British Journal of Educational Psychology (2004), 74, 323–342

www.bps.org.uk

Nature, nurture and academic achievement: A twin study of teacher assessments of 7-year-olds

Sheila O. Walker[1]*, Stephen A. Petrill[2], Frank M. Spinath[3] and Robert Plomin[1]

[1] Social, Genetic and Developmental Psychiatry Research Centre, Institute of Psychiatry, King's College, London, UK
[2] Pennsylvania State University, USA
[3] University of Bielefeld, Germany

Background. Twin research has consistently shown substantial genetic influence on individual differences in cognitive ability; however, much less is known about the genetic and environmental aetiologies of school achievement.

Aims. Our goal is to test the hypotheses that teacher-assessed achievement in the early school years shows substantial genetic influence but only modest shared environmental influence when children are assessed by the same teachers and by different teachers.

Sample. 1,189 monozygotic (MZ) and dizygotic (DZ) twin pairs born in 1994 in England and Wales.

Methods. Teachers evaluated academic achievement for 7-year-olds in Mathematics and English. Results were based on the twin method, which compares the similarity between identical and fraternal twins.

Results. Suggested substantial genetic influence in that identical twins were almost twice as similar as fraternal twins when compared on teacher assessments for Mathematics, English and a total score.

Conclusions. The results confirm prior research suggesting that teacher assessments of academic achievement are substantially influenced by genetics. This finding holds even when twins are assessed independently by different teachers.

Although the early school years are a critical developmental phase for children, very little is known about the interplay between nature and nurture as it relates to teacher assessments of academic achievement. Despite a realisation in the educational literature

* *Correspondence should be addressed to Sheila Walker, Box PO83, Institute of Psychiatry, DeCrespigny Park, Denmark Hill, London SE5 8AF, UK (e-mail: sheila.walker@iop.kcl.ac.uk).*

about the importance of teacher assessments (Hoge & Coladarci, 1989), behavioural genetic analyses have traditionally focused on standardized tests of academic achievement. For example, research using the twin design has suggested genetic influence on standardized tests of reading (Fisher & DeFries, 2002) and other aspects of language (Stromswold, 2001).

Although standardized tests represent a valuable source of information about children's abilities and disabilities, teachers are in a unique position to observe important motivational and personality characteristics in the 'real world' (as opposed to the controlled environment of standardized achievement tests) as the child learns and interacts with peers in a novel environment. How do genes and the environment interact during this important time in development in the classroom as assessed by teachers, and does this relationship change when children are in different classrooms and rated by different teachers?

We explored this question using the twin method, in which the similarity of identical and fraternal twins was compared in order to estimate the relative contribution of genes and the environment on individual differences in teacher-rated academic achievement. As discussed in the Method section, the twin method is widely used in the life sciences as a rough index of genetic influence by comparing resemblance for identical twins (monozygotic, MZ) who are, pair by pair, genetically identical, and for non-identical twins (dizygotic, DZ) who are only 50% similar genetically (Boomsma, Andreas, & Peltonen, 2002; Bouchard & Propping, 1993; Martin, Boomsma, & Machlin, 1997; Plomin, De Fries, McClearn, & McGuffin, 2001a).

We are aware of only one twin study that has investigated teacher assessments of academic achievement. In a study of 352 identical (MZ) and 668 fraternal (DZ) 13-year-old twin pairs in Sweden, MZ and DZ correlations were .81 and .48, respectively, for teacher assessments of arithmetic, .76 and .50 for writing, .72 and .57 for reading, and .80 and .51 for history (Husén, 1959). These patterns of twin correlations suggest substantial heritabilities (genetic effect sizes) accounting for about half (between 30% and 66%) of the variance of teacher assessments when both members of the twin pair were assessed by the same teacher. These substantial heritability estimates exceed heritabilities found for tests of academic achievement, which are typically about 30% in the early school years and about 40% later in development (e.g., Loehlin & Nichols, 1976; Thompson, Detterman, & Plomin, 1991).

Behavioural genetic studies also provide important information about environmental influences. Home environment influences on individual differences in academic achievement have traditionally been explored using family structure variables such as socioeconomic status, parental education, and family size (Bacete & Remirez, 2001; McDermott, 1995; Reynolds & Lee, 1991), and family process variables such as parental expectations and family affective climate (Christenson, Rounds, & Gorney, 1993; Menaghan, Kowaleski-Jones, & Mott, 1997). Such environmental factors would be shared by two children growing up in the same family and are called shared environmental influences, as contrasted with nonshared environmental influences that do not contribute to resemblance between children in the same family (Plomin, Asbury, & Dunn, 2001b). Shared environmental influence is estimated in twin studies as the similarity between twins that cannot be explained by genetics (Plomin *et al.*, 2001a). Shared environment can also be found outside of the home, for instance if children are in the same classroom. Moreover, children can be in different classrooms but still share the effects of the school environment. The influence of school environment on academic achievement has been explored by examining school quality and classroom climate (Eccles, Wigfield, & Schiefele, 1998), teacher skill and achievement incentives (Rutter, 1981), relationship with peers (Teo, Carlson, Mathieu, Egeland, & Sroufe, 1996), and classroom size (Nye, Hedges, & Konstantopoulos, 2001). The effect of such factors could be shared or nonshared – studies of siblings are needed to ask whether these factors contribute to sibling differences or similarities. Studies of twins are needed to disentangle genetic influences from shared and nonshared environmental sources of sibling differences and similarities. The only previous twin study of teacher assessments of academic achievement, mentioned earlier, yielded an average shared environment estimate of 25% after genetic influence was taken into account (Husén, 1959).

Twin studies of tests of academic achievement support the assumption that shared environmental factors contribute importantly to individual differences. For example, a study of academic achievement test scores in twins from 6 to 12 years of age yielded estimates of about 60% for shared environment (Thompson *et al.*, 1991), although a study of twins in high school estimated shared environment as about 30% (Loehlin & Nichols, 1976). This decrease in shared environmental influence with increased age may be due to the increased differentiation in experience that takes place as children grow older and spend more time outside of the home and live their separate lives (Dunn & Plomin, 1990).

The goal of the present study was to investigate whether the results of the Swedish twin study of 13-year-olds – that teacher assessments of academic achievement in the early school years show substantial genetic influence but only modest shared environmental influence – are also found for 7-year-old British twins at the end of the school year. We extend the Swedish study by asking whether these relationships hold even when twins are in different classrooms and their academic achievement is assessed independently by different teachers.

Method

Participants

The sampling frame for our study was the Twins Early Development Study (TEDS), a longitudinal population-based study of twins born in England and Wales in 1994 (Dale *et al.*, 1998). Of 3,121 families who received consent forms in February through June 2001 when the twins were 7 and completing their first year in primary school, 2,039 (65%) agreed in writing or verbally to participate in the 7-year study. Of those, 2,003 (98%) agreed in writing or verbally to allow us to contact the twins' teachers via postal questionnaire. Of the 2,236 teacher questionnaires sent to twins with the same teacher, 1,874 (84%) responded. Of the 1,278 teacher questionnaires sent to twins with different teachers, 1,104 (86%) responded. This yielded a cumulative total of 3,514 teachers contacted, with 2,978 (85%) responding. Despite attrition, it has been demonstrated that the TEDS sample continues to be reasonably representative in terms of parental education and ethnicity (Spinath, Ronald, Harlaar, Price, & Plomin, 2003). For example, UK census data (Office for National Statistics, 2002) indicate that 32% of mothers in UK households have one or more A-levels and that 93% of children are white; the comparable statistics for the current sample are 37% and 94%.

Physical similarity ratings by parents were used to determine the zygosity of the twins. This method was more than 95% accurate when validated with a sample of same-sex pairs using DNA markers (Price, Freeman, Craig, Ebersole, & Plomin, 2000). Twins with complete data within the normal range (± 3*SD*) included 194 pairs of identical or monozygotic (MZ) males, 240 pairs of MZ females, 174 pairs of fraternal or dizygotic (DZ) males, 209 pairs of DZ females, and 372 pairs of DZ opposite-sex twins. Of the pairs 790 were rated by the same teacher and 399 were rated by different teachers.

Measures

Teachers' assessments of academic achievement were based on UK National Curriculum (NC) criteria for Key Stage 1, which are designed for children age 5 through their first year of primary school at age 7 (QCA Key Stage 1 Handbook, 1999). The NC is the core academic curriculum developed by the Qualifications and Curriculum Authority (QCA) and the National Foundation for Educational Research. For Key Stage 1, the QCA provides teachers with NC curriculum material and test guidelines for three academic categories within Mathematics (Using and applying mathematics; Numbers; Shapes, space and measures) and three categories within English (Speaking and listening; Reading; Writing). These six measures provided the basis for our analysis of teacher-assessed academic achievement.

Key stage 1 NC scores are comprised of teacher-assessed performance at the end of the key stage, when children are 7 years old. The Qualifications and Curriculum Authority (QCA) is the public body accountable to the UK Secretary of State for Education that develops the NC curriculum and Scholastic Aptitude Tests (SAT) for students, and the grading key used by teachers (QCA Key Stage 1 Handbook, 1999). For Key Stage 1, the grading key stipulates five levels of achievement for each academic subject area, each level encompassing a broad range of skills. The child's final NC rating is subject to interpretation, as the teacher determines which level provides the best fit for the child's abilities. For example, a student receives a score of '1' for Key Stage 1 Writing if the pupil's writing communicates meaning through simple words and phrases, the pupil begins to show awareness of how full stops are used, and letters are usually clearly shaped and correctly orientated; a rating of '2' indicates that a pupil's writing communicates meaning in both narrative and non-narrative forms, uses appropriate and interesting vocabulary, and shows some awareness of the reader; and a rating of '3' if pupil's writing is often organized, imaginative and clear, and the basic grammatical structure is usually correct (QCA English Tasks Teacher's Handbook, 2002). We did not ask teachers to differentiate Level 2 into 2a, 2b and 2c which they are asked to do for Key Stage 1. A student performing below level 1 receives a score of 'not achieved' (NA), and a student performing above level 3 receives a 4+. We used this 5-point scale (NA, 1, 2, 3, 4+) as our measure of academic achievement.

The judgment of the teacher ultimately determines the final SAT score that is submitted back to the QCA at the end of the school year. The SAT at the end of Key Stage 1 is different from other key stages, as it is the only one in which the teacher bears responsibility for marking the exam (with guidelines provided by the QCA). Key Stage 2, 3 and 4 results (administered when children are 11, 14, and 16, respectively) are comprised of teacher-assessed classroom performance combined with a cumulative objective, externally graded SAT administered at the end of the key stage. SAT developers assert that at Key Stage 1 their aim was to measure something broader than objective test performance, but evidence for validity is lacking. However, we have recently used the TEDS sample to show that Key Stage 1 teacher-assessed reading correlates .68 with a brief test of early word recognition (Test of Early Word Reading

Efficiency; Torgesen, Wagner, & Rashotte, 1999) that we administered via the telephone for 5,808 7-year-olds, thus supporting the validity of the teacher assessments (Dale, Harlaar, & Plomin, 2003).

The Key Stage 1 NC assessments reported by the children's teachers are the academic achievement measure used in the current analysis. These scores were standardized to a zero mean and unit variance. A univariate analysis of variance (ANOVA) was performed in order to investigate mean differences between males and females and between zygosity groups. Principal component factor analysis and Promax rotation was applied in order to investigate the relationship among the six academic subject scores.

Analyses

The twin method

The twin method makes use of the natural experiment provided by MZ and DZ twins (Plomin *et al.*, 2001a). MZ twins share all of their genes but DZ twins are only 50% similar genetically as are other sibling pairs. For this reason, if one posits that genetic differences affect a particular trait, it is necessary to predict that MZ twins will be more similar on that trait than will DZ twins. If MZ twins are no more similar than DZ twins, this provides strong evidence that genetic influence is not important. Moreover, the increased resemblance of MZ twin pairs relative to DZ pairs provides a rough estimate of half the genetic influence on the variance of the outcome measure and doubling the difference in MZ and DZ twin correlations provides a rough estimate of 'heritability', the proportion of the observed (phenotypic) variance that can be attributed to genetic variance. The remaining within-pair similarity is accounted for by the shared environment, defined as environmental influences that make twins similar. Remaining variance not due to genes or shared environment is referred to as non-shared environment, which also includes measurement error.

It should be noted that in quantitative genetics, the word 'genetic' is defined very narrowly whereas the word 'environment' is defined very broadly. That is, 'genetic' refers to heritable DNA differences transmitted from parent to offspring. All members of the human species are identical for 99.9% of DNA nucleotide base sequences. Only the 0.1% of DNA that differs can be responsible for the phenotypic differences among us. Many aspects of development are too crucial to allow DNA variation which is winnowed out during natural selection. For example, nearly all members of our species are bipedal, have frontal vision which allows depth perception and easily learn to use language. Such species universals are constrained by DNA that does not vary. Quantitative genetics' definition limiting 'genetic' to the 0.1% of DNA that varies is what is commonly meant when people refer to inheritance, for example, the inheritance of eye colour or height.

In contrast to the narrow definition of genetic, the word 'environment' in quantitative genetics simply means all other sources of variance. It is in this sense that quantitative genetics decomposes variance into genetic and 'environmental' (which might be better labelled as 'non-genetic') components of variance. This very broad definition of environment includes not only the social sorts of environments that psychologists tend to study such as parenting but also biological events (such as prenatal and postnatal illnesses) and even non-inherited DNA events (such as somatic mutations). Even with this mutually exclusive definition of genetics and environment in which everything not genetic (by the narrow definition) is called 'environment', there remains the problem of genotype-environment correlation in which the effects of genes are mediated via the environment. To the extent that such genotype-environment correlation increases the resemblance of MZ twins more than DZ twins, these effects will be included in estimates of heritability based on the twin design (Plomin *et al.*, 2001a).

The twin design is widely accepted in the life sciences as a useful screen for genetic influence. For example, in research on common medical disorders, the twin method has been used to show that cancers such as breast cancer typically show little genetic influence in that MZ concordances are about 15% and DZ concordances are about 10% (Lichtenstein *et al.*, 2000); heart problems show moderate genetic influence with MZ and DZ concordances of about 30% and 15% respectively; and epileptic seizures show substantial genetic influence with MZ and DZ concordances of about 80% and 40% (Plomin, Owen, & McGuffin, 1994). In 1924, the twin method was first used in the behavioural sciences in a study of cognitive ability tests (Merriman, 1924). Dozens of twin studies of general cognitive ability involving more than 10,000 twin pairs have consistently yielded MZ correlations of about .85 and DZ correlations of about .60, suggesting substantial genetic influence (Plomin, 1999, 2003).

Heritability, or the magnitude of *genetic influence*, refers to anonymous components of variance. Heritability does not specify which genes are responsible for genetic influence, just as the environmental component of variance does not specify which environmental factors are responsible for environmental influence. Nor does quantitative genetics specify the physiological or psychological mechanisms by which these genes or environmental factors come to have their effects. Nonetheless, quantitative genetic theory assumes that heritability indexes the extent to which DNA differences can account for phenotypic differences. One of the most energetic areas of science today involves attempts to apply the harvest from the Human Genome Project to begin to identify some of the specific DNA responsible for heritable traits.

We use the word '*influence*' rather than words such as '*cause*' or '*determine*' because complex traits involve many genes as well as many environmental factors. There are thousands of single-gene disorders in which a mutation in a single gene is the necessary and sufficient cause for the disorder. For example, phenylketonuria (PKU) is a form of mental retardation caused by recessive mutations in a single gene (phenylalanine hydroxylase, PAH). That is, the PKU form of mental retardation only occurs in individuals with a double dose of the mutated form of the PAH gene. In contrast, for complex traits, because many genes are involved, their effects are probabilistic rather than deterministic.

Although probabilistic, the correlation between DNA differences and behavioural differences indicate a causal direction of effects from DNA to behaviour because the nucleotide structure of DNA cannot be changed by the behaviour. The expression (transcription and translation) of the DNA code can be changed by behaviour and the environment, but the code itself is not changed. In contrast, other correlations between biology (such as the anatomy and physiology of the brain) and behaviour are not causal because behaviour can change biology as well as biology changing behaviour. To the extent that heritability indexes these DNA differences, it also refers to causal influences from DNA to behaviour. The phrase 'correlation does not imply causation' should be replaced by 'correlation does not *necessarily* imply causation.' The correlational statistic merely describes the magnitude of an association – it is the experimental design that provides the heft of causality. The twin method is a quasi-experimental design in the sense that there are two groups of twins, MZ twins who are genetically identical pair by pair and DZ twins who are only 50% similar genetically. Although twins are not randomly assigned to these two groups (which is why the method is called quasi-experimental), if one hypothesizes that genetic influence is important for a particular trait, it is necessary to predict that MZ twins will be more similar than DZ twins for that trait.

The strengths and weaknesses of the twin method are discussed in detail elsewhere (Boomsma *et al.*, 2002; Bouchard & Propping, 1993; Martin, Boomsma, & Machlin, 1997; Plomin *et al.*, 2001a). The main concern with the twin method is the so-called equal environments assumption – the assumption that environmental similarity for MZ and DZ twins reared in the same family is similar (see review in Plomin *et al.*, 2001a). Violations of this assumption could inflate estimates of genetic influence. However, the equal environments assumption is supported by several studies (e.g., Bouchard & Propping, 1993; Loehlin & Nichols, 1976; Morris-Yates, Andrews, Howie, & Henderson, 1990).

One possible violation of the equal environments assumption is the fact that two-thirds of MZ twin pairs share the same chorion, whereas DZ twins never share the same chorion. Sharing a chorion might make MZ twins more similar than DZ twins. However, this hypothesis would lead to the prediction that MZ twins who share the same chorion should be more similar behaviourally than MZ twins who do not share a chorion but the evidence in support of this hypothesis is weak (Gutknecht, Spitz, & Carlier, 1999; Phelps, Davis, & Schwartz, 1997; Riese, 1999; Sokol *et al.*, 1995). It seems more likely that MZ twins may experience *greater* prenatal environmental differences than DZ twins. For example, MZ twins show greater birth weight differences than DZ twins do. The difference may be due to greater prenatal competition, especially for the majority of MZ twins who share chorion.

A postnatal test of the equal environments assumption involves twins whose zygosity has been incorrectly assigned. MZ twins who have been mistakenly labelled as DZ twins by their parents are as similar behaviourally as correctly labelled MZ twins (Kendler, Neale, Kessler, Health, & Eaves 1993; Scarr & Carter-Saltzman, 1979). Another way in which the equal environments assumption has been tested takes advantage of the fact that differences within pairs of MZ twins can only be due to environmental influences. The equal environments assumption would be shown to be violated if MZ twin pairs who are treated more differently than others behave more differently, but the evidence here also supports the reasonableness of the twin method (e.g., Loehlin & Nichols, 1976; Morris-Yates *et al.*, 1990). Another important piece of evidence supporting the reasonableness of the twin method is that adoption studies – such as studies of birth parents and their adopted-away children and adoptive parents and their adopted children – yield similar results suggesting genetic influence in the cognitive domain (Plomin, Fulker, Corley, & DeFries, 1997), despite the considerable differences between the twin design and the adoption design (Plomin *et al.*, 2001a).

Another issue about the twin method in addition to the equal environments assumption is the generalizability of results from the twin method to the general population. Twins that are different are often born premature and intrauterine environments can be adverse when twins share a womb (Phillips, 1993). Newborn twins are about 30% lighter at birth than singletons, a difference that disappears by middle childhood (MacGillivray, Campbell, & Thompson, 1988). Secondly, language develops somewhat more slowly in twins (Rutter & Redshaw, 1991), although most of this language deficit is recovered by the early school years (Wilson, 1983). Twins do not appear to be importantly different from singletons for other domains.

A final issue, discussed later, is that, in the cognitive domain, twins appear to share environmental experiences to a greater extent than non-twin siblings perhaps because twins are the same age and thus share more experiences than do non-twin siblings who differ in age (Koeppen-Schomerus, Spinath, & Plomin, 2003).

Twin analyses

For twin analyses, correlations for the six academic subject scores were calculated for MZ and DZ twin pairs using standardized residual scores that adjust for sex differences. In addition to presenting MZ and DZ twin correlations as an indication of genetic and environmental influence, standard maximum-likelihood model-fitting was applied to the twin variance-covariance matrices using the structural-equation modelling package Mx (22) (Neale, 1997). Univariate models were fit to the observed data using gender-corrected scores described above. Figure 1 shows the basic twin model.

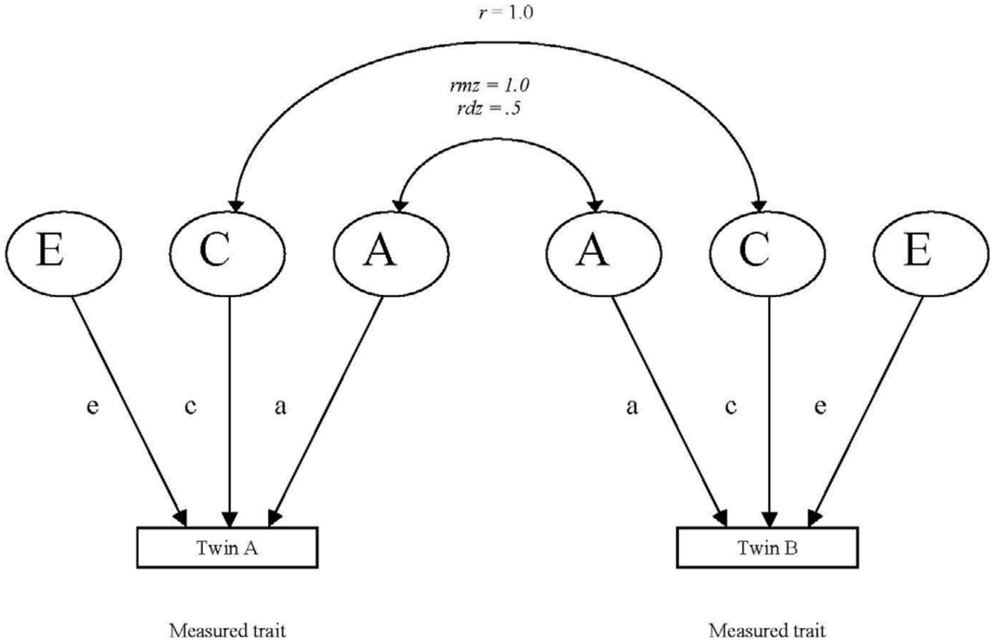

Figure 1. The Standard Twin Model. Twin A and Twin B = Twins in a Pair, A = Additive Genetic Influence, C = Shared Environment, E = Non-shared Environment, Paths a, c & e = Effects of A, C & E on a Trait, *rc* = Shared Environment Correlation, *rmz* = MZ Correlation, *rdz* = DZ Correlation

Resemblance in a measured trait for twins reared together can be due to additive genetic factors (A) or shared or common environment (C). The path coeffieients of latent variables A (genetic), C (shared environmental) and E (nonshared environmental, which includes error of measurement) factors are represented by the lower case letters a, c, and e, respectively. Genetic relatedness, or the genetic correlation (r_G) is 1.0 for MZ twins and 0.5 for DZ twins. Environmental relatedness, or the shared environmental correlation (r_C) is assumed to be 1.0 both for MZ and DZ twins. The latent E variable represents non-shared environmental influence, which contributes to differences between twins, and also contains measurement error. The full ACE model dissects the phenotypic variance into these three components of variance. (For details, see Plomin *et al.*, 2001a.)

In order to take possible sex differences into account in the ACE model, a sex-limitation model was used that is based on data from five zygosity groups: MZ male pairs, (MZM), MZ female pairs (MZF), DZ male pairs (DZM), DZ female pairs (DZF), and DZ opposite-sex pairs (DZO) (Galsworthy, Dionne, Dale, & Plomin, 2000). By comparing all five zygosity groups, A, C, and E can be estimated separately for males and females (quantitative differences). The model further tests whether the genetic correlation (r_G) or the shared environment correlation (r_C) for DZO is less than for DZ same-sex twins implicating the existence of factors that contribute to individual differences in one sex but not the other (qualitative differences). A series of 'nested' models can be tested and their fit can be compared by means of χ^2-difference tests. In the present study, four models were compared: (a) *a general or full sex-limitation model* allowing quantitative and qualitative differences between males and females and estimating either r_C or r_G, (b) a *common-effects sex-limitation model* allowing for quantitative differences between the sexes but fixing r_G or r_C to 1.0, (c) a *scalar effects sex-limitation model* which removes quantitative differences between males and females while still taking into account differences in variance, and (d) a *null model* which constrains all the parameters to be equal for males and females.

The above parameters and their 95% confidence intervals were estimated by fitting the models described to variance/covariance matrices using Mx (Neale, 1997). Three fit indices are reported: the χ^2-statistic, Akaike's information criterion (AIC = $\chi^2 - 2df$; Akaike, 1987), and the root mean square error of approximation (RMSEA) which is the most appropriate fit statistic for large sample sizes.

Results

Factor analyses

As shown in Table 1, phenotypic correlations among the six teacher-assessed academic subjects are substantial, suggesting the presence of a general factor. When phenotypic correlations were calculated separately for Twin 1 and Twin 2 groups, differences from the correlation matrix shown in Table 1 for the total sample were negligible.

Table 1. Phenotypic correlation matrix for teacher-rated academic achievement

	S&L	*R*	*W*	*M*	*N*	*SS&M*
Speaking & listening	1.00					
Reading	.67	1.00				
Writing	.61	.73	1.00			
Mathematics	.62	.66	.65	1.00		
Numbers	.59	.66	.61	.83	1.00	
Shapes, space & measures	.61	.67	.65	.83	.85	1.00

The factor analytic results are shown in Table 2. As expected from the average correlation of .68 in Table 1, principal component factor analysis yielded a first unrotated principal component that accounted for 71% of the variance in teacher-assessed academic achievement (See Table 2.) The first unrotated principal component represents what the tests have in common and can thus be considered as a general factor. The individual academic subject loadings on this general factor were uniformly high, suggesting that the six scores are well represented by a general factor of academic achievement. Although the eigenvalue for the second factor was only .65, we used an oblique rotation (Promax) in an attempt to identify separate English and Mathematics factor scores. As expected from the average correlation among the English subtests (.67), among the Mathematics subtests (.84), and the average correlation between the English and Mathematics subtests (.64), two highly correlated (.70) rotated factors emerged representing English and Mathematics. These loadings are also shown in Table 2. Although the English and Mathematics factors are highly correlated, in the following analyses we include results for English and Mathematics as well as for the general factor, creating unit-weighted scales by standardizing scores for each subtest and summing the scores for the three English subtests (English score), the three Mathematics subtests (Mathematics score), and all six subtests (General score).

Table 2. Factor loadings on the first unrotated principal component (general factor) and on two rotated factors representing Mathematics and English from an oblique (promax) rotation

	1st PC	*Rotated factors*	
	General	*Mathematics*	*English*
Speaking & listening	.77	.58	1.86
Reading	.84	.67	.89
Writing	.80	.63	.87
Mathematics	.89	.93	.69
Numbers	.88	.95	.65
Shapes, space & measures	.89	.94	.68

Descriptive statistics

Means and standard deviations for the three scores are listed in Table 3 separately by sex and zygosity. The distributions were significantly kurtotic with an excess of average ratings of 2. Univariate analysis of variance (ANOVA) of the general factor showed that females performed significantly better than males ($F = 9.75$, $p = .00$); however the effect size was negligible ($R_2 = .005$; mean standard scores = .04 for females and −.05 for males). Furthermore, females performed significantly better than males in English ($F = 36.38$, $p = .00$), a difference that accounted for less than 2% of overall achievement ($R_2 = .016$; mean standard scores = .11 for females and −.13 for males). No significant gender differences were found for Mathematics performance ($F = 1.56$, $p = .21$; mean standard scores = −.02 for females and .03 for males). We also compared the performance of twins in the same classroom with twins in different classrooms. While an independent samples t test revealed that MZ pairs in the same classroom performed better than MZ pairs in different classrooms ($t = 2.19$, $df = 866$, $p = .03$), the effect size was again negligible ($R_2 = .006$). There were no significant mean differences in achievement for DZ twins in the same versus different classrooms.

Table 3. Means and standard deviations for teacher-assessed academic achievement by gender and zygosity

		General		Maths		English	
	N	Mean	*SD*	Mean	*SD*	Mean	*SD*
MZM	195	−.12	1.11	−.02	1.11	−.19	1.08
MZF	243	.03	.98	−.04	.97	.10	.97
DZM	172	−.03	.99	.03	1.01	−.09	.98
DZF	209	.02	.95	−.04	.94	.08	.96
DZO	372	.05	.98	.05	.98	.03	.99

Twin correlations

Table 3 lists twin intraclass correlations for the General score as well as for English and Mathematics. MZ correlations for academic achievement consistently exceeded those of the DZS and DZO groups, suggesting that teacher-rated academic achievement is influenced by genetics. Twin correlations tend to be lower when the twins are rated by different teachers, which is to be expected not only because there are different raters but also because the twins are in different classrooms. Nonetheless, the average difference between MZ and DZ correlations, which is the essence of the estimate of genetic influence, is similar when twins are rated by the same teacher or by different teachers. As expected from the high correlations in Table 1 and the factor analytic results in Table 2, results are similar for the General score, English and Mathematics as shown in Table 4. Although the correlations for opposite-sex DZ twins are similar to those for same-sex DZ twins for different-teacher assessments, same-teacher assessments show somewhat lower correlations for opposite-sex twins than for same-sex twins. In other words, when the same teacher assesses both twins, the teacher appears to assess same-sex pairs less similarly than opposite-sex pairs but when opposite-sex twins are assessed by different teachers, they are assessed as similarly as same-sex pairs. This difference could be due to genetic or environmental (including rating bias) factors. However, as described in the following section, model-fitting results do not confirm this as a significant difference.

Table 4. Intraclass correlations for twins rated by same and different teachers for general factor (1st principal component), Mathematics and English

	N	Same Teacher	*N*	Different Teachers
General Factor				
MZ	283	r = .91 (CI: .88–.92)	151	r = .78 (CI: .71–.83)
DZS	253	r = .65 (CI: .58–.72)	130	r = .50 (CI: .36–.62)
DZO	254	r = .48 (CI: .38–.57)	118	r = .54 (CI: .40–.66)
Mathematics				
MZ	282	r = .86 (CI: .82–.89)	151	r = .67 (CI: .57–.75)
DZS	253	r = .60 (CI: .51–.67)	130	r = .38 (CI: .22–.52)
DZO	254	r = .43 (CI: .33–.53)	118	r = .49 (CI: .34–.62)
English				
MZ	282	r = .91 (CI: .89–.93)	151	r = .78 (CI: .71–.83)
DZS	253	r = .63 (CI: .55–.70)	130	r = .55 (CI: .42–.66)
DZO	254	r = .47 (CI: .37–.56)	118	r = .47 (CI: .31–.60)

Maximum-likelihood liability model-fitting analysis

Maximum-likelihood liability model-fitting estimates of genetic and environmental influences on academic achievement for same and different teachers are shown in Figure 2 for the General score, English and Mathematics.

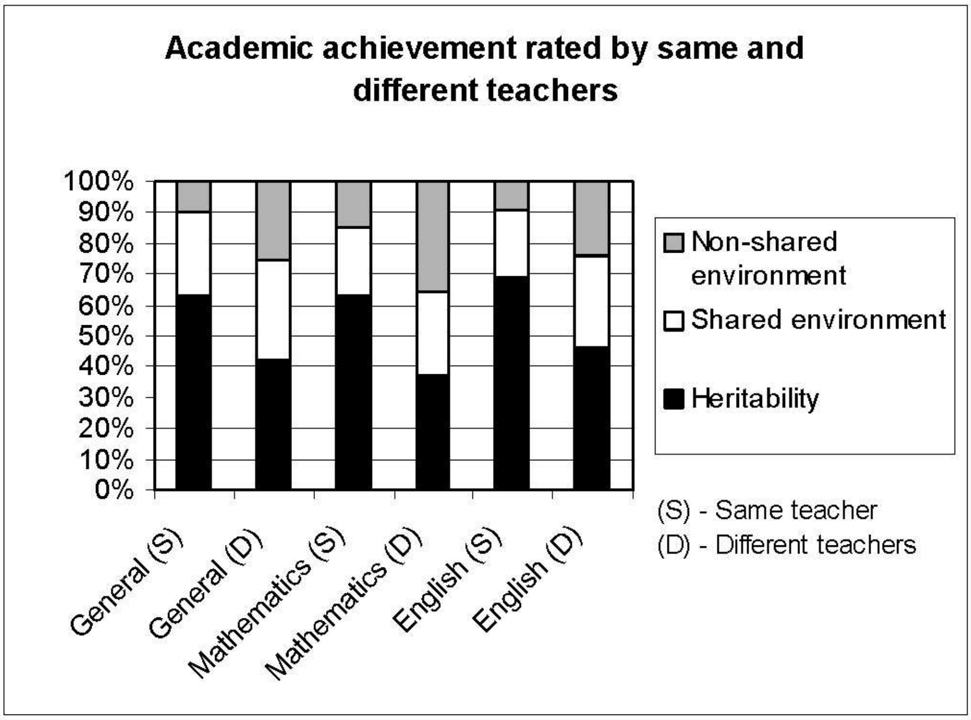

Figure 2. Parameter estimates for teacher-rated academic achievement
Note. All analyses are based on sex-corrected scores

The results were similar for all three scores. Using the General score as an example, the scalar sex differences model described above provided the most parsimonious fit with the data: χ^2 (22) = 28.67, p = .15, AIC = −15.34, RMSEA = 0.003, indicating that genetic and environmental parameter estimates can be equated for boys and girls. The variance was greater for males than for females, particularly when assessed by different teachers, and similar variance patterns emerged for both the Mathematics factor and for the English factor. Genetic influence for the general academic achievement factor is substantial (.63; 95% CI: .52–.75), shared environment is less (.27, 95% Cl: .15–.37), and non-shared environment is modest (.10; 95% Cl: .09–.12) when the same teacher assesses both members of a twin pair. When the twins are assessed by different teachers, heritability is considerable (.42; 95% Cl: .22–.63), shared environment is less (.33, 95% Cl: .14–.50), and non-shared environment is modest (.25, 95% Cl: .19–.31). Once again, the scalar sex differences model fit the Mathematics scale best (χ^2 (22) = 26.90, p = .22, AIC = −17.10, RMSEA = 0.000).

Results were similar for Mathematics and English. Mathematics when assessed by the same teacher showed sizeable genetic influence (.63, 95% Cl: .51–.76), modest shared environment (.22, 95% Cl: .09–.33), and less non-shared environment (.15, 95% Cl: .13–

.18). When assessed by different teachers, genetic influence was moderate (.37, 95% CI: .13–.62), non-shared environment was about the same (.36, 95% CI: .28–.45), and shared environment accounted for the balance (.27, 95% CI: .05–.46). Finally, the most parsimonious fit with the English factor was again the scalar sex differences model (χ^2 (22) = 33.00, p = .06, AIC = −11.03, RMSEA = .010). Genetic influence was sizeable when assessed by the same teacher (.69, 95% CI: .58–.81), shared environment was less (.22, 95% CI: .09–.33), and non-shared environment was negligible (.09, 95% CI: .08–.11). When the twins were assessed by different teachers, genetic influence accounted for just less than half of the variance (.46, 95% CI: .26–.67), shared environment was modest (.30, 95% CI: .10–.47), and non-shared environment made up the remainder (.24, 95% CI: .19–.31).

Discussion

The consistent pattern of higher MZ twin correlations relative to DZ twins suggests genetic influence on academic achievement, even when the twins were assessed by different teachers. Findings with regard to twins assessed by the same teacher supports an earlier Swedish twin study showing substantial genetic influence and modest shared environmental influence on academic performance (Husén, 1959).

The current study is the first behavioural genetic investigation of academic achievement in the early school years for twins assessed by different teachers as well as twins assessed by the same teacher and provides additional support for the hypothesis that genetics influences academic performance abilities. Model-fitting estimates were lower when twins were assessed by different teachers than by the same teacher. For example, for the general factor, model-fitting heritability estimates were .42 for different teachers and .64 for same teachers. However, the twin intraclass correlations shown in Table 3 indicate that this model-fitting result is caused by different results for same-sex and opposite-sex DZ twins. When rated by the same teacher, DZ twin correlations are greater for same-sex pairs (.65) than for opposite-sex pairs (.48). When rated by different teachers, DZ twin correlations are similar for same-sex pairs (.50) and opposite-sex pairs (.54). A similar pattern of results can be seen in Table 3 for Mathematics and English. In the 5-group model-fitting analysis, similarity for opposite-sex DZ twins was not found to be significantly different from similarity for same-sex DZ male and female pairs. The model-fitting averages out over the same-sex and opposite-sex DZ twins in order to estimate heritability, thus estimating greater heritability for same-teacher assessments than different-teacher assessments. If we ignore the opposite-sex twins and double the difference in the MZ and DZ correlations for the same-sex twins, the heritability estimates are highly similar for same-teacher assessments (.52) and different-teacher assessments (.56). Thus, the issue is not why heritability estimates are greater for same versus different teachers but rather why the opposite-sex DZ correlation is lower than the same-sex DZ correlation for same-teacher assessments but not for different-teacher assessments. One possibility is that same-teacher as contrasted with different-teacher assessments may be more subject to biases about sex differences in school achievement, especially in relation to the within-pair comparison of twins in the same classroom who are of opposite sex.

The larger issue is that both same-teacher and different-teacher assessments of academic achievement yield evidence for significant heritability, a comparison that has not been used previously in behavioural genetic research. Finding similar results for these two different types of data adds to the strength of the finding. Although it seems reasonable to assume that the results for different-teacher assessments are more valid, it is possible for example that seeing the two children together all day long makes same-teacher ratings more valid.

We emphasize that our conclusions are limited to teacher assessments of academic achievement. Although we did not obtain data about the reliability or validity of teacher assessments, the high correlations for the twins – even when rated by different teachers – provides strong evidence for both reliability and validity. Evidence exists that teacher assessments reveal some bias, if one accepts discrepancies with test scores as a sign of bias (Davies & Brember, 1994; Reeves, Boyle, & Christie, 2001). It would be interesting to compare teacher assessments and test scores, and this is what we are doing in the ongoing 9-year phase of TEDS. As noted earlier, we have recently reported a correlation of .68 between a brief test of early word recognition administered via the telephone (Torgesen *et al.*, 1999) and teacher KS1 assessments of reading (Dale *et al.*, 2003).

However, the use of teacher assessments rather than test scores is a strength as well as a limitation of the current study. Test scores are not without their own biases (Good & Salvia, 1988; Livingstone, 1995; Marks, 1990; Sharpley & Edgar, 1986). We believe that teachers – especially UK teachers following National Curriculum guidelines for performance rather than purely subjective ratings – have a broader perspective on students' performance than measures of test-taking performance. Teachers also assess performance over a long time in the 'real-world' contexts of the classroom. For these reasons, 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 predict long-term educational attainment and career outcomes (McCall, Evahn, & Kratzer, 1992). Furthermore, teacher judgments of children were shown to be good predictors of high school achievement test scores (Hoge & Butcher, 1984), and high school grades have been linked to subsequent academic achievement, as well as career success (Hauser, Sewell, & Alwin, 1976; Sewell, Hauser, & Wolf, 1980).

Regardless of opinions about the validity and value of teacher assessments, it should be noted that NC teacher assessments are the scores that affect children's lives in the British educational system. Even if teacher ratings were completely biased assessments, the results of our study would still be interesting – it needs to be explained why the heritability of these measures is so high even when different teachers provide the assessments. A more specific issue that warrants further investigation concerns the finding in our study, and the only other study of teacher assessments of academic achievement, that heritability estimates are greater than heritability estimates typically found in studies of achievement tests. Although more research is needed to demonstrate conclusively that heritability is greater for teacher assessments than for achievement tests, what could account for such a difference? It is possible that teachers detect phenotypic characteristics that are not captured by objective tests of academic achievement. Academic ability as measured by teacher assessments may present a more comprehensive picture of a child's abilities than test scores alone (i.e., street-smart children may outperform book-smart children in ways not measured by standardized tests). Furthermore, teacher assessments might capture important traits sometimes referred to as 'emotional intelligence' that are missed by achievement tests, traits that may play a significant role in children's long-term achievement in life, such as personality, motivation, sociability, resilience and disposition (Goleman, 1995). Complementary constellations of such characteristics might contribute importantly to long-term life successes and failures beyond the predictive power of test scores and they might also add sources of genetic variance independent of test scores.

Most educational research on inter-individual variation in academic achievement has focused on nurture-related explanations such as family and school environments (Eccles *et al.*, 1998; McDermott, 1995; Reynolds & Lee, 1991). Although the most novel aspect of our findings is that genetics contributes importantly to individual differences in academic achievement in the early school years, the results also provide strong support for environmental influence, suggesting that environmental influences account for more than one-third of the total variance. More specifically, shared environment accounted for a moderate amount of variance when twins were rated by different teachers as well as when twins were rated by the same teachers. This shared environmental influence could be due to family environment or to school environment, which twins share even if they are not in the same classroom. It is important to reiterate that shared environment extends beyond the home – it can include any experience that contributes to similarities rather than differences within twin pairs. One possibility is that twins share environmental experiences to a greater extent than non-twin siblings because twins are the same age and thus travel through life together (Koeppen-Schomerus *et al.*, 2003). We will be able to test this hypothesis because younger siblings of the TEDS twins will also be assessed by their teachers when the siblings reach 7 years of age. It is possible that the KS-1 SATS will change during the next few years, which could complicate the comparison by artificially lowering the correlation between the twins and their siblings.

Although shared environment can occur outside of the home, it is also important to recognize that perceptions are unique to the individual, and that similar environments are not experienced the same way by all children. Specifically, although school would seem by definition to be the 'same' environment for twins attending the same school, the twins' perceived or actual experiences at school could be very different which would lead to nonshared environment. Nonshared environment accounted for a somewhat greater proportion of the variance when the children were rated by different teachers (25%) than by the same teachers (10%). One might assume that this finding occurs for methodological reasons – when rated by different teachers, the twins are seen as less similar. In the current study, the finding that shared environmental estimates are highly similar for same-teacher and different-teacher ratings speaks against the hypothesis that identical twins are rated more similarly because they are treated alike. The greater nonshared environment estimate for different teacher ratings than for same teacher ratings reflects the lower heritability estimate for different teacher ratings. One possibility is that teachers are able to detect genetically influenced characteristics to a greater extent if they compare both twins than if one teacher rates one twin and another teacher rates the other. Another possibility is that twins in different classrooms may have more distinct genetically-mediated experiences, as children's genetic propensities interact in unique ways with each separate classroom environment. Distinctive gene-environment interactions, or genetic sensitivity to certain elements of the environment, would enhance individual differences and increase estimates of nonshared environment. Although genotype-environment interactions have been reported for adolescent conduct disorder (Cadoret, Yates, Troughton, & Woodworth, 1995) and adolescent antisocial behaviour (Cadoret, Cain, & Crowe, 1983; Crowe, 1974), elucidating gene-environment interactions with respect to cognitive abilities has been more elusive (Plomin, DeFries, & Fulker, 1988). Because a better understanding of differential responses to classroom experiences and educational interventions is essential in providing appropriate support to students, an important direction for future research is to investigate the interplay between nature and nurture in the classroom.

Behavioural genetic research can help to elucidate the role of environmental factors that operate in the school environment. For example, the present results suggest that environmental influences on academic achievement are equally divided between shared and nonshared factors. Shared factors for example would include shared family background factors, which are often considered as environmental contributors to academic achievement. However, less attention has been directed towards understanding the nonshared environmental experiences relevant to academic achievement. The key to understanding nonshared environmental experiences is to study siblings and to ask why siblings growing up in the same family have differences in academic experiences. Moreover, results from the current study highlight the need to evaluate environmentally-driven causal explanations for academic achievement in genetically sensitive designs that attempt to disentangle nature and nurture. That is, if individual differences in academic achievement are highly heritable, as our results suggest, this implies the possibility that genetic factors may contribute to relationships between academic achievement and family background factors.

An exploration of mean gender and zygosity differences yielded few significant results. While much research has focused on mean gender differences in academic performance, evidence has been somewhat mixed (Naglieri & Rojahn, 2001). A summary of some of the major findings show that girls have an advantage over boys in verbal fluency, reading and writing, foreign language, fine motor skills, maths calculation, and speech articulation, while boys have been found to do better in maths and science knowledge, mechanical reasoning and verbal analogies (Halpern, 1997). The present results support the view that mean gender differences are slight at 7 years old, at least on teacher assessments of academic achievement. For example, although girls outperformed boys in English and for the general factor, the difference accounted for less than 2% and for less than 1% of the variance, respectively. Concerning the origins of individual differences for boys and girls, which is an independent issue from mean gender differences, we find no evidence for quantitative or qualitative differences in genetic and environmental influences on boys and girls.

An issue of urgency for parents of twins is whether to separate the children in school in order to enhance their development as individuals (Bryan, 1992). A study of Australian twins found no important differences in individuality when twins were placed in different classrooms, even though over 90% of teachers considered this the main rationale for separating the twins (Gleeson, Hay, Johnston, & Theobald, 1990). Although the results of the current study pointed to a slightly higher performance for MZ pairs in same classroom versus different classrooms, the results were not significant and accounted for less than 1% of the variance. No significant difference in same versus different teacher-rated achievement scores was found for DZ pairs.

Although genetic contributions to behavioural traits are becoming widely accepted throughout the behavioural sciences even for cognitive abilities (e.g., Neisser, 1997), acceptance of nature as well as nurture appears to be slower in educational psychology (see Plomin & Walker, 2003). Genetic influence on disorders such as dyslexia, autism and hyperactivity is discussed in a few educational textbooks. However, there is little coverage of genetically-influenced individual differences within the normal range. Moreover, very few articles on genetics have been published in major journals of educational psychology (Plomin & Walker, 2003), despite the considerable body of evidence for the importance in educationally relevant behaviours such as cognitive abilities and disabilities and behaviour problems (Plomin *et al.*, 2001a). It is our hope that educational psychology will not re-live the nature-nurture battles fought two decades ago in other areas of the behavioural sciences. Rather, educational psychology can capitalise on coming late to genetics by embracing a more balanced position that acknowledges the importance of nature as well as nurture and uses genetic designs to ask questions that go beyond heritability, such as questions about the developmental interplay between nature and nurture in educational psychology.

The larger significance of the present finding of substantial genetic influence on academic achievement lies in beginning to bridge the wide gap between the field of education and the field of genetics. The field of education scarcely acknowledges genetics even though schools are the primary societal mechanism for fostering cognitive development, and cognitive development shows substantial genetic influence. Finding genetic influence will not denigrate the role of education; rather, it will suggest new ways of thinking about effective education, such as recognizing that children create their own experience within the educational process in part on the basis of their genetic propensities.

Some of the reluctance to accept genetics may be specific to the history and epistemology of education and educational psychology (Wooldridge, 1994). However, much of this reluctance is likely to involve general misconceptions about what it means to say that genetics is important (Rutter & Plomin, 1997). These misconceptions need to be overcome, especially the mistaken view that the word *genetic* connotes hard-wired defects. This deterministic misconception fuels a feeling of environmental nihilism, that is, if a disorder is heritable, there is nothing that we can do about it environmentally (Sternberg & Grigorenko, 1999). The myth of environmental nihilism feeds into a related myth that finding genetic influence will serve to justify social inequality. We do not accept this view. Knowledge alone does not account for societal and political decisions – values are just as important in the decision-making process and it is to be hoped that better decisions can be made with knowledge than without (Pinker, 2002).

Acknowledgements

We thank the parents of the twins in the Twins Early Development Study (TEDS) for making this study possible. TEDS is supported by a programme grant from the UK Medical Research Council.

References

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

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

Bacete, F.G., & Remirez, J.R. (2001). Family and personal correlates of academic achievement. *Psychological Reports, 88*, 533–547.

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

Bouchard, T. J., Jr., & Propping, P. (1993). *Twins as a tool of behavioral genetics*. New York: John Wiley & Sons.

Bryan, E. M. (1992). *Twins and higher multiple births: A guide to their nature and nurture*. London: Edward Arnold.

Cadoret, R. J., Cain, C. A., & Crowe, R. R. (1983). Evidence from gene-environment interaction in the development of adolescent antisocial behavior. *Behavior Genetics, 13*, 301–310.

Cadoret, R. J., Yates, W. R., Troughton, E., & Woodworth, G. (1995). Genetic-environmental interaction in the genesis of aggressivity and conduct disorders. *Archives of General Psychiatry, 52*, 916–924.

Christenson, S. L., Rounds, T., & Gorney, D. (1993). Family factors and student achievement: An avenue to increase students' success. *School Psychology Quarterly, 7*, 178–206.

Crowe, R. R., (1974). An adoption study of antisocial personality. *Archives of General Psychiatry, 31*, 785–791.

Dale, P. S., Harlaar, N., & Plomin, R. (2003). *A comparison of two nonconventional procedures for the research assessment of reading skills in 7-year-olds*. Manuscript submitted for publication.

Dale, P. S., Simonoff, E., Bishop, D. V. M., Eley, T. C., Oliver, B., Price, T. S., Purcell, S., Stevenson, J., & Plomin, R. (1998). Genetic influence on language delay in 2-year-old children. *Nature Neuroscience, 1*, 324–328.

Davies, J., & Brember, I. (1994). The first mathematics SATs at KS1. *British Educational Research Journal, 20*, 35–40.

Dunn, J. F., & Plomin, R. (1990). *Separate lives: Why siblings are so different*. New York: Basic Books.

Eccles, J., Wigfield, A., & Schiefele, U. (1998). Motivation to succeed. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology: Vol 3. Social, emotional, and personality development* (5th ed.) (pp. 1017–1095). New York: Wiley.

Fisher, S. E., & DeFries, J. C. (2002). Developmental dyslexia: Genetic dissection of a complex cognitive trait. *Nature Reviews: Neuroscience, 3*, 767–782.

Galsworthy, M. J., Dionne, G., Dale, P. S., & Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development. *Developmental Science, 3*, 206–215.

GHS (2000/01). *General Household Survey*. Great Britain Office of Population Censuses and Surveys.

Gleeson, C., Hay, D. A., Johnston, C. J., & Theobald, T. M. (1990). 'Twins in School'. An Australian-wide program. *Acta Genet Med Gemellol, 39*, 231–244.

Goleman, D. (1995). *Emotional intelligence*. New York: Bantam Books.

Good, R. H., & Salvia, J. (1988). Curriculum bias in published, norm-referenced reading tests: Demonstrable effects. *School Psychology Review, 17*, 51–60.

Gutknecht, L., Spitz, E., & Carlier, M. (1999). Long-term effect of placental type on anthropometrical and psychological traits among monozygotic twins: A follow up study. *Twin Research, 1999*, 212–217.

Halpern, D. F. (1997). Sex differences in intelligence: Implications for education. *American Psychologist, 52*, 1091–1102.

Hauser, R., Sewell, W., & Alwin, D. (1976). High school effects on achievement. In W. Sewell, R. Hauser & D. Featherman (Eds.), *Schooling and achievement in American society* (pp. 309–341). New York: Academic Press.

Hoge, R. D., & Butcher, R. (1984). Analysis of teacher judgments of pupil achievement levels. *Journal of Educational Psychology, 76*, 777–781.

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

Husén, T. (1959). *Psychological twin research*. Stockholm: Almqvist & Wiksell.

Kendler, K. S., Neale, M. C., Kessler, R. C., Heath, A. C., & Eaves, L. J. (1993). A test of the equal-environment assumption in twin studies of psychiatric illness. *Behavior Genetics, 23*, 21–27.

Koeppen-Schomerus, G., Spinath, F. M., & Plomin, R. (2003). Twins and non-twin siblings: Different estimates of shared environmental influence in early childhood. *Twin Research*, *6*(2), 97–105.

Lichtenstein, P., Holm, N. V., Verkasalo, P. K., Iliadou, A., Kaprio, J., Koskenvuo, M., Pukkala, E., Skytte, A., & Hemminki, K. (2000). Environmental and heritable factors in the causation of cancer. *New England Journal of Medicine, 343*, 78–85.

Livingstone, D. W. (1995). For whom The Bell Curve tolls. *Alberta Journal of Educational Research, 41*, 335–341.

Loehlin, J. C., & Nichols, J. (1976). *Heredity, environment, and personality*. Austin, TX: University of Texas Press.

MacGillivray, I., Campbell, D. M., & Thompson, B. (1988). *Twinning and twins*. Chichester: John Wiley & Sons.

Marks, D. (1990). Cautions in interpreting district-wide standardized mathematics achievement test results. *Journal of Educational Research, 83*, 349–354.

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

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

McDermott, P. A. (1995). Sex, race, class and other demographics as explanations for children's ability and adjustment: A national appraisal. *Journal of School Psychology, 33*, 75–91.

Menaghan, E. G., Kowaleski-Jones, L., & Mott, F. L. (1997). The intergenerational costs of parental social stressors: Academic and social difficulties in early adolescence for children of young mothers. *Journal of Health & Social Behavior, 38*(1), 72–86.

Merriman, C. (1924). The intellectual resemblance of twins. *Psychological Monographs, 33*, 1–58.

Morris-Yates, A., Andrews, G., Howie, P., & Henderson, S. (1990). Twins: A test of the equal environments assumption. *Acta Psychiatrica Scandinavica, 81*, 322–326.

Naglieri, J. A., & Rojahn, J. (2001). Gender differences in planning, attention, simultaneous, and successive (PASS) cognitive processes and achievement. *Journal of Educational Psychology, 93*(2), 430–437.

Neale, M. (1997). *Mx: Statistical modeling*. Box 126 MCV, Richmond, VA 23298: Department of Psychology.

Neisser, U. (1997). Never a dull moment. *American Psychologist, 52*, 79–81.

Nye, B., Hedges, L. B., & Konstantopoulos, S. (2001). Are effects of small classes cumulative? Evidence from a Tennessee experiment. *Journal of Educational Research, 94*(6), 336–345.

Office for National Statistics (2002). *Living in Britain: Results from the 2000/01 General Household Survey*. Retrieved November 28, 2002 from http://www.statistics.gov.uk/lib/resources/fileAttachments/GHS2000.pdf

Phelps, J. A., Davis, O. J., & Schwartz, K. M. (1997). Nature, nurture and twin research strategies. *Current Directions in Psychological Science, 6*, 117–121.

Phillips, D. I. W. (1993). Twin studies in medical research: Can they tell us whether diseases are genetically determined? *Lancet, 341*, 1008–1009.

Pinker S. (2002). *The blank slate: The modern denial of human nature*. New York: Penguin.

Plomin, R. (1999). Genetics and general cognitive ability. *Nature, 402*, C25–C29.

Plomin, R, (2003). General cognitive ability. In R. Plomin, J. C. DeFries, I. W. Craig & P. McGuffin (Eds.), *Behavioral genetics in the postgenomic era* (pp. 183–201). Washington, DC: APA Books.

Plomin, R., & Walker, S. O. (2003). Genetics and Educational Psychology. *British Journal of Educational Psychology, 73*, 3–14.

Plomin, R., DeFries, J. C., & Fulker, D. W. (1988). *Nature and nurture during infancy and early childhood*. Cambridge: Cambridge University Press.

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

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

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

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.

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

QCA (1999). *QCA Key Stage 1: Assessment and reporting arrangements* (1999). Great Britain: Qualifications and Curriculum Authority.

QCA (2002). *QCA Key Stage 1: English tests teachers guide* (2002). Qualifications and Curriculum Authority.

Reeves, D. J., Boyle, W. F., & Christie, T. (2001). The relationship between teacher assessments and pupil attainments in standard test tasks at Key Stage 2, 1996–98. *British Educational Research Journal, 27*, 141–160.

Reynolds, A. J., & Lee, J. S. (1991). Factor analyses of measures of home environment. *Educational & Psychological Measurement, 51*, 181–192.

Riese, M. L. (1999). Effects of chorion type on neonatal temperament differences in monozygotic pairs. *Behavior Genetics, 29*, 87–94.

Rutter, M. (1981). Epidemiological – longitudinal approaches to the study of development. In W.A. Collins (Ed.), *Minnesota Symposia on Child Psychology: Vol 15. The concept of development* (pp. 105–144). Hillsdale, NJ: Erlbaum.

Rutter, M., & Plomin, R. (1997). Opportunities for psychiatry from genetic findings. *British Journal of Psychiatry, 171*, 209–219.

Rutter, M., & Redshaw, J. (1991). Annotation: Growing up as a twin: Twin-singleton differences in psychological development. *Journal of Child Psychology and Psychiatry, 32*, 885–895.

Scarr, S., & Carter-Saltzman, L. (1979). Twin method: Defense of a critical assumption. *Behavior Genetics, 9*, 527–542.

Sewell, W., Hauser, R., & Wolf, W. (1980). Sex, schooling and occupational status. *American Journal of Sociology, 86*, 551–583.

Sharpley, C. F., & Edgar, E. (1986). Teachers' ratings vs. standardized tests: An empirical investigation of agreement between two indices of achievement. *Psychology in the Schools, 23*, 106–111.

Sokol, D. K., Moore, C. A., Rose, R. J., Williams, C. J., Reed, T., & Christian, J. C. (1995). Intrapair differences in personality and cognitive ability among young monozygotic twins distinguished by chorion type. *Behavior Genetics, 25*, 457–466.

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.

Sternberg, R. J., & Grigorenko, E. L. (1999). Myths in psychology and education regarding the gene-environment debate. *Teachers College Record, 100*, 536–553.

Stromswold, K. (2001). The heritability of language: A review and metaanalysis of twin, adoption and linkage studies. *Language, 77*, 647–723.

Teo, A., Carlson, E., Mathieu, P., Egeland, B., & Sroufe, L. A. (1996). A prospective longitudinal study of psychosocial influence on achievement. *Journal of School Psychology, 34*, 285–306.

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.

Torgeson, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency (TOWRE)*. Austin, TX: Pro-Ed.

Wilson, R. S. (1983). The Louisville Twin Study: Developmental synchronies in behavior. *Child Development, 54*, 298–316.

Wooldridge A (1994). *Measuring the mind: Education and psychology in England, c. 1860–c. 1990*. Cambridge: Cambridge University Press.

Received 22 May, 2002; revised version received 20 March, 2003