment in Australian twins: sex differences and secular changes. Behav Genet 26:89–102

Bartels M, Rietveld MJH, Van Baal CGM, Boosma DI (2002) Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. Twin Res 5:544–553

Benyamin B, Wilson V, Whalley LJ, Visscher PM, Deary IJ (2005) Large, consistent estimates of the heritability of cognitive ability in two entire populations of 11-year-old twins from Scottish Mental Surveys of 1932 and 1947. Behav Genet 35:525–34

Calvin C, Fernandes C, Smith P, Visscher PM, Deary IJ (2009) Is there still a cognitive cost of being a twin in the UK? Intelligence 37:243–248

Calvin CM Deary IJ, Fenton C, Roberts BA, Der G, Leckenby N, Batty GD (2011) Intelligence in youth and all-cause-mortality: systematic review with meta-analysis. Int J Epidemiol 40:626-644

Davies G, Tenesa A, Payton A, Yang J, Harris SE, Liewald D et al (2011) Genome-wide association studies establish that human intelligence is highly heritable and polygenic. Mol Psychiatry 16:996-1005

Deary IJ, Johnson W (2010) Intelligence and education: causal perceptions drive analytic processes and therefore conclusions. Int J Epidemiol 39:1362–1369

Deary IJ, Johnson W, Houlihan LM (2009) Genetic foundations of human intelligence. Hum Genet 126:215–232

Deary IJ, Spinath FM, Bate TC (2006) Genetics of intelligence. Eur J Hum Genet 14:690–700

Deary IJ, Strand S, Smith P, Fernandes C (2007) Intelligence and educational achievement. Intelligence 35:13–21

Devlin B, Daniels M, Roeder K (1997) The heritability of IQ. Nature, 388:468–471

Directgov (2011) National Curriculum teacher assessments and key stage tests. Webpage retrieved on July 7, 2011: http://www.direct.gov.uk/en/Parents/Schoolslearninganddevelopment/ExamsTestsAndTheCurriculum/DG_10013041

Driessen G, Van Langen A, Oudenhoven X (1994) De toetsen voor de cohort Primair onderwijs, verantwoording. Nijmegen ITS.

Driessen G, Van Langen A, Vierke H (2004) Basisrapportage PRIMA-

cohortonderzoek, Vijfde meting 2002–2003. (Report on PRIMA-longitudinal research project, Survey 2002–2003). Nijmegen.
Dunn A, Macfarlane A (1996) Recent trends in the incidence of multiple births and associated mortality in England and Wales. Arch Dis Child Fetal Neonatal Ed 75:F10-19
Fisch GS (2009) Models of human behaviour: talking to the animals. In Y-K Kim (Ed.), Handbook of Behavior Genetics. New York, US: Springer.
Gilmour AR, Gogel BJ, Cullis BR, Thompson R (2009) ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, HP1 1ES, UK. www.vsni.co.uk
Hart SA, Petrill SA, Thompson LA, Plomin R (2009) The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. J Ed Psych 101:388-402
Haworth CMA, Wright MJ, Luciano M, Martin NG, de Geus EJC, van Beijsterveldt CEM et al (2010) The heritability of general cognitive ability increases linearly from childhood to young adulthood. Mol Psychiatry 15:1112–1120
Hoekstra RA, Bartels M, Boomsma DI (2007) Longitudinal genetic study of verbal and nonverbal IQ from early childhood to young adulthood. Learn Individ Differ 17:97–114
Imaizumi Y (2003) A comparative study of zygotic twinning and triplet rates in eight countries, 1972-1999. J Biosoc Sci 35:287-302
Johnson W, McGue M, Iacono WG (2006) Genetic and environmental influences on academic achievement trajectories during adolescence. Dev Psychol 42:514–532
Kaprio J (2011) Specific advantages of twin registries and biobanks. In JM Walker (Ed.). Methods in Molecular Biology. New York, US: Springer.
Kovas Y, Haworth CMA, Dale PS, Plomin R (2007) The genetic and environmental origins of learning abilities and disabilities in the early school years. Monogr Soc Res Child Dev Serial No. 288, 72(3)
Lindahl E, Lindahl M, Oosterbeek H, Webbink D (2007) The effect of extra funding for disadvantaged pupils on achievement. Rev Econ Stat 89:721–736
Lleras-Muney A (2005) The relationship between education and adult mortality in the United States. Rev Econ Stud 72:189–221
Luciano M, Smith GA, Geffen GM, Geffen LB, Martin NG (2001) Genetic covariance among measures of information processing speed, working memory, and IQ. Behav Genet 31:581–592

Luciano M, Wright MJ, Duffy DL, Wainwright MA, Zhu G, Evans DM et al (2006) Genome-wide scan of IQ finds significant linkage to a quantitative trait locus on 2q. Behav Genet 36:45–55

Luo D, Thompson LA, Detterman DK (2003) Phenotypic and behavioural genetic covariation between elemental cognitive components and scholastic measures. Behav Genet 33:221–246

Lykken DT, McGue M, Tellegen A (1987) Recruitment bias in twin research: The rule of two-thirds reconsidered. Behav Genet 17:343–362

Martin NG (1975) The inheritance of scholastic abilities in a sample of twins. Ann Hum Genet 39:219–29

Neale MC (2009) Biometrical models in behavioural genetics. In Y-K Kim (Ed.), Handbook of Behavior Genetics. London: Springer.

Neale MC, Cardon LR (1992) Methodology for Genetic Studies of Twins and Families. Dordrecht: Kluwer.

Petrill SA, Wilkerson B (2000) Intelligence and achievement: a behavioural genetic perspective. Educ Psychol Rev 12:185–197

Plomin R, DeFries JC, McClearn GE, McGuffin P (2009) Behavioral Genetics (fifth edition). Worth Publishers, New York.

Posthuma D, Luciano M, de Geus EJC, Wright MJ, Slagboom PE, Montgomery GW, Boomsma DI, Martin NG (2005) A genomewide scan for intelligence identifies quantitative trait loci on 2q and 6p. Am J Hum Genet 77:318–326

Richards M, Sacker A (2011) Is education causal? Yes. Int J Epidemiol 40:516–518

Rietveld MJH, van Baal GCM, Dolan CV, Boosma DI (2000) Genetic Factor Analyses of Specific Cognitive Abilities in 5-Year-Old Dutch Children. Behav Genet 30:29–40

Scarr-Salapatek S (1971) Race, social class, and IQ. Science 174:1285–1295