Developmental Origins of Low Mathematics Performance and Normal Variation in Twins from 7 to 9 Years

Claire M. A. Haworth,[1] Yulia Kovas,[1] Stephen A. Petrill,[2] and Robert Plomin[1]

[1] Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, De Crespigny Park, London, United Kingdom

[2] Department of Human Development and Family Science, Ohio State University, Columbus, Ohio, United States of America

A previous publication reported the etiology of mathematics performance in 7-year-old twins (Oliver et al., 2004). As part of the same longitudinal study we investigated low mathematics performance and normal variation in a representative United Kingdom sample of 1713 same-sex 9-year-old twins based on teacher-assessed National Curriculum standards. Univariate individual differences and DeFries-Fulker extremes analyses were performed. Similar to our results at 7 years, all mathematics scores at 9 years showed high heritability (.62–.75) and low shared environmental estimates (.00–.11) for both the low performance group and the full sample. Longitudinal analyses were performed from 7 to 9 years. These longitudinal analyses indicated strong genetic continuity from 7 to 9 years for both low performance and mathematics in the normal range. We conclude that, despite the considerable differences in mathematics curricula from 7 to 9 years, the same genetic effects largely operate at the two ages.

Some children understand concepts they are taught much faster than their peers. When in the same schools and following the same curriculum, what is it that makes some children capable while others struggle to grasp even the basics? Research into the origins of academic abilities has centered on reading and general cognition, and the associated problems that children may have with these skills. Little research has investigated the etiology of mathematical skills, and even less research has focused on those children who are not able to acquire the mathematical skills appropriate for their age group. The relative contributions of nature and nurture to mathematical performance throughout development have yet to be thoroughly investigated.

Research Into Mathematics Performance

Previous studies have reported varying estimates of the contribution of genetics to mathematics performance that ranged from .20 (Thompson et al., 1991) to .90 (Alarcón et al., 2000). Estimates of group heritability of low mathematics performance have also been varied, ranging from .38 (Alarcón et al., 1997) to .65 (Oliver et al., 2004). These differing estimates for both mathematics and low mathematics performance may be due to methodological issues such as the use of different age groups, measures and samples. More research is needed on the etiology of mathematics performance to clarify these discrepancies using a large representative sample. Also unknown is the extent to which the etiology changes during development, as, for example, the heritability of general cognitive ability increases during childhood (Plomin et al., 1997). In addition to age differences in heritability, no longitudinal analyses have been reported that examine genetic and environmental contributions to age-to-age change and continuity for mathematics.

In a previous report using the same large twin sample as the present study, the genetic and environmental etiology of early mathematics performance at 7 years was investigated both for the entire range of variation and for the low extreme of performance (Oliver et al., 2004). This research was based on teacher reports of year-long observations of children's performance in mathematics. In England and Wales, schools must follow the National Curriculum (NC), which sets standards for achievement in each subject; these standards are arranged into four key stages that children progress through between the ages of 5 and 16 years. Because the children were 7 years old, teachers were asked to report performance based on specific NC criteria for Key Stage 1. The etiology of mathematics performance was analyzed using the twin method for the entire spectrum of a representative community sample of more than 2000 pairs of twins, and also for mathematics performance in the lowest 15% of the same sample. Mathematics performance at 7 years as assessed by teachers' reports based on NC criteria showed high heritability both

for the entire spectrum of performance (66%) and for the lowest mathematics performance (65%). These results suggest that the same genetic factors affect the entire normal distribution and the low extreme of the distribution (Plomin & Kovas, 2005). In other words, low mathematics performance appears to be the quantitative extreme of the same genetic factors responsible for variation throughout the distribution. A second finding was that shared environmental influence was minimal (less than 10%) both for the entire sample and the low extreme. This finding implies that growing up in the same family, going to the same schools, and in most cases being in the same classroom contributes only marginally to individual differences in mathematical performance at 7 years. In other words, the environment is important for mathematical development, accounting for as much as a third of the total variance, but whatever these environmental factors may be, they are of the nonshared variety.

The Present Study

Is there something peculiar about early mathematical performance at 7 years that would yield such results? The purpose of the present study is to conduct similar analyses when the same sample was 9 years old and evaluated on NC Key Stage 2 criteria.

Between the ages of 7 and 9 years, children in schools in England and Wales transfer between two key stages in the NC and make great advances in their mathematical understanding. At the age of 7 years (Key Stage 1), children are taught to count and to read and write numbers. By the age of 9 years (Key Stage 2), children are expected to manipulate numbers and understand relationships between fractions and derive division facts from their times tables (DfEE Publications, 1999). These great strides in mathematics learning could produce different genetic and environmental influences at 7 and 9 years. As noted, at 7 years, we found that low mathematics performance and variation in the normal range are both highly heritable and that environmental effects are mainly of the nonshared rather than the shared type of environmental influence. The major advances in mathematics development from 7 to 9 years and the results at 7 years showing such high heritability and low shared-environmental influence led us to predict that we would find different results at 9 years than we found at 7 years. A novel feature of the present study is that it provides the first longitudinal genetic analysis of mathematics development for both the normal distribution and the extremes. Again, given the considerable changes in mathematics learning from 7 to 9 years, we expected to find that different genetic and environmental factors contribute to mathematics ability and disability at 7 and 9 years. That is, we expected that genetic and environmental factors would contribute to change in mathematics performance from 7 to 9 years.

Materials and Methods

Sample

The sampling frame for the present study was the Twins' Early Development Study (TEDS), a study of twins born in England and Wales in 1994, 1995, and 1996 (Oliver & Plomin, 2007; Trouton et al., 2002). The TEDS sample has been shown to be reasonably representative of the population (see Oliver et al., 2004 for details of the 7-year sample). The 9-year sample is similar to the 7-year sample, with 80% of the sample the same. TEDS is described in more detail elsewhere (Oliver & Plomin, 2007; Spinath et al., 2003).

From the 1994 and 1995 cohorts of the TEDS sample 4077 families agreed to participate in the 9-year testing. Ninety-five per cent of this sample (3859 families) agreed to allow us to contact the current teachers of the twins, and provided school details. Teachers were contacted when the children were towards the end of their fourth year of primary school so that the teachers would be familiar with the children's performance during the school year. Teacher forms for both members of a twin pair were distributed at the same time. When the same teacher assessed both twins in a pair, responses for the twins were received simultaneously; when different teachers assessed members of a twin pair, responses were usually received within a few days of each other, although some pairs were assessed a few weeks apart. As expected, the correlations between the date of the teacher questionnaire being returned and the mathematics scores were low (–.003 to .02), indicating negligible effects of time of teacher assessment. Teachers were sent a covering letter with the background and aims of TEDS, as well as an explanation that we had obtained consent from the twins' parents to ask teachers for information about the child's performance at school. Teachers were asked to check one of five boxes to indicate level of attainment in terms of the NC criteria (see Measures). Of the teacher questionnaires sent, 5836 individual forms (76%) were returned complete. Three hundred and sixty-six families were excluded as we received questionnaires from only one member of the twin pair, leaving 2735 families (5470 children). Of the questionnaires returned, 62% of the twins were assessed by the same teacher. There was no bias toward monozygotic (MZ) twins being kept together in the same classroom: 63% of MZ twins versus 62% of dizygotic (DZ) twins had the same teacher.

For the purposes of the current study, we excluded 265 families in which at least one member of the twin pair had a specific medical syndrome or was an extreme outlier for perinatal problems such as extreme low birthweight, or for whom zygosity information was not available. We further excluded 132 twin pairs, for the individual differences analysis, when at least one twin scored more than 3 standard deviations below or above the mean on any of the three mathematics scales or the composite score so that the

disproportionate influence of extreme outliers was negated. After these exclusions the sample included 4940 children (2470 families). We further excluded 757 families with opposite-sex twins as the DeFries-Fulker (DF) extremes analysis technique (see Analyses section) allows for same-sex pairs only.

Thus, after exclusions, the entire sample for the 9-year analysis was 1713 pairs of twins, consisting of 790 boy pairs and 923 girl pairs, and 893 MZ and 820 DZ same-sex twin pairs. The mean age of the twins when questionnaires were returned from the teachers was 9.04 (range = 8.46–10.54). Zygosity was assessed initially through a parent questionnaire of physical similarity, which has been shown to be over 95% accurate when compared to DNA testing (Price et al., 2000). Subsequently as DNA has been obtained for most twin pairs, zygosity was confirmed using DNA markers (Freeman et al., 2003).

Probands

In order to investigate the etiology of low mathematics performance at 9 years of age and its links with variability in the normal range, we selected twin pairs where at least one (or both) twin(s) scored in the lowest 15% on each of the three standardized mathematics scores and a composite score (described below). The cut-off was chosen in part because it provides a reasonable sample size in terms of power for DF extremes analysis. This cut-off yields a group of children with clear problems in mathematics development according to their teachers. For example, using the raw scores (i.e., uncorrected for age and sex), in the Shapes, Space and Measures subtest 1.6% of the sample were given a score of 1, indicating achievement well below the expected standard for 9-year-olds, and 18.7% of the sample scored below a score of 3, indicating achievement below, or well below, the expected standard.

Although the majority of children scoring in the low group for one mathematics dimension also scored in the low group for the other dimensions, some children were selected as low probands for one mathematics score, but not for another. For the low 'using and applying mathematics' group there were 384 families (548 probands), 187 MZ and 197 DZ same-sex pairs. For the low 'numbers' group there were 396 families (535 probands), 198 MZ and 198 DZ same-sex pairs. For the low 'shapes, space and measures' group there were 389 families (537 probands), 193 MZ and 196 DZ same-sex pairs. For the mathematics composite measure, the low group included 394 families (539 probands) in 200 MZ and 194 DZ same-sex pairs.

Longitudinal Sample

The sample used for the 7-year data has been described previously (Oliver et al., 2004) and consisted of 2178 pairs of twins; 1027 male pairs and 1151 female pairs, of whom 1146 were MZ twins and 1032 were DZ same-sex twins. For the individual differences longitudinal analysis all pairs were used; 2178 pairs at 7 years and 1713 pairs at 9 years. For the DF extremes longitudinal analysis only those probands at each age with co-twin data at the other age were used, although cut-off variables were calculated on the whole sample. Two types of DF extremes longitudinal genetic analyses were conducted: prospective and retrospective. For the prospective analysis, low-scoring probands were selected at 7 years and co-twins' scores were compared at 9 years; the sample consisted of 187 MZ probands and 209 DZ probands. For the retrospective analysis, low-scoring probands were selected at 9 years and compared to co-twins' scores at 7 years; there were 221 MZ probands and 163 DZ probands.

Measures

As for all children, the twins' mathematics performance was assessed throughout the fourth year of school by their teachers, using the assessment materials of the NC for England and Wales, the core academic curriculum developed by the Qualifications and Curriculum Authority (QCA). Assessment at the end of key stages involves two types of measurement, NC Direct Testing and NC Teacher Assessments. The NC Teacher Assessments consist of teachers giving a score from a 5-point scale on the basis of the child's performance throughout the school year. For the purposes of the current study, these NC Teacher Assessments at Key Stage 2 were used, which are familiar to teachers and are designed for children aged 8 to 11 years. For Key Stage 2, the QCA provides teachers with NC material and assessment guidelines for three strands of mathematics which directly map on to areas in mathematics that are taught throughout the NC at this stage: using and applying mathematics; numbers and algebra; and shapes, space and measures (DfEE Publications, 1999). (See http://www.ncaction.org.uk/subjects/maths/levels.htm for the 5-point NC criteria used by teachers to indicate achievement levels in each of the three areas of mathematics.) Along with the NC Direct Testing score, the NC Teacher Assessments score given by the teacher on these NC criteria for a particular child ultimately determines the final score that is submitted to the QCA for that child at the end of the key stage. For the purposes of the present study, teachers were asked to check one of five boxes to indicate the child's NC Teacher Assessments score. Reminders of the NC criteria used to select the appropriate attainment level were provided as part of the questionnaire. Further details about these measures have been published previously (Walker et al., 2004).

It should be emphasized that the present study is limited to the NC Teacher Assessments, which are the teachers' perceptions of mathematics performance. Although it would have been desirable to include objective tests as well as these year-long teacher assessments, we were unable to do so for this large

sample. However, objective test data will be available on this sample at 10 years. There is evidence for the validity of teacher assessments. In a meta-analysis of 16 studies comparing teacher assessments and standardized test results, a median correlation of .66 was found despite great variations in the methods used for teacher assessments (Hoge & Coladarci, 1989; see Oliver et al., 2004 for further support of the use of teacher assessments). Moreover, we have shown that teacher assessments of reading performance in this sample correlate highly (.68) with a telephone-administered test of word and nonword reading (Dale et al., 2005) and that a general factor of NC Teacher Assessments of academic achievement (including mathematics) correlates highly (.58) with a general factor of telephone-administered tests of verbal and nonverbal cognitive abilities (Spinath et al., in press). We expect a similar pattern to emerge between teacher-reported mathematics and tests of mathematics, but this needs to be directly assessed. We aim to compare directly assessed mathematics with NC measures when the twins are 10 years old.

The use of teacher assessments of mathematics performance as indicated by the NC Teacher Assessments used here (hereafter referred to as Teacher Assessments [TA]) rather than test scores is a strength as well as a limitation of our study, since there is some evidence to support the hypothesis that teacher assessments are likely to add to achievement tests in predicting long-term outcomes. For example, after controlling for socioeconomic status, preschool teachers' overestimates and underestimates of intelligence relative to IQ scores at 4 years significantly predicted high school grades and Scholastic Aptitude Test results 14 years later (Alvidrez & Weinstein, 1999). A similar study of teacher assessments of underachieving students predicted long-term educational attainment and career outcomes (McCall et al., 1992).

TA scores on the three measures were standardized to a mean of zero and a standard deviation of 1 on the basis of the entire sample of twins (with children with major perinatal and medical problems excluded as described earlier), and provided the basis for our analysis. All of the scales are normally distributed with skewness well under 1. The three scales are highly correlated — their average intercorrelation is .86 and a factor analysis of the three scales indicated that the principal component accounted for 90.6% of the variance. Thus we also computed a composite mathematics score by summing scores for the three scales and restandardizing the composite score. All measures were residualized for age and sex effects using a regression procedure. Standardized residuals were used because the age and sex of twins is perfectly correlated across pairs, and variation within age at the time of testing and variation within sex might contribute to the correlation between twins, and thus be misrepresented as environmental influences shared by the twins (Eaves et al., 1989; McGue & Bouchard, Jr., 1984).

Analyses

Individual Differences Analysis

In the current study, we performed model-fitting analyses for each of the three mathematics scores — using and applying; numbers and algebra; and shapes, space and measures — as well as for the composite score. Mx software for structural equation modeling was used to perform standard model-fitting analyses using raw data (Neale et al., 1999). Two fit indices are reported: Chi-square (χ^2) and Akaike's information criterion, AIC (Akaike, 1987). The best-fitting model was chosen on the basis of a change in χ^2 not representing a significant worsening of fit.

Prospective longitudinal analysis for the composite mathematics score was performed using a bivariate Cholesky decomposition model. A Cholesky decomposition model was fit to raw data to test for common and independent genetic and environmental effects on variance and covariance in mathematics performance at 7 and 9 years. Figure 1 shows the design of the Cholesky decomposition model for this analysis. Estimates from the Cholesky model can be transformed to obtain genetic, shared environmental and nonshared environmental correlations between mathematics performance at 7 and 9 years. The genetic correlation estimates the extent to which the same genetic effects operate at the two ages. Bivariate heritability and environmentality were also estimated. Bivariate heritability is an estimate of the proportion of phenotypic covariance between 7 and 9 years that can be attributed to genetic covariance between the two ages.

Extremes Analysis

A major goal of this study was to examine the etiology of low mathematics performance. As indicated earlier, the twin pairs where at least one member of the pair scored below the 15th percentile for each mathematics measure were selected to be in each of the low mathematics groups. Probandwise concordance (number of probands in concordant pairs as a ratio of the total number of probands) was calculated, which indicates the risk that a co-twin of a proband also meets criteria for low performance. If genetic influences are indicated, MZ concordance will exceed DZ concordance.

DF extremes analysis (DeFries & Fulker, 1985) was used to test a model of low performance that incorporates continuous variation. DF extremes analysis uses the mean trait score difference between probands (lowest 15%), the probands' co-twins, and the population for MZ and DZ twins to estimate genetic and environmental sources of the mean difference between probands and the population. Scores are standardized and transformed to adjust for proband mean differences between MZ and DZ groups so that genetic and environmental parameters can be estimated from structural equation model fitting based on the regression:

$$C(M) = B_1 P(M) + B_2 R + A$$

Claire M. A. Haworth, Yulia Kovas, Stephen A. Petrill, and Robert Plomin
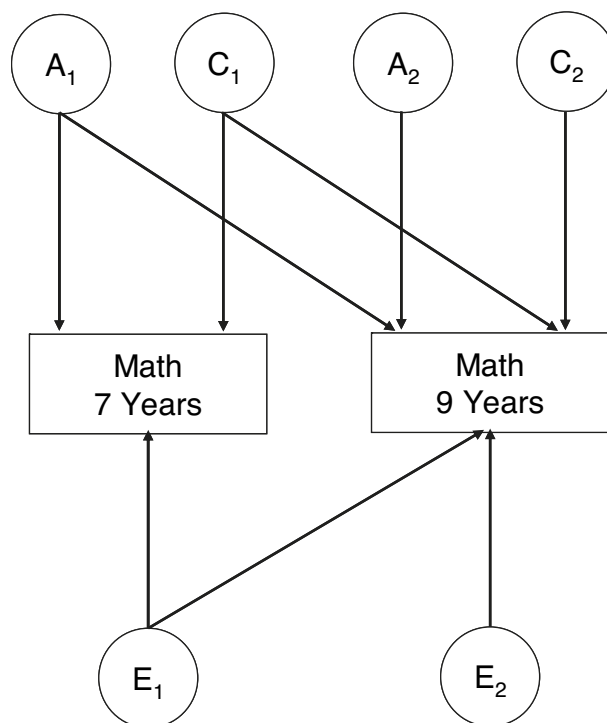
Figure 1

Bivariate Cholesky decomposition model.

Note: The two measured variables (rectangles) are mathematics performance at 7 years and mathematics performance at 9 years.

A = additive genetic influence; C = shared environmental influence; E = nonshared environmental influence.

where the co-twin's mathematics score, $C(M)$, is predicted from the proband's mathematics score, $P(M)$, and the coefficient of relatedness (R), which is 1.0 for MZ (genetically identical) and .5 for DZ twins (who are on average 50% similar genetically). The regression weight B_2 estimates group heritability because it tests whether proband and co-twin mean similarity varies as a function of the degree of genetic relatedness (R). This heritability is called *group* heritability as it refers to genetic influence on the mean difference between the proband group and the population, in contrast to the usual estimate of heritability, which could be called *individual differences* heritability, which refers to genetic influence on individual differences throughout the distribution.

For the longitudinal extremes analysis, probands were selected at 7 years and analyzed in comparison to their co-twin score at 9 years. Probands were also selected at 9 years and analyzed with their 7-year co-twin scores. Analysis in both directions is required to estimate a DF extremes genetic correlation, using the following formula:

$$r_{g_{(xy)}} = \sqrt{\frac{(B_2xy)(B_2yx)}{(B_2x)(B_2y)}}$$

where B_2xy is the group heritability from x to y (i.e., from 7 to 9 years) and B_2yx is the group heritability from y to x (i.e., from 9 to 7 years), B_2x is the group heritability at x (i.e., univariate group heritability at 7 years) and B_2y is the group heritability at y (i.e., univariate group heritability at 9 years; see Knopik et al., 1997 for further details).

Results

The 9-year means and standard deviations for the three mathematics scores and for the composite are presented in Table 1. Two-by-two (sex BY zygosity) ANOVA analyses on the mathematics measures yielded a significant main effect of sex on two of the four scores (using and applying: $p = .069$, $\eta^2 = 0.001$; numbers and algebra: $p = .001$, $\eta^2 = 0.002$; shapes, space and measures: $p = .408$, $\eta^2 < 0.001$; composite: $p = .04$, $\eta^2 = 0.001$), with boys performing significantly better than girls. However, the significance of these effects is attributable to the large sample size because the effect size was very small, accounting for less than 1% of the variance. Similarly, there were significant main effects of zygosity on all of the mathematics measures (using and applying: $p = .004$, $\eta^2 = 0.002$; numbers and algebra: $p = .012$, $\eta^2 = 0.001$; shapes, space and measures: $p = .002$, $\eta^2 = 0.002$; composite: $p = .003$, $\eta^2 = 0.002$), with DZ twins performing significantly better than MZ

Table 1

Means (and Standard Deviations) for 9-Year Teacher Assessments of Mathematics (Adjusted for Age), by Zygosity and Sex; and ANOVA Results Showing Significance and Effect Size, by Sex and Zygosity

Measure	Zygosity			Sex		ANOVA		
	All (n = 5339–5377)	MZ (n = 1927–1940)	DZ (n = 1754–1773)	Female (n = 2804–2826)	Male (n = 2535–2551)	Zygosity	Sex	Zygosity*Sex
Using and applying	0.00 (0.99)	−0.06 (0.99)	0.01 (1.00)	−0.03 (0.96)	0.04 (1.05)	$p = .004$ $\eta^2 = 0.002$	$p = .069$ $\eta^2 = 0.001$	$p = .123$ $\eta^2 < 0.001$
Numbers and algebra	0.00 (0.99)	−0.05 (0.99)	0.01 (1.00)	−0.05 (0.96)	0.06 (1.04)	$p = .012$ $\eta^2 = 0.001$	$p = .001$ $\eta^2 = 0.002$	$p = .201$ $\eta^2 < 0.001$
Shapes, spaces and measures	0.00 (0.99)	−0.06 (0.99)	0.04 (1.01)	−0.02 (0.96)	0.02 (1.04)	$p = .002$ $\eta^2 = 0.002$	$p = .408$ $\eta^2 < 0.001$	$p = .210$ $\eta^2 < 0.001$
Composite	0.00 (0.99)	−0.06 (0.99)	0.02 (1.00)	−0.04 (0.95)	0.04 (1.05)	$p = .003$ $\eta^2 = 0.002$	$p = .040$ $\eta^2 = 0.001$	$p = .154$ $\eta^2 < 0.001$

Note: Each score was adjusted for age and standardized on the basis of the whole sample before exclusions specific to this study (but after medical exclusions). For this reason, the standardized means shown all slightly exceed 0.00.

MZ = monozygotic; DZ = dizygotic; η^2 = eta squared (effect size).

twins. Again, the effect sizes of these significant effects were very small, accounting for less than 1% of the variance. Whether the twins had the same or different teacher did not significantly affect their mathematics scores. All analyses were repeated separating the sample into twins assessed by the same or different teacher; however, the findings remained unchanged and there was minimal difference between the two subsamples. Therefore, these results are not presented here.

Genetic Analysis of Individual Differences for the Entire Sample

The intraclass twin correlations for the four scores are shown in Table 2. For the entire sample, doubling the difference between these correlations indicates that genetics substantially influences all mathematics scores: .66 for using and applying; .64 for numbers and algebra; .62 for shapes, space and measures; and .68 for the composite score. Estimates of the shared environment are consistently modest (average .09).

Univariate model-fitting analyses were carried out for each of the three mathematics measures and the composite. The results of these univariate analyses for the entire sample are shown in Table 3.

For the full ACE models, shared environmental (C) estimates range from .08 to .11, and heritabilities range from .62 to .68. For all measures, the parameter estimates from the model fitting are highly similar to the estimates made from the intraclass correlations (Table 2). Furthermore, in line with our estimates from the correlations, the best-fitting and most parsimonious model for every measure is the AE model. For this best-fitting AE model, heritability estimates are greater than for the full ACE model because the A in the AE model tends to subsume the small amount of variance due to shared environment in the full ACE model. The remainder of the variance is attributed to nonshared environment plus error of measurement.

Genetic Analysis of Low Mathematics Performance

Probandwise concordances are shown in Table 4. As with the intraclass correlations, these concordance rates suggest genetic influence on the risk of being a proband because, in every case, concordance rates are substantially higher for MZ twins than those for DZ twins. Average MZ and DZ concordances across the four scores are .69 and .38, respectively, suggesting substantial genetic influence.

Table 4 also presents the results from the DF extremes analysis, which gives estimates of group heritability and group environmental influences. When calculating these results, group heritability cannot exceed MZ group correlation, in such cases group heritability was constrained. The results of the DF extremes analyses are highly similar to the results of the individual differences analyses for the entire

Claire M. A. Haworth, Yulia Kovas, Stephen A. Petrill, and Robert Plomin

Table 2

Intraclass Correlations and Estimated A, C and E Parameters for Three Mathematics Measures and the Composite for Twins by Zygosity at 9 Years

Measure	MZ (*n* = 899–909)	DZ (*n* = 825–844)	A	C	E
Using and applying	.73**	.41**	.64	.09	.27
Numbers and algebra	.71**	.40**	.62	.09	.29
Shapes, spaces and measures	.73**	.42**	.62	.11	.27
Composite	.77**	.43**	.68	.09	.23

Note: MZ = monozygotic; DZ = dizygotic; A = additive genetic influence; C = shared environmental influence; E = nonshared environmental influence.

**p* < .01.

Table 3

Individual Differences Univariate Model Fitting for all Measures for the Entire Sample: Model Fit and Parameter Estimates

Measure	Model	$\Delta\chi^2$	Δdf	*p*	AIC	A	C	E
Using and applying	ACE	1.296	3	.73	−4.704	0.65 (0.54–0.75)	0.08 (0.00–0.19)	0.26 (0.24–0.29)
	CE	144.309	1	.00	142.309		0.58 (0.55–0.61)	0.42 (0.39–0.45)
	AE	2.133	1	.14	0.133	0.74 (0.71–0.76)		0.26 (0.24–0.29)
Numbers and algebra	ACE	0.435	3	.93	−5.565	0.62 (0.51–0.73)	0.09 (0.00–0.20)	0.29 (0.26–0.32)
	CE	123.009	1	.00	121.009		0.57 (0.53–0.60)	0.43 (0.40–0.47)
	AE	2.548	1	.11	0.548	0.72 (0.69–0.74)		0.28 (0.26–0.31)
Shapes, spaces and measures	ACE	1.526	3	.68	−4.474	0.63 (0.51–0.74)	0.11 (0.00–0.21)	0.27 (0.24–0.30)
	CE	135.294	1	.00	133.294		0.58 (0.55–0.61)	0.42 (0.39–0.45)
	AE	3.539	1	.06	1.539	0.74 (0.71–0.76)		0.26 (0.24–0.29)
Composite	ACE	1.26	3	.74	−4.74	0.68 (0.57–0.78)	0.09 (0.00–0.20)	0.23 (0.21–0.26)
	CE	180.508	1	.00	178.508		0.61 (0.58–0.64)	0.39 (0.36–0.42)
	AE	2.752	1	.10	0.752	0.77 (0.75–0.79)		0.23 (0.21–0.25)

Note: Two fit indices are reported: Δchi-square (χ^2) and Akaike's information criterion (AIC; Akaike, 1987). CE and AE models are nested within the ACE model; the ACE model is nested in the fully saturated model. The best fitting model (in boldface) was chosen on the basis of a change in χ^2 not representing a significant worsening of fit (for a change of *df* of 1, the statistically significant change in χ^2 is 3.84).

Δdf = change in degrees of freedom; A = additive genetic influence; C = shared environmental influence; E = nonshared environmental influence.

95% confidence intervals in parentheses.

Table 4

Low Mathematics Performance: MZ and DZ Probandwise Concordances and Results of DF Extremes Analysis (Twin Group Correlations and h^2g and c^2g Parameter Estimates) Using a 15% Cutoff

|	Probandwise concordance		Twin group correlation		DF estimates			
	MZ	DZ	MZ	DZ	h^2g	SE	c^2g	SE
Using and applying	.72	.46	.74	.39	.70	.11	.04	.10
Numbers and algebra	.65	.36	.69	.32	.69	.11	.00	.09
Shapes, spaces and measures	.72	.33	.74	.29	.74	.11	.00	.09
Composite	.68	.36	.75	.33	.75	.11	.00	.10

Note: MZ = monozygotic; DZ = dizygotic; h^2g = group heritability; c^2g = group shared environment.

sample (Tables 2 and 3). For example, for the mathematics composite, MZ and DZ twin group correlations (Table 4) are .75 and .33, respectively. The group heritability estimate for the mathematics composite (Table 4) is .75 and the influence of shared environment is .00. Results for the three mathematics scales are similar to those for the composite measure, yielding substantial estimates of group heritability (.69 to .75 in Table 4) and modest estimates of group shared environment (.00 to .04).

Longitudinal Analysis of Individual Differences for the Entire Sample

Longitudinal data were analyzed for the composite mathematics scores at 7 and 9 years. The phenotypic correlation between the mathematics composites at 7 and 9 years is .60. Table 5 presents the results from
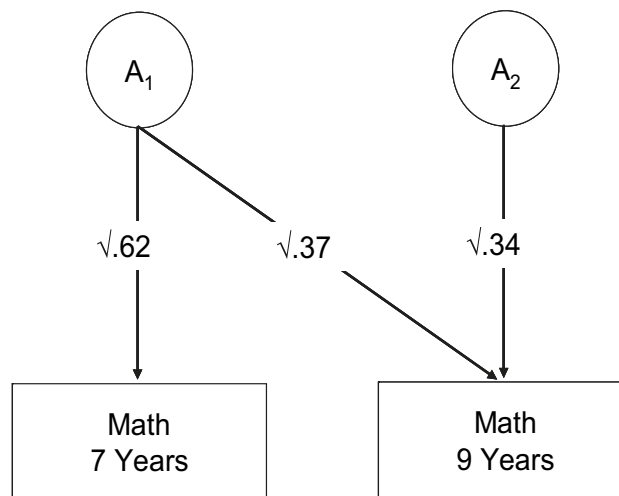
Figure 2

Genetic results from bivariate Cholesky decomposition. Results for additive genetic effects that are common and independent for mathematics performance (composite score) at 7 and 9 years-of-age.

Note: Model fit of Cholesky Decomposition model: –2 times log-likelihood of data = 27,309.351 (df = 10996); AIC = 5317.351.

the longitudinal Cholesky model, which decomposes the variance and covariance of the mathematics composite scores at 7 and 9 years into common and independent additive genetic, shared environmental and nonshared environmental influences. Figure 2 displays the Cholesky results for the genetic influences that are common and independent between the mathematics composite at 7 and 9.

The genetic contribution to the phenotypic correlation can be estimated from this longitudinal model as the product of the paths to the latent variable A_1, which represents genetic influences in common across the two ages. Thus, the genetic contribution to the phenotypic correlation of .60 is .48 (i.e., $\sqrt{.62} * \sqrt{.37}$). In other words, 80% of the phenotypic correlation is mediated genetically (i.e., .48 / .60 = .80), which is bivariate (longitudinal) heritability. As shown in Table 6, the model-fitting estimate of bivariate heritability is .81 (.67–.95 95% confidence interval [CI]). Although most of the phenotypic continuity between 7 and 9 years is mediated genetically, there is genetic variance at each age that is unique to that age. For example, Figure 2 shows that the model-fitting estimate of heritability at age 7 is .62; the path estimate of .37 from A_1 to 9 years indicates that roughly half of this genetic influence at 7 years also affects scores at 9 years. Similarly, heritability at 9 years is estimated as .71, the sum of the two paths to 9 years (.37 + .34). Again, about half of the genetic influence at 9 years is in common with 7 years and about half is independent.

As noted earlier, this longitudinal genetic analysis also provides an estimate of the genetic correlation from 7 to 9 years. The genetic correlation is independent of heritabilities at each age: The genetic correlation can be low even if the heritabilities are high and vice versa. It indicates the extent to which the same

genes are operating at each age regardless of the magnitude of their effect on the phenotype. As shown in Table 6, the genetic correlation is estimated as .72 (.64–.82 CI), indicating substantial genetic overlap between mathematics performance at 7 and 9 years. The genetic correlation can be gleaned from Figure 2: The genetic contribution to the phenotypic correlation (.48) mentioned above is the product of the square roots of the two heritabilities and the genetic correlation (Plomin & DeFries, 1979). Because we know the heritabilities at 7 years (.62) and 9 years (.71), we can solve for the genetic correlation, $.48/(\sqrt{.62} * \sqrt{.71}) = .72$.

Of the phenotypic correlation of .60 from 7 to 9 not explained by genetics, .13 (22%) can be attributed to shared environment and .06 (10%) to nonshared environment. Although shared environment does not account for much variance at either age, what shared

Table 5

Standardized Cholesky Squared Path Estimates (95% Confidence Intervals) for Mathematics at 7 and Mathematics at 9 Indicating Proportions of Genetic (A), Shared Environmental (C) and Nonshared Environmental (E) Influences on Each Trait That Are Shared With or Independent of Other Traits

Path	Mathematics at 7 Years	Mathematics at 9 Years
A_1	.62 (.55–.70)	.37 (.27–.50)
C_1	.16 (.09–.22)	.04 (.00–.13)
E_1	.22 (.20–.23)	.01 (.00–.01)
A_2		.34 (.23–.40)
C_2		.01 (.00–.09)
E_2		.23 (.23–.26)

Note: The two-factor model allows for (1) A, C, and E effects on Mathematics at 7, some of which also influence Mathematics at 9 (A_1, C_1, E_1); (2) A, C, E effects on Mathematics at 9 that are unique to Mathematics at 9 (A_2, C_2, E_2).

Claire M. A. Haworth, Yulia Kovas, Stephen A. Petrill, and Robert Plomin

Table 6

Genetic, Shared Environment, and Nonshared Environment Correlations for Mathematics Composite at 7 years and Mathematics Composite at 9 years; and Proportion of Phenotypic Correlation Between These Variables Mediated by A, C, and E

	Mathematics at 7 and Mathematics at 9
Correlation	
r_A	.72 (.64–.82)
r_C	.87 (–1.0–1.0)
r_E	.16 (.08–.23)
Mediation of r_P	
A ($a_x a_y r_A / r_P$)	.81 (.67–.95)
C ($c_x c_y r_C / r_P$)	.13 (.00–.26)
E ($e_x e_y r_E / r_P$)	.06 (.03–.09)

Note: r_A, r_C, r_E = genetic, shared environment, nonshared environment correlations; r_P = phenotypic correlation.

95% confidence intervals in parentheses.

environment exists is largely in common between 7 and 9 years. In contrast, nonshared environmental influences are largely independent at the two ages.

Longitudinal Analysis of Low Mathematics Performance

DF extremes longitudinal analyses based on the lowest 15% of each sample also indicated substantial genetic continuity from 7 to 9 years for low mathematics performance. Prospective group heritability, with probands defined at 7 years of age and compared to co-twins' 9-year scores, was .81 (SE = .14). Retrospective group heritability, with probands selected at 9 years and compared to co-twins' 7-year scores, was .66 (SE = .15). From these two longitudinal group heritabilities it is possible to calculate a genetic correlation between low mathematics performance at 7 and 9 years (Knopik et al., 1997). The genetic correlation was high and in fact exceeded 1; although confidence intervals for this group genetic correlation have not been worked out, they would obviously be very large. As found in the individual differences longitudinal analyses, nonshared environmental influences largely contributed to change in low mathematics performance from 7 to 9 years. Group shared environmental influences were not significant in this longitudinal analysis.

Discussion

At 7 years, we were surprised to find so much genetic influence and so little shared environmental influence for both low mathematics performance and normal variation in the early school years (Oliver et al., 2004). Given the major changes in mathematics development from 7 to 9 years, we expected that longitudinal analyses from 7 to 9 years would show that different genetic and environmental effects contributed to mathematics performance at the two ages. To the contrary, we found results at 9 years that were

highly similar to those at 7 years. For the entire sample, heritability was .68 at 9 years for the mathematics composite as compared to .66 at 7 years; shared environmental influence was .09 at both ages. For the low mathematics group, group heritability was .75 at 9 years (.65 at 7 years); shared environmental influence was .00 at 9 (.09 at 7). Our longitudinal analyses from 7 to 9 years showed substantial genetic continuity. For the entire sample, 80% of the phenotypic correlation of .60 from 7 to 9 years for the mathematics composite was mediated genetically. Moreover, the genetic correlation from 7 to 9 years was .72, which suggests that the same genes largely affect mathematics performance at 7 and 9 years. Similar results emerged from longitudinal analyses for low mathematics performance.

The only way in which our expectation of change from 7 to 9 years was met involved finding new genetic influence at 9 years. That is, despite the genetic correlation of .72 between 7 and 9 years, significant genetic variance appeared at 9 years that was independent of genetic effects at 7 years. This apparent paradox — a high genetic correlation from 7 to 9 but independent genetic variance at 7 and 9 years — can be explained by the high heritabilities at 7 and 9 years and the moderate phenotypic correlation of .60 from 7 to 9 years: Although genetic effects largely account for the phenotypic correlation of .60, genetic effects remain that are unique to each age.

Results for the three mathematics scales (using and applying mathematics; numbers and algebra; shapes, space and measures) were similar to those for the composite measure; this is unsurprising considering their high intercorrelations. For the entire sample, heritability estimates for the three scales were .65, .62, and .63, respectively; shared environmental estimates were .08, .09, and .11. For the low mathematics group, group heritability estimates were .70, .69 and .74; group shared environmental estimates were .04, .00, and .00. Together with finding such high phenotypic correlations among the three scales (.86 on average), this result suggests that the three scales are substantially linked genetically. However, a multivariate genetic analysis is needed to demonstrate this conclusively. The first multivariate genetic analysis of diverse aspects of mathematics found that they were highly correlated genetically (Kovas et al., in press).

The similarity of our estimates for individual differences heritability of mathematics performance (.68) and group heritability of low mathematics performance (.75) suggests that it is likely that the same genetic factors are largely responsible for low performance and for performance throughout the distribution. Moreover, DF group heritability itself suggests genetic links between the normal and abnormal because group heritability requires a genetic link between the proband group selected on the basis of low mathematics scores and quantitative scores of the co-twins (see Plomin & Kovas, 2005). Finding genetic

overlap between low performance and normal variation supports the quantitative trait locus (QTL) model of molecular genetics (Plomin et al., 1994). The basis of the QTL model is that many genes of small effect are responsible for the heritability of common disorders and continuous (quantitative) traits and this implies that common disorders are the quantitative extreme of the same genetic effects that operate throughout the distribution. That is, when genes are found that contribute to the heritability of low mathematics performance, the same genes would be expected to contribute to the heritability of normal variation in mathematics performance. Of course there are rare monogenetic disorders that result in low mathematics performance, for example, individuals with Prader-Willi syndrome generally present with, among other problems, poor mathematical abilities (Bertella et al., 2005). Such individuals have been excluded from this analysis, and therefore the genetic effects found here are not likely to be the result of such severe and rare monogenic disorders.

Limitations

A specific limitation of this study is its use of teacher assessments. Teachers rated the children's mathematics performance in three areas over an entire year, based on NC criteria, which teachers in schools in England and Wales must follow. Teachers are given strict guidelines, even to the level of lesson planning, and a timetable of expectations throughout the school year. The NC was gradually introduced throughout the early 1990s, and teacher training is focused on following the NC standards. Therefore, all teachers are very familiar with the guidelines and rating criteria. The use of teacher-report data has been supported in previous research (Oliver et al., 2004; Walker et al., 2004). Of particular importance for the low mathematics performance is that if a teacher deems that the child's performance is not adequate, the child will not take tests but will have their entire performance assessed by the teacher. Despite the practical importance of teacher assessments, a multi-method approach that incorporates tests would help to triangulate on the developmental origins of mathematics performance. As part of the TEDS project, a substudy has recently been completed of normal variation in mathematics performance at 10 years using online web-based tests to assess diverse aspects of mathematics similar to those assessed by teachers in the present study (Kovas et al., in press). Similarly, low estimates of shared environmental influence were found (.15 on average) but heritability estimates were lower than for teacher assessments (.40 on average). If confirmed, these results leave open the possibility that teachers are assessing somewhat different aspects of mathematics performance than are tests. What is needed is a multivariate genetic analysis of the links between mathematics assessments by teachers and by tests.

A general limitation is the use of the twin method, which makes several assumptions, such as the equal environments assumption. The validity of twin data is discussed in detail elsewhere (Boomsma et al., 2002; Martin et al., 1997; Plomin, DeFries, et al., 2001). Confirmation by the adoption method would be useful because the adoption method makes different assumptions (Plomin, DeFries, et al., 2001). Ultimately, the most convincing evidence will come from molecular genetic studies that identify the genes responsible for heritability and group heritability.

Implications

Although our quantitative genetic results require molecular genetic studies for ultimate confirmation, at the same time the results help to chart the course for molecular genetic research in three ways. First, they suggest that mathematics performance at 7 and 9 years as assessed by teachers is a good target for molecular genetic research as they are highly heritable. Second, the finding that, despite the high genetic correlation between 7 and 9 years, substantial unique genetic variation exists at each age recommends conducting molecular genetic research on longitudinal phenotypes, for example, combining scores at 7 and 9 years. Third, our results supporting the QTL model suggest that molecular genetic research would benefit from considering mathematics performance as a continuous trait rather than focusing on low mathematics performance. Additionally, our results suggest but do not prove that most of the genetic action is general across diverse mathematics processes, which would recommend that molecular genetic research would profit by focusing on general rather than specific mathematics processes; however, a multivariate genetic analysis of mathematics processes is needed to prove this. In addition, we will need to conduct similar research later into adolescence when children begin to study more sophisticated mathematics, such as algebra and geometry, to determine whether these types of mathematics are also genetically linked.

These results also have general implications for education as we believe that this research will prove fundamental in terms of leading to both genetic and environmental risk indicators that can promote early intervention studies for those children who might otherwise struggle. This study has confirmed our earlier finding at 7 years that individual differences in mathematical performance across the spectrum are substantially influenced by genetic factors. However, this in no way implies that the teaching environment is not important. Mathematics is a taught skill; it is not a purely innate ability and can only be learned if the child is exposed to it. For instance, there are individual differences in how well children perform in mathematics, and we find evidence for the importance of environmental factors in these differences: For the composite mathematics score, a third (32%) of the variance could be attributed to environmental factors;

9% to shared environment and 23% to nonshared environment. What is particularly interesting here is that it is the nonshared environmental influence that has the greater effect, even though most of these children are in the same schools. There are also no significant differences between twins in the same versus different classrooms. The classroom environment in England and Wales is likely to be similar due to the highly structured NC; this suggests that children are experiencing the same environment differently, and gives support to the theory that environmental influences operate on an individual-by-individual basis and not generally on a family-by-family basis (Plomin, Asbury, et al., 2001).

It may be that education should follow the trend towards individualization, by adopting specific learning plans for each child. Those children with special educational needs already have a certain level of personalized teaching plans (Department for Education and Skills, 2002). Further research into the environmental factors that are most relevant and their correlation or interaction with genetic effects could enlighten the options available for individualized learning for all children.

Acknowledgments

References

Akaike, H. (1987). Factor analysis and AIC. *Psychometrika, 52,* 317–332.

Alarcón, M., DeFries, J. C., Light, J. G., & Pennington, B. F. (1997). A twin study of mathematics disability. *Journal of Learning Disabilities, 30,* 617–623.

Alarcón, M., Knopik, V. S., & DeFries, J. C. (2000). Covariation of mathematics achievement and general cognitive ability in twins. *Journal of School Psychology, 38,* 63–77.

Alvidrez, J., & Weinstein, R. S. (1999). Early teacher perceptions and later student academic achievement. *Journal of Educational Psychology, 91,* 731–746.

Bertella, L., Girelli, L., Grugni, G., Marchi, S., Molinari, E., & Semenza, C. (2005). Mathematical skills in Prader-Willi Syndrome. *Journal of Intellectual Disability Research, 49,* 159–169.

Boomsma, D., Busjahn, A., & Peltonen, L. (2002). Classical twin studies and beyond. *Nature Reviews Genetics, 3,* 872–882.

Dale, P., Harlaar, N., & Plomin, R. (2005) Telephone testing and teacher assessment of reading skills in 7-year-olds: I. Substantial correspondence for a sample of 5808 children and for extremes. *Reading and Writing: An Interdisciplinary Journal, 18,* 385–400.

DeFries, J. C., & Fulker, D. W. (1985). Multiple regression analysis of twin data. *Behavior Genetics, 15,* 467–473.

Department for Education and Skills. (2002). *Special educational needs in England.* January, 2002 Bulletin.

DfEE Publications. (1999). *The National Numeracy Strategy.* Cambridge: Cambridge University Press.

Eaves, L. J., Eysenck, H., & Martin, N. G. (1989). *Genes, culture, and personality: An empirical approach.* London: Academic Press.

Freeman, B., Smith, N., Curtis, C., Huckett, L., Mill, J., & Craig, I. W. (2003). DNA from buccal swabs recruited by mail: Evaluation of storage effects on long-term stability and suitability for multiplex polymerase chain reaction genotyping. *Behavior Genetics, 33,* 67–72.

Hoge, R. D., & Coladarci, T. (1989). Teacher-based judgments of academic achievement: A review of literature. *Review of Educational Research, 59,* 297–313.

Knopik, V. S., Alarcón, M., & DeFries, J. C. (1997). Comorbidity of mathematics and reading deficits: Evidence for a genetic etiology. *Behavior Genetics, 27,* 447–453.

Kovas, Y., Petrill, S. A., & Plomin, R. (in press). The origins of diverse domains of mathematics: Generalist genes but specialist environments. *Journal of Educational Psychology.*

Martin, N., Boomsma, D. I., & Machin, G. (1997). A twin-pronged attack on complex trait. *Nature Genetics, 17,* 387–392.

McCall, R. B., Evahn, C., & Kratzer, L. (1992). *High school underachievers: What do they achieve as adults?* Pittsburgh, PA: Sage.

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

Neale, M. C., Boker, S. M., Xie, G., & Maes, H. (1999). *Mx: Statistical modeling* (5th ed.). Richmond, VA: Department of Psychiatry.

Oliver, B., Harlaar, N., Hayiou-Thomas, M. E., Kovas, Y., Walker, S. O., Petrill, S. A., Spinath, F. M., Dale, P. S., & Plomin, R. (2004). A twin study of teacher-reported mathematics performance and low performance in 7-year-olds. *Journal of Educational Psychology, 96,* 504–517.

Oliver, B. R., & Plomin, R. (2007). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through

adolescence. *Twin Research and Human Genetics, 10,* 96–105.

Plomin, R., Asbury, K., & Dunn, J. (2001). Why are children in the same family so different? Nonshared environment a decade later. *Canadian Journal of Psychiatry, 46,* 225–233.

Plomin, R., & DeFries, J. C. (1979). Multivariate behavioral genetic analysis of twin data on scholastic abilities. *Behavior Genetics, 9,* 505–517.

Plomin, R., DeFries, J. C., McClearn, G. E., & McGuffin, P. (2001). *Behavioral genetics* (4th ed.). New York: Worth Publishers.

Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, nurture and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science, 8,* 442–447.

Plomin, R., & Kovas, Y. (2005). Generalist genes and learning disabilities. *Psychological Bulletin, 131,* 592–617.

Plomin, R., Owen, M. J., & McGuffin, P. (1994). The genetic basis of complex human behaviors. *Science, 264,* 1733–1739.

Price, T. S., Freeman, B., Craig, I. W., Petrill, S. A., Ebersole, L., & Plomin, R. (2000). Infant zygosity can be assigned by parental report questionnaire data. *Twin Research, 3,* 129–133.

Spinath, F. M., Ronald, A., Harlaar, N., Price, T. S., & Plomin, R. (2003). Phenotypic 'g' early in life: On the etiology of general cognitive ability in a large population sample of twin children aged 2 to 4 years. *Intelligence, 31,* 195–210.

Spinath, F. M., Walker, S. O., Saudino, K. J., & Plomin, R. (in press). To what extent is genetic influence on teacher-assessed academic achievement due to genetic influence on test-assessed general cognitive ability? A study of 1812 pairs of 7-year-old twins. *Intelligence.*

Thompson, L. A., Detterman, D. K., & Plomin, R. (1991). Associations between cognitive abilities and scholastic achievement: Genetic overlap but environmental differences. *Psychological Science, 2,* 158–165.

Trouton, A., Spinath, F. M., & Plomin, R. (2002). Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behaviour problems in childhood. *Twin Research, 5,* 444–448.

Walker, S. O., Petrill, S. A., Spinath, F. M., & Plomin, R. (2004). Nature, nurture and academic achievement: A twin study of teacher ratings of 7-year-olds. *British Journal of Educational Psychology, 74,* 323–342.