



Published in final edited form as:

J Learn Disabil. 2007 ; 40(6): 554–567.

Mathematical Ability of 10-Year-Old Boys and Girls: Genetic and Environmental Etiology of Typical and Low Performance

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

Abstract

The genetic and environmental etiologies of 3 aspects of low mathematical performance (math disability) and the full range of variability (math ability) were compared for boys and girls in a sample of 5,348 children age 10 years (members of 2,674 pairs of same-sex and opposite-sex twins) from the United Kingdom (UK). The measures, which we developed for Web-based testing, included problems from 3 domains of mathematics taught as part of the UK National Curriculum. Using quantitative genetic model-fitting analyses, similar results were found for math disabilities and abilities for all 3 measures: Moderate genetic influence and environmental influence were mainly due to nonshared environmental factors that were unique to the individual, with little influence from shared environment. No sex differences were found in the etiologies of math abilities and disabilities. We conclude that low mathematical performance is the quantitative extreme of the same genetic and environmental factors responsible for variation throughout the distribution.

Compared to reading disabilities (Pennington & Olson, 2005), relatively little is known about the genetic and environmental etiologies of mathematics abilities and disabilities, the extent to which these etiologies differ for low performance and typical performance, and the extent to which these etiologies differ for male and female students. The present study provides the first analysis of these issues using a sample of twins large enough to investigate the origins of low mathematics performance (which we will refer to as *math disability*) in the context of normal variation (or ability) and to consider sex differences in both abilities and disabilities.

Four previous twin studies of general mathematics performance have yielded a wide range of results (see Plomin & Kovas, 2005). Our previous research on mathematics disabilities and abilities, based on teacher ratings using the United Kingdom (UK) National Curriculum criteria, was the first study to compare mathematics disabilities and abilities and found similar genetic and environmental results for math disabilities and abilities (Oliver et al., 2004). This work focused on three aspects of mathematics identified by the UK National Curriculum for 7-year-old children: Using and Applying Mathematics; Number and Algebra; and Shapes, Space and Measures. Substantial genetic influence (65% heritabilities) was found both for mathematics disabilities (lowest 15%) and for mathematics abilities (the rest of the distribution). Moreover, the results indicated not only that the magnitude of genetic and environmental influences was similar but also that the same genetic and environmental factors drive individual differences in the whole of the distribution of ability—at the low end as well as the rest of the distribution (see Method section for an explanation of the relevant analytic methods). At the level of the environment, the position of an individual on the mathematics continuum depends on a cumulative effect of many risk and protective environments. Similarly, at the genetic level, the position of an individual on an ability continuum depends on a cumulative effect of risk or protective variants of many genes. With many environmental and

genetic effects and possible interactions and correlations between them contributing to the position of an individual on an ability distribution, the influence of each effect is likely to be small and probabilistic, leading to a normal distribution.

Moreover, for both math disabilities and abilities, similarity between twins in the same family appeared to stem almost entirely from similarity in their genetic makeup rather than similarity in their environments. That is, relevant environmental influences were of the nonshared type that make children in the same family different from each other. This previous study included only same-sex twins and did not examine possible sex differences in genetic and environmental etiologies for mathematics disabilities and abilities. The present study includes opposite-sex twins, which provide unique power for investigating etiological issues concerning sex differences.

Although there is considerable evidence for the validity of teacher ratings based on UK National Curriculum criteria, it is important to confirm these results using test scores. However, a practical problem in conducting genetic research with individually administered tests is that large samples of twins are needed, but it is expensive to test large samples in person, especially when the twins are distributed over a wide area. We have addressed this problem by developing Web-based tests that make it possible to assess large samples efficiently and economically. We assessed three aspects of mathematics based on the UK National Curriculum for 10-year-old children:

1. understanding of numerical and algebraic processes to be applied when solving problems,
2. understanding of non-numerical mathematical processes and concepts, and
3. fact retrieval and computation.

A preliminary study when the data from only one cohort (i.e., half of the sample) were available also assessed two other aspects of mathematics (interpreting information and using mathematics in problem-solving situations); an article focused on the multivariate genetic relationship among these five aspects of mathematics concluded that there was substantial overlap, thus leading to the decision to drop these two tests when assessing the rest of the sample (Kovas, Petrill & Plomin, 2007). This preliminary study only addressed individual differences across the entire sample for boys and girls combined—the sample was not large enough to investigate mathematics disability, to compare the etiologies of disabilities and abilities, or to examine sex differences. The present article reports results at 10 years, using a sample more than twice as large, which is sufficiently large to compare results for disabilities and abilities and for boys and girls.

Concerning the comparison between math disabilities and abilities, our previous research based on teacher ratings (Oliver et al., 2004) and similar findings for reading disabilities and abilities (Plomin & Kovas, 2005) led us to predict similar results for the present study based on individually administered tests of mathematics performance. In respect to sex differences, we examined for the first time sex differences in the etiologies of diverse aspects of mathematics for both disabilities and abilities. Although mean sex differences in mathematical ability have been reported in the past (Maccoby & Jacklin, 1974), recently a “gender similarities” hypothesis has emerged, suggesting that male and female students are similar on most psychological variables, including mathematics (Hyde, 2005; Spelke, 2005). Although we consider sex differences in means and variances, our focus is on sex differences in genetic and environmental etiologies of mathematical abilities and disabilities, which are independent of mean sex differences. Research is not available to suggest hypotheses in relation to sex differences in genetic and environmental etiologies in mathematics, or whether a different

picture of sex differences emerges for math disabilities and abilities and for different aspects of mathematics.

To summarize, the present study uses a large sample of twins that allows us to provide the first analysis of three important issues:

1. genetic and environmental etiologies of mathematics abilities and disabilities in 10 year-old children,
2. the extent to which these etiologies differ for low performance and typical performance, and
3. the extent to which these etiologies differ for boys and girls.

Method

Participants and Procedure

Participants were part of the Twins' Early Development Study (TEDS), a longitudinal study involving a representative sample of all twins born in England and Wales in 1994, 1995, and 1996 (Trouton, Spinath, & Plomin, 2002). When the twins were 10 years old, data on a wide variety of cognitive measures were collected from 7,442 children born between January 1994 and August 1996. This number refers to all children who took part in the battery after the following exclusion criteria were applied: specific medical syndromes, such as Down syndrome and other chromosomal anomalies, cystic fibrosis, cerebral palsy, hearing loss, autism spectrum disorder, organic brain damage, extreme outliers for birth weight, gestational age, maternal alcohol consumption during pregnancy, and special care after birth. We excluded from the sample 274 children (137 pairs) for whom English was not the only language spoken at home. After these exclusions, the sample consisted of 7,168 children (3,584 pairs). From this sample, data for mathematics were available for 5,348 individuals (2,674 pairs of same-sex and opposite-sex twins). Due to the nature of the test administration, each of these children had complete data on all three sub-tests. To examine whether this attrition would affect further results, we compared National Curriculum math scores for children who completed the math battery with those who did not. Using standard scores based on the entire TEDS sample (after medical exclusions), we showed that children who completed the Web-based math battery performed slightly better than average on math ($M = 0.08$, $SD = 0.92$). Those who did not complete the math battery had slightly lower than average math scores ($M = -0.18$, $SD = 1.03$). These mean differences account for less than 1% of the variance. Zygosity was ascertained by parental ratings, with an error rate not exceeding 5%, as validated by DNA genotyping (for details, see Freeman et al., 2003).

Measures

Informed consent was obtained in writing from all of the families who agreed to take part in the study. To assess mathematics skills, we developed a Web-based battery that included questions from three different components of mathematics. The items were based on the National Foundation for Educational Research 5–14 Mathematics Series, which is linked closely to curriculum requirements in the UK and the English Numeracy Strategy (NFER-Nelson, 1994; NFER-Nelson, 1999; NFER-Nelson, 2001). Such curriculum-based assessment alleviates some of the potential biases associated with other achievement tests (Good & Salvia, 1988). From Booklets 6, 7, 8, 9, 10, and 11 (referring to age of students), a total of 77 target items were chosen. The items were organized by mathematical category (see hereafter) and level of difficulty. The level of difficulty was based on the National Curriculum level and the standardization sample percentage correct for each item (reported in the Group Record Sheets; NFER-Nelson). A set of adaptive branching rules was developed separately for each of the five categories, so that all the children started with the same items, but then were branched to

easier or harder items depending on their performance. The presentation of items was streamed, so that items from different categories were mixed, but the data recording and branching were done within each category. Participants could attempt each item only once.

As with many psychological tests that use branching (e.g., third edition of the *Wechsler Intelligence Scale for Children*; WISC-III-UK; Wechsler, 1992), the generic scoring rules were as follows: 1 point was recorded for each correct response, for each unadministered item preceding the child's starting point, and for each item skipped through branching to harder items. After a certain number of failures, a discontinuation rule was applied within each category, and no points were recorded for all items after discontinuation. Thus, for each of the 77 items, a score of 1 or 0 was recorded for each child. For example, in the Computation and Knowledge category (total number of items = 31), all children started at Item 10. The following rules were then applied:

1. If Items 10, 11, and 12 were all answered incorrectly, the child was branched to Item 1 and had to continue with the test attempting all remaining items, or until the discontinuation criterion was met.
2. If Items 10, 11, and 12 were all answered correctly, the child received credit for all preceding items (1–9), and was branched to Item 24. If Items 24, 25, and 26 were all answered incorrectly, the child was branched back to Item 13 and had to continue with the test (skipping all items previously administered), attempting all remaining items, or until the discontinuation criterion was met. If Items 24, 25, and 26 were not all answered incorrectly, the child received credit for all preceding items (13–23) and then continued with the test, attempting Items 27, 28, 29, 30, and 31, or until the discontinuation criterion was met.
3. If Items 10, 11, and 12 were not all answered incorrectly or correctly, the child received credit for all preceding items (1–9) and then had to continue with the test, attempting at all remaining items or until the discontinuation criterion was met.
4. Discontinuation criterion was 3 incorrect answers in a row (did not apply across branching points).

As with other psychological tests with items of increasing difficulty and using similar rules, this scoring system is equivalent to that in which all children attempt all items, allowing us to calculate the total number and proportion of correct responses for each child for each category and to test the internal consistency of each category (see hereafter). Specific branching and discontinuation rules and the number of skipped (credited) items for each category are available from the authors.

The items were drawn from the following three categories:

- **Understanding Number** (27 items) requires an understanding of the numerical and algebraic process to be applied when solving problems (e.g., understanding that multiplication and division are inverse operations). For example, “Look at the number 6,085. Change the order of the figures around to make the biggest number possible.” Another example is, “Type the missing number in the box: $27 + 27 + 27 + 27 + 27 = 27 \times \underline{\quad}$.”
- **Non-Numerical Processes** (19 items) requires understanding of non-numerical mathematical processes and concepts, such as rotational or reflective symmetry and other spatial operations. The questions do not have any significant numerical content that needs to be considered by the pupils. Three examples follow: “Which is the longest drinking straw? Click on it.” “One of these shapes has corners that are the

same. Click on this shape.” “Which card appears the same when turned upside down? Click on it.”

- **Computation and Knowledge** (31 items) assesses the ability to perform straightforward computations using well-rehearsed pencil-and-paper techniques and the ability to recall mathematical facts and terminology. These questions are either mechanistic or rely on memorizing mathematical facts and terminology. The operation is stated or is relatively unambiguous. Three examples follow. “Type in the answer: $76-39 = \underline{\quad}$.” “All 4-sided shapes are called? Click on the answer (squares rectangles parallelograms kites quadrilaterals).” “Type in the answer: $149 + 785 = \underline{\quad}$.”

In creating this battery, we aimed to choose tests that would between them cover a wide range of mathematical abilities of 10-year-old children, without subscribing to a particular theoretical position on the structure of the mathematical domain. Some overlap in what these tests measure is inevitable, as they each make demands on overlapping cognitive and performance factors (e.g., attention, motivation, memory), and it is never possible to achieve a completely “pure” measure of any of these components. In fact, the three tests correlated .62, and a principal components analysis of the three tests yielded a general factor (un-rotated first principal component) that explained 75% of the total variance. Nonetheless, we present the results for the three tests separately for the following reasons. Although one can never completely control for the multiple influences on performance of a mathematical test, it is nevertheless possible to choose tests that emphasize one component of mathematics more than another. Most important, the three aspects of mathematics that we assessed were not arbitrarily chosen: They are used in the UK National Curriculum to assess aspects of mathematics.

As noted previously, the use of Web-based assessment facilitates data collection because it allows data from large, widely dispersed samples to be collected quickly, cheaply, and reliably. Web-based data collection is less error prone because it does not require human transcription and data entry (Kraut et al., 2004; Naglieri et al., 2004). Another positive aspect of Web testing is that the social pressure or embarrassment that might be present in face-to-face testing is reduced (Birnbaum, 2004; Kraut et al., 2004). Moreover, several recent empirical studies have found that Web-based findings generalize across presentation formats and are consistent with findings from traditional methods (e.g., Gosling, Vazire, Srivastava, & John, 2004).

In TEDS, 80% of the families have daily access to the Internet (based on a pilot study with 100 randomly selected TEDS families), which is similar to the results of market surveys of UK families with adolescents. Most children without access to the Internet at home have access in their schools and local libraries.

In designing our Web-based battery, we guarded against potential problems associated with research on the Internet. The Web page and testing were administered by a secure server in the TEDS office (see Note). We used a secure site for data storage, and identifying information was kept separately from the data. Appropriate safeguards were in place that prevented children from answering the same item more than once. We provided technical support and other advice to parents and children who were advised to call our toll-free telephone number in case of any problems or questions.

Parents supervised the testing by coming online first with a user name and password for the family, examining a demonstration test and completing a consent form. Then parents allowed each twin to complete the test in turn. Parents were urged not to assist the twins with answers, and we are confident that most parents complied with this requirement.

The internal consistency of the Web-administered measures was examined, yielding high Cronbach's alpha coefficients (Understanding Number, $\alpha = .88$; Non-Numerical Processes, $\alpha = .78$; Computation and Knowledge, $\alpha = .93$). Moreover, we compared children's overall Web-based performance in mathematics at 10 years to their overall mathematics performance in the classroom as assessed by their teachers on the National Curriculum criteria when the children were 10 years old. We found a correlation of .53 ($p < .001$, $n = 1,878$). Only one twin from each pair was randomly selected for this analysis; a similar correlation of .50 was found for the other half of the sample.

As a direct test of the reliability and validity of the Web-based measures, we conducted a test-retest study in which 30 children (members of 15 twin pairs) who had completed the Web-based testing completed the tests in person using the standard 12-year paper-and-pencil version of the test (NFER-Nelson, 2001). Stratified sampling was used to ensure coverage of the full range of ability. The interval between test and retest was 1 to 3 months, with an average of 2.2 months. The total math score from our Web-based tests correlated .92 with the total score from the in-person testing for the total sample of 30 children; generalized estimation equations (GEE) that took into account the nested covariance structure also yielded a correlation of .93. For the three subtests reported in this article, the correlations between the Web-based and the paper-and-pencil scores were .77, .64, and .81 for Understanding Number, Non-Numerical Processes, and Computation and Knowledge, respectively. These results demonstrate that our Web-based testing is both highly reliable and valid.

Results

Descriptive Statistics and Further Exclusions

The data were first explored using descriptive statistics analyses in SPSS. Descriptive statistics for the three categories (Understanding Number, Non-Numerical Processes, and Computation and Knowledge) are summarized in Table 1. The means and standard deviations (% of correct response scores) for boys and girls, for monozygotic (MZ) and dizygotic (DZ) twins, and for same-sex and opposite-sex DZ twins were very similar.

An analysis of variance (ANOVA) was performed to assess the mean effects of sex and zygosity on mathematical ability in our sample. In the analysis, zygosity had 3 levels: MZ, DZ same-sex, and DZ opposite-sex. No significant effect of zygosity was found for any of the variables. The main effect of sex was significant for all categories, with boys on average performing better than girls. However, this effect was negligible, accounting for less than 1% of the variance. No significant sex by zygosity interactions were found. The three means (and standard deviations) were very similar, suggesting that the categories were of similar difficulty.

For subsequent analyses, the scores from each test were separately standardized using the means and standard deviations of the entire sample (after the medical and ethnic exclusions described in the Method section) so that each test had zero mean and unit variance for the total sample of 5,348 twins. The scores were corrected for age so that age does not contribute to twin resemblance, which is standard in analyses of twin data (McGue & Bouchard, 1984). Although all children were tested at 10 years of age, the results could be affected even by small differences in age at the time of testing at this important stage of development, which would inflate estimates of shared environment because members of a twin pair are of exactly the same age. For the analyses of individual differences, the scores were also corrected for sex differences. This was not done for the extremes analyses in order not to affect the representativeness of groups at low-ability cutoffs.

In addition to the exclusions described in the Method section, we also excluded from further analyses all pairs in which one or both twins had missing data; this was done separately for

each category. For the individual difference analyses, to avoid the possibility that our results were affected by very extreme scores, all pairs in which one or both twins scored 3 or more standard deviations below or above the mean were excluded from each category.

Genetic Analysis

The twin method, one of the major tools of quantitative genetic research, addresses the origins of individual differences by estimating the proportion of variance that can be attributed to genetic, shared environment, and non-shared environment factors (Plomin, DeFries, McClearn, & McGuffin, 2001). In the case of complex traits that are likely to be influenced by multiple factors, the genetic component of variance refers to the influence of alleles at all gene loci that affect the trait. The similarity between twins for any particular trait may be due to a common set of genes. It may also be due to the shared environment, which refers to environmental influences that vary in the population and that contribute to the similarity between co-twins. For example, twins experience similar conditions during gestation, have the same socioeconomic status, live in the same family, and usually go to the same school. These factors could reasonably be expected to increase similarity between co-twins. *Non-shared environment* refers to any aspect of environmental influence that makes co-twins different from each other, including measurement error. Such influences involve aspects of the environment that are specific to an individual, such as traumas and diseases, idiosyncratic experiences, different peers, differential treatment by the parents and teachers, and different perceptions of such influences.

Genetic influence can be estimated by comparing intraclass correlations for monozygotic (MZ) twins, who are genetically identical, and dizygotic (DZ) twins, whose genetic relatedness is on average .50. For this reason, the phenotypic variance of a trait is attributed to genetic variance (called heritability) to the extent that the MZ twin correlation exceeds the DZ twin correlation. The relatedness for shared (common) environmental influences is assumed to be 1.0 for both MZ and DZ twin pairs who grow up in the same family, because they experience similar prenatal and postnatal environments. Shared environmental influences are indicated to the extent that DZ twins' correlation is more than half of the MZ correlation.

Twin intraclass correlations for the three measures are shown in Table 2. The results for all categories are similar and are consistent with those previously reported for teacher-assessed global measures of mathematics in the same sample when the children were 7 years old (Kovas, Harlaar, Petrill, & Plomin, 2005; Oliver et al., 2004). The twin correlations of .51–.59 for MZ and .24–.42 for DZ twins suggest at least moderate genetic influences, with environmental factors being primarily of nonshared nature. Correlations for same-sex and opposite-sex DZ twins are similar, suggesting the absence of sex differences in genetic and environmental parameter estimates.

To assess whether being in the same class and having the same teacher increased similarity between co-twins and affected the genetic findings, we reran our correlational analyses, splitting the data by same versus different teachers. The two groups were equal in size. The correlations were highly similar for the two groups; for example, for the Understanding Numbers measure, the MZ correlations were .70 and .71 for same and different teacher, respectively; for DZ twins, the correlations were .45 and .49, respectively.

Model Fitting

Structural equation model fitting is a comprehensive way of estimating variance components based on the principles described earlier. In this study, structural equation modeling was used to examine (a) the genetic and environmental etiology of individual differences across the range of ability; (b) the genetic and environmental etiology of low mathematical performance

(extremes analyses); and (c) the question of whether the same or different factors influence individual differences in boys and girls across the whole range of ability and in the low-ability group.

Individual Difference Analyses

Variations of the so-called ACE model can be used for analyses of individual differences. The ACE model apportions the phenotypic variance into genetic (A), shared environmental (C), and non-shared environmental (E) components, assuming no effects of nonadditive genetics or nonrandom mating. The ACE parameters and their confidence intervals can be estimated by fitting the models to variance/covariance matrices using the model fitting program *Mx* (Neale, 1997).

There are three possibilities (see Table 3) with respect to the causes of individual differences in boys and girls, regardless of the mean differences between the sexes (Neale & Maes, 2003). The first possibility is that different genetic and environmental factors are responsible for individual differences in mathematics for boys and girls—these are called *qualitative* differences. Such sex-specific effects are not limited to genes on the X chromosome but can also involve genes on the autosomal chromosomes that affect boys and girls differently—for example, because the genes interact with sex hormones. The second possibility, which is not mutually exclusive with the first, is that the same etiological influences affect individual differences in boys and girls, but that they do so to a different extent—these are known as *quantitative* differences. The third possibility is that even if there are mean differences, there are no differences in the etiology of individual differences for boys and girls—the same genes and environments operate to the same extent in both sexes. That is, boys as a group may exhibit a disadvantage in mathematics, but the factors that make one boy different from another are the same as those that make one girl different from another girl. These three possibilities (qualitative differences, quantitative differences, and no differences) were assessed using sex-limited structural equation modeling.

Each possibility is associated with a set of parameters in the models. Qualitative differences are reflected in the genetic correlation (r_g) between DZ opposite-sex twins. In DZ same-sex pairs, the assumption is that on average the twins share 50% of their varying DNA, and the coefficient of genetic relatedness (i.e., the genetic correlation between the two children) is therefore .5. If there are qualitative differences in etiology between boys and girls (i.e., different genetic and environmental factors), the genetic correlation in DZ opposite-sex twins will be less than .5. If there are quantitative differences (i.e., the same factors, but exerting different magnitudes of effect) rather than qualitative differences, the genetic correlation for DZ opposite-sex pairs will be .5, but the parameter estimates for the A, C, and E components will be significantly different for male–male pairs and female–female pairs. If there are no differences between boys and girls, the DZ opposite-sex (DZ-os) pairs will have a genetic correlation of .5 and the A, C, and E estimates for male–male and female–female pairs will be the same, although the phenotypic variance might nonetheless differ for the two sexes because mean differences are often associated with variance differences (i.e., higher means have higher variances).

Using the model-fitting program *Mx* (Neale, 1997), we first tested the full model, which allows all parameters to vary: r_g in the DZ opposite-sex pairs; A, C, and E estimates; and variance estimates. This model was fit to variance/covariance matrices derived from the data. A series of nested models was then tested. The first nested model was a common-effects sex-limitation model, which fixes r_g to .5 in the DZ-os, but allows different A, C, E and variance estimates. The second nested model was a scalar-effects sex-limitation model, which constrains the r_g in the DZ-os, as well as the A, C, and E parameters, but allows differences in phenotypic variance between male and female twins by modeling the variance in one sex to be a scalar multiple of

the variance in the other sex. The third and final nested model tests the null hypothesis and constrains all the parameters to be equal for male and female twins.

For each model, the ACE parameters and their confidence intervals were estimated. The overall fit of each model was evaluated using three indices. The χ^2 statistic, where degrees of freedom equal the number of observed correlations minus the number of estimated parameters, indicates the fit of the full model and also tests the fit of nested models, with a lower value indicating better fit (with degrees of freedom equal to the difference in degrees of freedom between the full and nested models). The other two indices used were Akaike's information criterion ($AIC = \chi^2 - 2df$; Akaike, 1987) and the root mean square error of approximation (RMSEA), with lower values representing better fitting models.

The results of the model fitting are summarized in Table 4. All model comparisons favored the null model. In other words, fixing ACE parameters and variances to be the same for boys and girls and fixing the genetic correlation for DZ opposite-sex twin pairs to .5 resulted in nonsignificant changes in the fit of the model. This suggests that the quantity and quality of genetic and environmental effects are the same in boys and girls. For this reason, only the parameter estimates from the null model are presented in Table 4.

As suggested by the twin correlations in Table 2, the overall pattern of results suggests that all measured aspects of mathematics are moderately heritable (.32–.45). Shared environment has a much smaller influence on these abilities (.07–.23), although the effect is significant for two of the tests, as indicated by the 95% confidence intervals that do not include zero. Non-shared environmental influence is moderate for all measures (.42–.48), which is partly due to the inclusion of measurement error in this component. Although the heritability of the NonNumerical Processes test is somewhat lower (.32) than the heritability of the other two tests (.42 and .45), the overlapping 95% confidence intervals indicate that this difference is not significant. We obtained similar results (available from the authors) when a composite of the three measures was analyzed.

Extremes Analyses

Probands and Probandwise Concordances—For each of the measures, we defined probands as either 5% and 15% of the whole sample, identifying statistically low performance on that measure. Table 5 shows the means, ranges, and standard deviations (total number of correct items) for probands for each of the three measures for both proband cutoffs. Table 5 also refers to National Curriculum standardization information in relation to these scores. These data suggest that the probands in our sample (at both cutoffs) represent a group of children performing below their grade expectation and failing items that are solved correctly by the majority of much younger children.

Probandwise concordances (i.e., the ratio of the number of probands in concordant pairs to the total number of probands) were calculated separately for each measure and each of the five sex–zygosity groups. Probandwise concordances represent the risk that a co-twin of a proband is affected (Plomin et al., 2001). Table 6 shows that concordances for MZ twins are higher than for DZ twins, suggesting a genetic influence. The results are similar for boys and girls and for same-sex and opposite-sex DZ twins.

Liability–Threshold Analysis—Unlike twin correlations, twin concordances cannot be used to estimate genetic and environmental parameters, because they do not in themselves include information about the population incidence. The liability–threshold model, which is a natural extension of quantitative genetic models for quantitative traits, is widely used in genetics to analyze concordance data for dichotomous traits (Sham, 1998). The model assumes an underlying continuous liability that has a normal distribution, with a mean of 0 and a variance

of 1 in the general population. If the liability to a disorder is quantitative rather than categorical, the disorder is assumed to be present in all individuals whose liability is above a certain threshold value and to be absent in all other individuals. The value of the threshold can be estimated from the population frequency of the disorder. The liability is not measured directly but is estimated from the observed categorical data. For the purposes of this study, the data from the entire twin sample were organized into 2×2 contingency tables, where the cells represent pairs in which both twins are unaffected, both twins are probands, and two discordant cells where one twin or the other is a proband. These data can be used to quantify genetic and environmental sources of variation in liability in the population. In this study, a structural equation model was fit to the contingency tables by maximum likelihood, using the *Mx* program to estimate ACE parameters (Neale, 1997).

A full sex-limited liability–threshold model and a series of nested models were tested and compared; the full model, common-effects model, and scalar-effects model used were the same as those used in the individual difference modeling. Moreover, the null model in this case equated thresholds for male and female twins; the threshold corresponds to the proportion of affected individuals for the two sexes. Whether the same threshold could be fit for male and female twins was tested to determine whether rates of disorder differed for male and female twins.

Comparisons for all except one model favored the null-effects sex-limitation model that constrains genetic and environmental parameters and the thresholds of prevalence to be equal between the sexes. In other words, as was the case for the individual difference analyses, fixing ACE parameters to be the same for boys and girls and fixing the genetic correlation for DZ opposite-sex twin pairs to .5 resulted in nonsignificant changes in the fit, suggesting that the quantity and quality of genetic and environmental effects are the same in boys and girls. The prevalence of mathematical disability was also generally the same for boys and girls.

Table 7 summarizes the results of the best-fit models for the low extremes of the three measures. As expected from the twin concordances in Table 6, and similar to the individual difference results, the three measures show significant and substantial heritability, modest shared environmental influence that reaches significance in only one of the six analyses, and moderate but significant non-shared environmental influence. There is some suggestion that heritability may be greater at the very extreme end of the distribution (bottom 5% of our sample), although the sample sizes are very small for this extreme cutoff, resulting in very wide confidence intervals. The overlapping confidence intervals for heritability estimates at the 5% and 15% cutoffs indicate that the heritability estimates do not differ significantly.

DF Extremes Analysis—The liability–threshold model assumes a continuous dimension, even though it is based on dichotomous data (i.e., presence or absence of disability). The liability–threshold analysis converts dichotomous diagnostic data to poly-choric correlations (Smith, 1974) to analyze the hypothetical construct of a threshold with an underlying continuous liability. In contrast, *DF* extremes analysis assesses rather than assumes a continuum. If all of the assumptions of the liability–threshold model are correct for a particular disorder, it will yield results similar to *DF* extremes analysis to the extent that the quantitative dimension assessed is linked genetically to the qualitative disorder (Plomin, 1991).

DF extremes analysis assesses genetic links between disability and ability by bringing together dichotomous diagnoses of disability and quantitative traits of ability. Rather than assessing twin similarity in terms of individual differences on a quantitative trait of ability or in terms of concordance for a diagnostic cutoff, *DF* extremes analysis assesses twin similarity as the extent to which the mean standardized quantitative trait score of co-twins is as low as the mean standardized score of selected extreme or diagnosed probands (see Plomin & Kovas, 2005, for

a detailed explanation of *DF* extremes analysis and for discussion on contrasting the three methods). This measure of twin similarity is called a *group* twin correlation (or *transformed co-twin mean*) in *DF* extremes analysis, because it focuses on the mean quantitative trait score of co-twins rather than on individual differences. Genetic influence is implied if group twin correlations are greater for MZ than for DZ twins. Doubling the difference between MZ and DZ group twin correlations estimates the genetic contribution to the average phenotypic difference between the probands and the population. The ratio between this genetic estimate and the phenotypic difference between the probands and the population is called *group heritability*. It should be noted that group heritability does not refer to individual differences among the probands—the question is not why one proband has a slightly greater disability than another, but rather why the probands as a group have a much greater disability than the rest of the population.

Although *DF* extremes group heritability can be estimated by doubling the difference in MZ and DZ group twin correlations (Plomin, 1991), *DF* extremes analysis is more properly conducted using a regression model (DeFries & Fulker, 1988). The *DF* extremes model fits standardized scores for MZ and DZ twins to the regression equation, $C = B_1P + B_2R + A$, where C is the predicted score for the co-twin, P is the proband score, R is the coefficient of genetic relatedness (1.0 for MZ twins and .5 for DZ twins), and A is the regression constant. B_1 is the partial regression of the co-twin score on the proband—an index of average MZ and DZ twin resemblance independent of B_2 . The focus of *DF* extremes analysis is on B_2 . B_2 is the partial regression of the co-twin score on R independent of B_1 . It is equivalent to twice the difference between the means for MZ and DZ co-twins adjusted for differences between MZ and DZ probands. In other words, B_2 is the genetic contribution to the phenotypic mean difference between the probands and the population. Group heritability is estimated by dividing B_2 by the difference between the means for probands and the population.

Finding group heritability implies that disability and ability are both heritable and that there are genetic links between the disability and the normal variation in the ability. That is, group heritability itself, not the comparison between group heritability and the other estimates of heritability, indicates genetic links between disability and ability. If a measure of extremes (or a diagnosis) were not linked genetically to a quantitative trait, group heritability would be zero. For example, it is possible that a severe form of learning disability is due to a single-gene disorder that contributes little to the normal variation in learning ability. However, most researchers now believe that common disorders are caused by common genetic variants—the common disease/common variant hypothesis (Collins, Euyer, & Chakravarti, 1997)—rather than by a concatenation of rare single-gene disorders. To the extent that the same genes contribute to learning disability and normal variation in learning ability, group heritability can be observed, but the magnitude of group heritability depends on the individual heritability for normal variation and the heritability of disability gleaned from concordances for disability.

In this study, scores were standardized and transformed to adjust for proband mean differences between MZ and DZ groups, so that genetic and environmental parameters could be estimated from model fitting, on the basis of the regression, $C(M) = B_1P(M) + B_2R + A$, where $C(M)$, the co-twin's mathematics score, is predicted from $P(M)$, the proband's mathematics score, 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 twin similarity varies as a function of the degree of genetic relatedness (R).

The results from the *DF* extremes analysis are presented in Table 8. Group heritabilities for all measures are significant and substantial, suggesting a genetic link between ability and disability. Shared environmental estimates are modest and generally not significant. Similar

to the results from the liability–threshold analysis, there is again a hint of greater genetic influence for disability for two of the three measures, but again these differences are not significant.

Finally, Figure 1 summarizes the genetic results of the study, presenting point estimates and confidence intervals for heritability from analyses of individual differences in ability (“continuous ability”) and for disability using two cutoffs (5% and 15%) as analyzed by *DF* extremes analyses and by the liability–threshold model. It can be seen that all estimates are similar, although heritability is somewhat greater for the 5% cutoff for two of the measures.

Discussion

The first aim of our study was to investigate the pattern of genetic and environmental contributions to the typical range of individual differences for three aspects of mathematics. The results showed that the relative contributions of genes and environments to individual differences in diverse aspects of mathematics are similar. Genetic influence was moderate for all three measures. Consistent with previous research, we found that most environmental influences on diverse aspects of mathematics were nonshared, although this estimate also includes measurement error. This means that the environmental factors that shape individual differences in mathematics are mostly not shared between co-twins (even MZ twins) in the same family. In other words, having the same parents, the same socioeconomic status, and going to the same school does not contribute to the similarity between the two children beyond the similarity due to their shared genes. The analysis of the data by same versus different teachers showed no differences in correlations for the two groups. These findings suggest that being in the same class does not contribute to children’s similarity in mathematical performance. That is, classroom environments affect mathematical ability in different children (even twins and even MZ twins) in different ways; these effects are subsumed under the nonshared environmental estimate.

The second aim of the study was to investigate whether the etiology of mathematical difficulties is similar and is linked to that of typical performance in mathematics. Despite some apparent differences in the estimates between individual differences and extremes (see Tables 4 and 7), heritability estimates from the individual difference analyses never fell outside the 95% confidence intervals of the estimates from the analyses of extremes. Overall, it can be concluded that the ACE estimates for individual differences and extremes are similar, at least within the limits of the power provided by this largest yet twin study. That is, liability to disability in different aspects of mathematics and individual differences in mathematical abilities are influenced by moderate genetic influences and modest shared environmental influences. Moreover, group heritabilities from *DF* extremes analysis were substantial, which suggests strong links between genetic etiologies of mathematical difficulties and variation in mathematical ability in the typical range. There is some suggestion that genetic factors may be more important at the very low end of performance. This might mean that beyond a certain threshold of accumulated genetic risk factors, the influence of some otherwise relevant environmental factors is suppressed.

The third aim of the study was to investigate any possible sex differences in the etiology of typical and low mathematics performance. The results showed negligible mean sex differences for all measures, explaining less than 1% of the variance. In terms of etiology, no sex differences were found in either the whole range of ability or the low extremes. The results suggest that the same factors affect individual differences in boys and girls to the same extent (quantitative effects). Similarly, the same factors are responsible for low performance for both boys and girls for all three measures of mathematics (qualitative effects).

Similar results for the three aspects of mathematics assessed in this study and their intercorrelations (.62 on average) suggest that the same genetic and environmental factors might affect them. Previous multivariate genetic research into learning abilities and disabilities has suggested that the observed overlap within and between different abilities is largely due to common genes (Plomin & Kovas, 2005). Our analyses of the first half of the present dataset (Kovas et al., 2007) suggested that genetic effects are largely shared for the three measures. What makes them different are nonshared environments, specific to each aspect of mathematics.

The obvious limitation of this study is that no specific genes or environments were considered. A next step for genetic research is to find specific genetic markers associated with mathematical disability, which is now made more feasible using genome-wide association strategies that test hundreds of thousands of DNA markers simultaneously (Plomin, 2005). Our results suggest that when such genetic associations are identified, they will be associated with variation in the whole range of ability, not just with disability. Similarly, our results suggest that the same nonshared environmental factors will be associated with ability and disability. When specific genetic and environmental factors are identified, they will facilitate a direct test of our conclusion that disability is merely the quantitative extreme of a continuum of ability. We predict that associations will be found throughout the normal distribution, as much for high performance as for low performance.

Although the results of this study have no immediate implications for the ways in which mathematics is taught, we believe that they provide grounds for better conceptualizing several important issues. For example, our findings suggest that common mathematical disability is etiologically the low end of a quantitative continuum, rather than being driven by unique genetic and environmental factors that are not involved in individual differences. This finding has implications for defining mathematical disability as well as for research into factors that are responsible for low mathematical ability.

The finding that factors that make one boy different from another boy are the same as those that make one girl different from another girl is also important in that it suggests, for example, that the same remediation and prevention of mathematical difficulties will be successful for both sexes. Finally, another far-reaching implication is that environmental factors important for mathematical ability seem to operate mostly on an individual basis, interacting with a child's genetic makeup and other individual-specific environmental influences. It is too early to say whether this calls for a more individualized education or for the recognition that the same methods or teaching styles will not be optimal for all students. Moreover, though small, shared environmental influences were significant for some measures; therefore, research is also needed that attempts to identify the specific factors responsible for these influences. What this study demonstrates is that more genetically sensitive research is needed in this area, as it can make important contributions to the study of mathematical ability and disability, especially when the specific genes and nonshared environmental factors are identified.

References

- Akaike H. Factor analysis and AIC. *Psychometrika* 1987;52:317–332.
- Birnbaum MH. Human research and data collection via the Internet. *Annual Review of Psychology* 2004;55:803–832.
- Collins FS, Euyer MS, Chakravarti A. Variations on a theme: Cataloguing human DNA sequence variation. *Science* 1997;278:1580–1581. [PubMed: 9411782]
- DeFries JC, Fulker DW. Multiple regression analysis of twin data: Etiology of deviant scores versus individual differences. *Acta Geneticae Medicae et Gemellologicae* 1988;37:205–216.

- Freeman B, Smith N, Curtis C, Hockett L, Mill J, Craig IW. 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* 2003;33:67–72. [PubMed: 12645823]
- Good RH III, Salvia J. Curriculum bias in published, norm-referenced reading tests: Demonstrable effects. *School Psychology Review* 1988;17(1):51–60.
- Gosling SD, Vazire S, Srivastava S, John OP. Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *The American Psychologist* 2004;59:93–104. [PubMed: 14992636]
- Hyde JS. The gender similarities hypothesis. *The American Psychologist* 2005;60:581–592. [PubMed: 16173891]
- Kovas Y, Harlaar N, Petrill SA, Plomin R. “Generalist genes” and mathematics in 7-year-old twins. *Intelligence* 2005;33:473–489. [PubMed: 19319204]
- Kovas Y, Petrill SA, Plomin R. The origins of diverse domains of mathematics: Generalist genes but specialist environments. *Journal of Educational Psychology* 2007;99(1):128–139. [PubMed: 19756208]
- Kraut R, Olson J, Banaji M, Bruckman A, Cohen J, Couper M. Psychological research online: Report of the Board of Scientific Affairs’ Advisory Group on the conduct of research on the Internet. *The American Psychologist* 2004;59:105–117. [PubMed: 14992637]
- Maccoby, EE.; Jacklin, CN. *The psychology of sex differences*. Stanford, CA: Stanford University Press; 1974.
- McGue M, Bouchard TJ. Adjustment of twin data for the effects of age and sex. *Behavior Genetics* 1984;14:325–343. [PubMed: 6542356]
- Naglieri JA, Drasgow F, Schmit M, Handler L, Prifitera A, Margolis A, et al. Psychological testing on the Internet: New problems, old issues. *The American Psychologist* 2004;59:150–162. [PubMed: 15222858]
- Neale, MC. *Mx: Statistical modeling*. Richmond, VA: Department of Psychiatry; 1997.
- Neale, MC.; Maes, HHM. *Methodology for genetic studies of twins and families*. Dordrecht, the Netherlands: Kluwer; 2003.
- NFER-Nelson. *Maths 5–14 series*. London: Author; 1994.
- NFER-Nelson. *Maths 5–14 series*. London: Author; 1999.
- NFER-Nelson. *Maths 5–14 series*. London: Author; 2001.
- Oliver B, Harlaar N, Hayiou-Thomas ME, Kovas Y, Walker SO, Petrill SA, et al. A twin study of teacher-reported mathematics performance and low performance in 7-year-olds. *Journal of Educational Psychology* 2004;96:504–517.
- Pennington, BF.; Olson, RK. Genetics of dyslexia. In: Snowling, M.; Hulme, C., editors. *The science of reading: A handbook*. Oxford: Black-well; 2005. p. 453-472.
- Plomin, R. Genetic risk and psychosocial disorders: Links between the normal and abnormal. In: Rutter, M.; Casaer, P., editors. *Biological risk factors for psychosocial disorders*. Cambridge, UK: Cambridge University Press; 1991. p. 101-138.
- Plomin R. Finding genes in child psychology and psychiatry: When are we going to be there? *Journal of Child Psychology and Psychiatry and Allied Disciplines* 2005;46:1030–1038.
- Plomin, R.; DeFries, JC.; McClearn, GE.; McGuffin, P. *Behavioral genetics*. 4. New York: Worth; 2001.
- Plomin R, Kovas Y. Generalist genes and learning disabilities. *Psychological Bulletin* 2005;131:592–617. [PubMed: 16060804]
- Sham, P. *Statistics in human genetics*. London: Arnold; 1998.
- Smith C. Concordance in twins: Methods and interpretation. *American Journal of Human Genetics* 1974;26:454–466. [PubMed: 4858380]
- Spelke ES. Sex differences in intrinsic aptitude for mathemaics and science? A critical review. *The American Psychologist* 2005;60:950–958. [PubMed: 16366817]
- Trouton A, Spinath FM, Plomin R. Twins Early Development Study (TEDS): A multivariate longitudinal genetic investigation of language, cognition and behaviour problems in childhood. *Twin Research* 2002;5:444–448. [PubMed: 12537874]
- Wechsler, D. *Wechsler intelligence scale for children*. 3. London: Psychological Corp; 1992.

Biographies

Yulia Kovas, PhD, is a researcher at the Social, Genetic and Developmental Psychiatry Research Centre (SGDP), Institute of Psychiatry, London. Her current interests include the genetic and environmental etiology of individual differences in learning abilities and disabilities and the etiology of comorbidity between different learning disabilities.

Claire M. A. Haworth, BA, is a PhD student at the SGDP, Institute of Psychiatry. Her current interests include the genetic and environmental etiology of individual differences in academic abilities and disabilities and the etiology of change and continuity throughout childhood.

Stephen A. Petrill, PhD, is a professor of biobehavioral health and assistant director of the Center for Developmental and Health Genetics at the Pennsylvania State University. His interests involve understanding the genetic and environmental influences on the development of and covariance between early reading, math, and language skills.

Robert Plomin, PhD, is professor of behavioral genetics at the Institute of Psychiatry, where he is deputy director of the SGDP. His current interest is in harnessing the power of molecular genetics to identify genes for psychological traits.

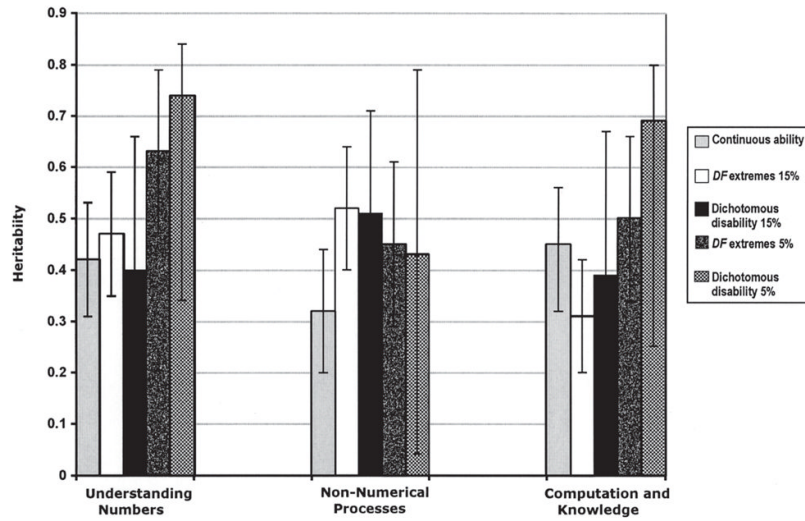


FIGURE 1. Heritability estimates for continuous ability, DF extremes analysis, and dichotomous diagnoses of disability at 5% and 15% cutoffs with corresponding confidence intervals for the three measures. The heritability estimate refers to group heritability for DF extremes analysis, liability heritability for dichotomous disability analysis, and individual difference heritability for continuous ability analysis.

Means and Standard Deviations for Percentage of Correct Items and ANOVA Results by Sex and Zygosity for All Three Mathematics Measures

TABLE 1

Measure	MZ ^a		DZ ^b		Male ^c		Female ^d		MZ-m ^e		DZ-m ^f		MZ-f ^g		DZ-f ^h		DZ-opp ⁱ		Sex		Zygosity		Sex × Zygosity		
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD	p	η^2	p	η^2	p	η^2	
Understanding Number	.76	.18	.77	.18	.78	.18	.75	.18	.78	.18	.79	.18	.75	.18	.75	.18	.77	.18	.000	.008	.000	.377	.001	.458	.001
Non-Numerical Processes	.69	.18	.70	.17	.71	.17	.68	.17	.70	.18	.72	.16	.68	.17	.68	.17	.70	.17	.009	.003	.009	.867	.000	.097	.002
Computation and Knowledge	.81	.19	.82	.19	.83	.19	.81	.19	.83	.18	.84	.19	.80	.20	.80	.20	.82	.19	.002	.003	.002	.794	.000	.243	.001

Note. Descriptive statistics are reported on the whole sample after medical exclusions, $N = 5,348$. MZ = monozygotic twins; DZ = dizygotic twins (same and opposite sex); MZ-m = monozygotic male twins; DZ-m = dizygotic male twins; MZ-f = monozygotic female twins; DZ-f = dizygotic female twins; DZ-opp = dizygotic opposite-sex twins; η^2 = proportion of the total variance attributed to an effect.

^a $n = 1,941$.

^b $n = 3,407$.

^c $n = 2,413$.

^d $n = 2,935$.

^e $n = 803$.

^f $n = 777$.

^g $n = 1,138$.

^h $n = 926$.

ⁱ $n = 1,704$.

TABLE 2
 Intraclass Correlations and Numbers of Pairs by Sex and Zygosity for All Three Mathematics Measures

Measure	MZ ^a		DZ ^b		MZ-m ^c		DZ-m ^d		MZ-f ^e		DZ-f ^f		DZ-opp ^g	
	r	n	r	n	r	n	r	n	r	n	r	n	r	n
Understanding Numbers	.58*	885	.37*	1,559	.57*	362	.35*	352	.59*	523	.42*	433	.35*	774
Non-Numerical Processes	.56*	873	.39*	1,553	.57*	358	.39*	354	.54*	515	.39*	430	.39*	769
Computation and Knowledge	.51*	888	.29*	1,557	.51*	362	.32*	355	.52*	526	.36*	425	.24*	777

Note. Scores were standardized on the whole sample (after medical exclusions). All correlations are based on age- and sex-corrected scores. Numbers of pairs are after the exclusion of pairs in which at least one twin was an outlier or had missing data separately for each category. MZ = monozygotic twins; DZ = dizygotic twins (same and opposite sex); MZ-m = monozygotic male twins; DZ-m = dizygotic male twins; MZ-f = monozygotic female twins; DZ-f = dizygotic female twins; DZ-opp = dizygotic opposite-sex twins.

^a n = 1,941.

^b n = 3,407.

^c n = 803.

^d n = 777.

^e n = 1,138.

^f n = 926.

^g n = 1,704.

* p < .01.

TABLE 3

Overview of Possible Types of Individual Differences in Boys and Girls Regardless of Mean Differences Between the Sexes

Type of difference	Explanation	Possible contributing factors
Qualitative	Different genetic and environmental factors are responsible for individual differences in mathematics for boys and girls	<ul style="list-style-type: none"> • Genes on the sex chromosomes • Genes on the autosomal chromosomes affecting boys and girls differently, for example, because the genes interact with sex hormones • Teachers treating boys and girls differently in terms of their expectations or additional help
Quantitative	The same etiological influences affect individual differences in boys and girls, but that they do so to a different extent.	As above, but the differences are in the quantity of effects.
No difference	The same genes and environments operate to the same extent in both sexes.	Boys as a group may exhibit a disadvantage in mathematics, but the factors that make one boy different from another are the same as those that make one girl different from another girl.

Individual Difference Analyses: Indices of Fit, Parameter Estimates, and Confidence Intervals of Best-Fit Model for All Three Measures

TABLE 4

Measure	Fit indices													
	$\chi^2(13)$	<i>p</i>	AIC	RMSEA	PE	CI	A			C		E		
Understanding Numbers	6.66	.919	-19.344	.000	.42	.31-.53	.16	.07-.25	.42	.38-.46				
Non-Numerical Processes	9.90	.702	-16.103	.000	.32	.20-.44	.23	.14-.32	.45	.40-.50				
Computation and Knowledge	13.63	.401	-12.374	.006	.45	.32-.56	.07	.00-.17	.48	.44-.53				

Note. The best-fit model (ACE model) did not allow for sex differences—that is, male and female variance component estimates were constrained to be equal (quantitative differences removed) and both genetic correlations and shared environment correlations were fixed to .5 and 1, respectively (qualitative differences removed). AIC = Akaike's information criterion (Akaike, 1987); RMSEA = root mean square error of approximation; A = additive genetic influence; C = shared environmental influence; E = unshared environmental influence; PE = parameter point estimate; CI = confidence interval.

TABLE 5

Means, Standard Deviations, and Ranges of Proband Scores at Two Different Cutoffs for All Three Measures

Measure	Range	<i>M</i>	<i>SD</i>	National Curriculum level and standardization information
Understanding Numbers^a				
5% cutoff	0–11	8.65	3.10	Fail all age-appropriate items. Children with lowest scores are failing items from Levels 1 and 2, recommended for children ages 6, 7, 8, and 9 years. None of the children get beyond easiest items at Level 4, which are successfully solved by more than 50% of 8- and 9-year-olds.
15% cutoff	0–15	12.95	3.15	Most children are failing items that are successfully solved by more than 50% of 8- and 9-year-olds. None of the children get beyond easiest items at Level 4, which are successfully solved by more than 50% of 10-year-olds.
Non-Numerical Processes^b				
5% cutoff	0–7	5.32	1.99	Children with lowest scores are failing items from Levels 1 and 2, recommended for children ages 6, 7, 8, and 9 years. None of the children get beyond easiest items at Level 4, which are successfully solved by more than 50% of 8- and 9-year-olds.
15% cutoff	0–10	7.60	2.11	Children with lowest scores are failing items from Levels 1 and 2, recommended for children ages 6, 7, 8, and 9 years.
Computation and Knowledge^c				
5% cutoff	0–13	10.19	3.53	Children with lowest scores are failing items from Levels 1 and 2, recommended for children ages 6, 7, 8, and 9 years. None of the children get beyond easiest items at Level 4, which are successfully solved by more than 50% of 8- and 9-year-olds.
15% cutoff	0–18	13.11	3.06	Children with lowest scores are failing items from Levels 1 and 2, recommended for children ages 6, 7, 8, and 9 years. None of the children get beyond easiest items at Level 4, which are successfully solved by more than 50% of 9- and 10-year-olds.

Note. National Curriculum uses 7 levels (Level 8 = exceptional performance). Children at Key Stage 2 (7- to 11-year-olds) are expected to work with Levels 2–5, and most children are expected to attain Level 4 by the end of the year during which we tested our sample. Within each level, items can be organized by difficulty. We used easy and hard sublevels within each level to identify items that were successfully solved by less or more than 50% of children within that school year. For each item, we had 3 difficulty criteria: National Curriculum level, age (school year) at which an item is presented, and difficulty of the item according to the standardization sample percentage correct for that age (school year; reported in the Group Record Sheets, NFER-Nelson, 1994, 1999, 2001).

^a27 items in category.

^b19 items in category.

^c31 items in category.

Probandwise Twin Concordances and Number of Probands for Low and High Mathematical Ability at Two Cutoffs by Sex and Zygosity for All Three Measures

TABLE 6

Measure	<u>MZ-m^a</u>		<u>DZ-m^b</u>		<u>MZ-f^c</u>		<u>DZ-f^d</u>		<u>DZ-opp^e</u>	
	%	n	%	n	%	n	%	n	%	n
Understanding Numbers										
5% cutoff	35	34	14	28	49	61	19	53	17	72
15% cutoff	53	102	36	88	46	170	38	146	33	234
Non-Numerical Processes										
5% cutoff	26	47	8	24	56	57	32	50	20	69
15% cutoff	50	119	24	83	46	169	31	147	32	220
Computation and Knowledge										
5% cutoff	45	31	22	36	35	62	26	47	6	70
15% cutoff	41	107	35	98	47	170	40	149	25	212

Note. All extremes analyses were done on age- but not sex-regressed scores. MZ-m = monozygotic male twins; DZ-m = dizygotic male twins; MZ-f = monozygotic female twins; DZ-f = dizygotic female twins; DZ-opp = dizygotic opposite-sex twins.

^a n = 803.

^b n = 777.

^c n = 1,138.

^d n = 926.

^e n = 1,704.

Liability-Threshold Model: Indices of Fit, Parameter Estimates, and Confidence Intervals for All Three Measures at Two Cutoffs

TABLE 7

Measure	χ^2	df	P	Fit indices			A			C			E				
				AIC	RMSEA	PE	CI	PE	CI	PE	CI	PE	CI				
Understanding Numbers																	
5% cutoff	6.27	13	.936	-19.735	.000	.74	.34-.84	.00	.00-.33	.26	.16-.39						
15% cutoff	7.58	12 ^a	.817	-16.418	.000	.40	.13-.66	.25	.04-.45	.35	.26-.45						
Non-Numerical Processes																	
5% cutoff	20.75	13	.078	-5.251	.012	.43	.04-.79	.28	.00-.59	.29	.19-.43						
15% cutoff	15.32	13	.288	-10.685	.009	.51	.23-.71	.12	.00-.34	.38	.29-.48						
Computation and Knowledge																	
5% cutoff	11.58	13	.562	-14.416	.000	.69	.25-.80	.00	.00-.35	.31	.20-.47						
15% cutoff	17.39	13	.182	-8.615	.005	.39	.11-.67	.20	.00-.41	.41	.32-.52						

Note. In the best-fit model, male and female parameter estimates were constrained to be equal (quantitative differences removed) and both genetic correlation and shared environment correlations were fixed to .5 and 1, respectively (qualitative differences removed). Thresholds were also equated, indicating no sex differences in prevalence for most measures. AIC = Akaike's information criterion (Akaike, 1987); RMSEA = root mean square error of approximation; A = additive genetic influence; C = shared environmental influence; E = unshared environmental influence; PE = parameter point estimate; CI = confidence interval.

^aFor the 15% cutoff for the Understanding Numbers category, the thresholds could not be equated for male and female probands, indicating that this group contained significantly more girls.

TABLE 8

DF Extremes Analysis: Probandwise Twin Concordances, Twin Group Correlations, and Parameter Estimates at Two Cutoffs

Measure	Probandwise concordance				Twin group correlation			Parameter estimates		
	MZ		DZ-ss		MZ	DZ-ss	h^2_g	SE	c^2_g	SE
	%	n	%	n						
Understanding Numbers										
5% cutoff	44	95	17	81	.63	.32	.63	.16	.00	.13
15% cutoff	49	272	38	234	.60	.37	.47	.12	.13	.10
Non-Numerical Processes										
5% cutoff	42	104	24	74	.60	.38	.45	.16	.15	.14
15% cutoff	48	288	29	230	.59	.33	.52	.12	.07	.1
Computation and Knowledge										
5% cutoff	39	93	24	83	.59	.34	.50	.16	.09	.13
15% cutoff	45	277	38	247	.48	.33	.31	.11	.17	.09

Note. Only the same-sex pairs of twins were included in these analyses. Combining the same-sex and opposite-sex DZ twins did not change the results (data available from the first author). MZ = monozygotic twins; DZ-ss = dizygotic same-sex twins; h^2_g = group heritability; c^2_g = group shared environment.