Published in final edited form as:

J Child Psychol Psychiatry. 2009 September ; 50(9): 1113–1120. doi:10.1111/j.1469-7610.2009.02087.x.

Sex Differences and Science: The Etiology of Science Excellence

Claire M.A. Haworth¹, Philip S. Dale², and Robert Plomin¹

¹Social, Genetic and Developmental Psychiatry Centre, Institute of Psychiatry, King's College London, UK

²Department of Speech and Hearing Sciences, University of New Mexico, Albuquerque, USA

Abstract

Background—Are there sex differences in the etiology of high performance in science in childhood that could contribute to the under-representation of women in scientific careers? In this study the relative contributions of genetic and environmental influences on high performance in science in both boys and girls were assessed using standard twin analyses.

Methods—The sample included 3000 twin pairs from the UK Twins Early Development Study (TEDS). Science performance ratings based on the UK National Curriculum were collected from teachers when the twins were 9, 10 and 12 years old. Science excellence was defined as performing above the 85th percentile. Sex-limitation liability threshold models were used to assess sex differences in etiology.

Results—We found no evidence for quantitative or qualitative sex differences in the etiology of science excellence, which was moderately heritable (30–50%), and influenced by both shared (40–56%) and non-shared (10–13%) environmental influences.

Conclusions—Although boys and girls do not differ genetically in relation to school science performance per se, the under-representation of women in scientific careers may be due to attitudes rather than aptitudes.

Keywords

Science excellence; Sex differences; Behavior genetics

The domain of high abilities is one of the last areas of research to realize that the nature-nurture controversy is well and truly over. Research in this area has been polarized by the ideas of 'innate' ability versus deliberate and extensive practice, and few research programs have considered that *both* nature *and* nurture are likely to be important for the development of exceptional ability. Although the strengths and weakness of both extreme positions have been extensively discussed (e.g., Ericsson, Roring, and Nandagopal, 2007; Howe, Davidson, and Sloboda, 1998), there have been only a few studies that have empirically addressed both

Address for Correspondence: Claire M.A. Haworth, SGDP Centre P080, Institute of Psychiatry, King's College London, De Crespigny Park, London SE5 8AF, UK. Tel: +44(0)2078485621, Fax: +44(0)2078480866, Claire.Haworth@iop.kcl.ac.uk.

Conflict of Interest: The authors declare no conflict of interest.

the genetic and environmental contributions to exceptional performance, and even fewer have considered exceptional performance in cognitive abilities, focusing rather on artistic and sporting abilities.

Could there be genetic and environmental factors that differentially influence the performance of males and females in science, especially at the high end? Are men just intrinsically better at science than women? We are not aware of any empirical evidence regarding sex differences in the genetic and environmental influences on scientific excellence. A large body of literature on cognitive sex differences exists, with increasing support for the Gender Similarities Hypothesis (Hyde, 2005; Hyde and Linn, 2006), which based on meta-analyses of gender differences studies, states that males and females are much more alike than is generally portrayed. In fact, effect sizes for the influence of gender on cognitive variables are close to zero. A recent review has claimed that the under-representation of women in scientific careers is not due to sex differences in aptitude (Spelke, 2005), suggesting that other non-cognitive factors must explain the under-representation. Although mean differences in aptitude do not appear to fully explain the under-representation of women in scientific careers (Halpern et al., 2007), it is possible that underlying etiology of aptitude differs for boys and girls and this contributes to the under-representation of women. For example, if performance were more genetically driven for males than females this might contribute to males persisting in pursuing scientific careers.

The purpose of our study is to assess whether there are any sex differences in the genetic and environmental influences on science excellence during the school years. Using the twin method we can assess *what* causes differences in performance, and whether the same causes impact on men and women equally. The causes behind differences in means and individual differences (variance) are not necessarily the same. Even if males and females show no mean difference in performance, there could still be differences in the relative contributions of genes and environments to the variance in boys and girls, as well as different genes and environments impacting on boys and girls.

Our interest here is in the variance of science performance, and whether the variance in boys and girls is explained to a similar extent by the same genetic and environmental influences.

Etiology of Typical Science Performance

We conducted the first twin study of science performance when the twins in our Twins Early Development Study (TEDS) were 9 years old. Results indicated that genetic influences accounted for more than 60% of the variance, and shared environmental influences – that is environmental influences that make members of a twin pair similar – accounted for 14% of the variance (Haworth, Dale, and Plomin, 2008). In contrast to the findings for many other abilities, the heritability of science actually decreases from 9 years to 10 and 12 years (from 60% at 9 years to 49% and 44%) and the shared environment makes a greater contribution to the variance (from 14% to 27% and 33%) (Haworth, Kovas, Dale, and Plomin, 2008; Haworth, Dale, and Plomin, in press). Our findings for science performance at 12 years are consistent with results from the only other twin study looking at science-related phenotypes. Loehlin and Nichols (1976) studied the etiology of reading comprehension of passages

involving social and natural science in a sample of bright high school students. Heritability was estimated as 40%, and shared environmental influences were of a similar magnitude (30%) to those found for the 12-year-olds in our study. These results suggest that the shared environment may be particularly important for science performance in adolescence.

We reported the first sex-limitation model-fitting analyses of science performance at 9 and 10 years in TEDS (Haworth et al., 2008), where we tested the significance of qualitative and quantitative sex differences. Quantitative sex differences (e.g., differences in the level of heritability), can be tested by estimating genetic and environmental influences for males and females separately and then equating these estimates and assessing relative fit of these models to the data. Qualitative sex differences would indicate different genes or different environments impact males and females. Opposite-sex fraternal twins provide a good test of these differences, because if different genes and environments are important for males and females, the within-pair similarity of opposite-sex pairs will be reduced compared to same-sex fraternal pairs. Again, structural equation models allow the assessment of the best-fitting model to explain the data. At 9 and 10 years we found no evidence for quantitative or qualitative differences between boys and girls (Haworth et al., 2008; Kovas et al., 2007). However, it is possible that different results could be found at the high extreme of performance, particularly as males are typically more variable than females. Different results could also be found at later ages. The current paper focuses on the presence of sex differences in the etiology of science excellence and at 12 years in addition to 9 and 10 years.

Etiology of Science Excellence

Proponents of the extensive and deliberate practice theory might predict that environmental influences would be more important at the high extreme, because only those in the most stimulating environments would maintain the high level of practice and commitment needed for high performance. Those that support the idea of ‘innate’ talent might suggest that high ability would be more heritable than typical performance. Of course this is not a nature versus nurture scenario; both are almost certainly important. Our question is whether the relative contributions of genetic and environmental influences on science excellence are similar for males and females (quantitative differences), and also whether these are the same kinds of influence in both sexes (qualitative differences). Our previous analysis that found no quantitative or qualitative sex differences at 9 and 10 years for the entire distribution leads us to predict that we will also find no etiological differences for science excellence, even at the later assessment at 12 years.

Methods

Sample

The sampling frame was the Twins Early Development Study (TEDS), a study of twins born in England and Wales between 1994–96 (Oliver and Plomin, 2007). TEDS is reasonably representative of the general population in terms of parental education, ethnicity and employment status (Kovas et al., 2007). Zygosity was assessed through a parent questionnaire of physical similarity, which is over 95% accurate when compared to DNA

testing (Price et al., 2000). For cases where zygosity was unclear, DNA testing was conducted.

At 9 and 10 years, TEDS families from the 1994 and 1995 cohorts were invited to participate in the teacher assessments. Of the teacher rating forms sent at 9 years, 5836 forms were returned complete (76% of those contacted). At 10 years 6129 forms (79%) were returned complete. At age 12, twins born in 1994, 1995 and 1996 were invited to participate, and 9905 forms were returned complete (78%). For each assessment we obtained informed consent from the parents of the twins. The mean age at each assessment was 9.04 (sd=.29); 10.11 (sd=.29); and 11.54 (sd=.66). Not all teachers provided National Curriculum levels, and for some twin pairs we only received a completed form for one member of the pair. In Table 2 we present the number of complete twin pairs split by sex and zygosity at each age.

Measures

National Curriculum Teacher Reports—As for all children in the U.K., the twins' academic performance was assessed throughout the school year by their teachers, using the assessment materials of the National Curriculum for England and Wales (NC). Teachers were contacted in the second half of the school year so that they would be familiar with the children's performance. For the 9- and 10-year studies, the NC Teacher Assessments at Key Stage 2 were used, which are designed for children age 8 through their sixth year of primary school at age 11. Teachers assessed three broad areas of science ability: Scientific inquiry, Life processes and living things, and Physical processes. For the 12-year study, the NC Teacher Assessments at Key Stage 3 were used, which are designed for children aged 11–14 years old. At 12 years, teachers assessed four broad areas of science ability: Scientific inquiry, Life processes and living things, Physical processes, and Materials and their properties. Teachers provide a rating from 1–8 to indicate performance. By the end of Key Stage 2 (age 11), children are expected to reach level 4, and by the end of Key Stage 3 at age 14, children are expected to reach levels 5/6. Note that these judgments were not made specifically for the present study, but rather form the continuing assessment of each child that ultimately leads to the final NC Teacher Assessment score submitted at the end of the Key Stage to indicate the child's academic achievement during that period. At 9 and 10 years, 63% and 54% of the twins were rated by the same teacher. We do not have this information at 12 years. However, we have shown that the effect of same teacher versus different teachers has little effect on our ACE estimates (Kovas et al., 2007). The science scales are highly correlated, with an average inter-correlation of .85 at 9 years, .87 at 10 years, and .93 at 12 years. A factor analysis indicated that the first principal component accounted for 90% of the variance at 9 years, and 91% and 95% of the variance at 10 and 12 years. We therefore present results of a composite science score calculated by taking a mean of the three scores. Further information about the National Curriculum is available at <http://curriculum.qca.org.uk>.

Analyses

The twin method allows researchers to investigate the genetic and environmental influences on individual differences (variance) in observed traits. By comparing the twin intra-class correlations it is possible to estimate genetic effects, shared environmental effects (which

makes twins growing up in the same family more similar) and non-shared environmental effects (which do not contribute to familial resemblance). These three effects are typically labeled A, C and E respectively. A more comprehensive way of estimating the ACE parameters is maximum likelihood model fitting analysis (Neale, Boker, Xie, and Maes, 2006; Rijdsdijk et al., 2002), which makes assumptions explicit, tests the fit of the entire model to the data, tests the relative fit of alternative models, and provides confidence intervals for the parameter estimates. Discussion of the use of maximum likelihood model fitting analyses can be found elsewhere (Boomsma, Busjahn, and Peltonen, 2002; Martin et al., 1997; Plomin et al., 2008; Rijdsdijk et al., 2002). Mx software for structural equation modeling was used for standard analyses using raw data (Neale et al., 2006). In the present analysis, high ability was treated as a separate category. Individuals in the top 15% of the distribution were selected as high performers. We selected a 15% cut-off because this provided a reasonable balance between extreme scores and sample size; we would have limited power at more extreme cut-offs. For example to have 80% power to detect 50% heritability for a trait with 10% prevalence one would need 3659 pairs (Neale, Eaves, & Kendler, 1994). Our sample of 3000 pairs provides us with at least 80% power to detect heritability estimates of 40–50% for a prevalence of 15%. The standard liability threshold model was applied, which uses categorical data and concordance rates to assess the relative contributions of genetic and environmental influences (Falconer, 1965; Smith, 1974). Liability threshold modeling compares the concordance for a trait across families to that within families for different groups (in this case high performance and typical performance). Liability threshold modeling is the categorical equivalent of continuous twin model-fitting analyses, and it allows the assessment of sex differences in the etiology of categorical traits.

A full model that allowed qualitative, quantitative and prevalence differences between the sexes was compared to subsequent sub-models that modeled fewer parameters. The fits of the models are compared using a likelihood ratio test. A change in chi square of 3.84 is significant for a 1 degree of freedom test. We also report an additional fit statistic, Akaike Information criterion (Akaike, 1987), with lower values indicating better fit.

Results

Descriptive statistics

Whole sample—The means and standard deviations (*SD*) at 9, 10 and 12 years of one randomly selected member of each twin pair are presented in Table 1. The results of the *t*-tests, also shown in Table 1, indicate no significant effects of sex on any of the teacher report measures. This was despite the large sample size, which provides 90% power to detect mean differences as small as 0.1 *SD* (i.e., $d=.1$) that account for just 0.2% of the variance (Cohen, 1988).

High-performing subsample—At 9 years the mean standardized score of the high performance group was 1.71 (minimum score=1.10); at 10 years 1.34 (minimum=0.98); and at 12 years 1.54 (minimum=0.88). In terms of National Curriculum levels, the high performance group is on average one whole NC level above the typically performing group. At 9 years the high performance group have an average NC level of 3.97 (typical

group=2.81); at 10 years 4.21 (typical=3.22); and at 12 years 5.76 (typical=4.22). The high performance group at 10 years is at the same level as the typical performance at 12 years. Children are expected to reach level 4 at the age of 11 -- our high performers reach this at age 9; and level 5/6 at age 14 – which our high performers achieve by age 12. By these standards, our high performance group does represent a group of children who are excelling in science education.

Even though there are no sex differences for the entire sample, sex differences could emerge at the high extreme of the distribution. The high performance group was selected based on an 85% cut-off. This group scores roughly one standard deviation above the typical performers (below 85th percentile). At 9 years this group consisted of 808 individuals (388 female, 420 male); at 10 years 848 individuals (450 female, 398 male); and at 12 years 1143 individuals (594 female, 549 male). These prevalence differences, in favor of males at 9 years and in favor of females at 10 and 12 years, will be tested formally in the liability threshold model.

Tetrachoric twin correlations

The present paper focuses on quantitative and qualitative sex differences in the genetic and environmental origins of individual differences regardless of mean phenotypic sex differences.

For both males and females, tetrachoric twin correlations (Table 2) were higher in MZ than DZ pairs indicating genetic influence. Most important for the present paper is the pattern of MZ and DZ same-sex correlations for males and females. Across the three ages, the average MZ and DZ twin correlations, respectively, are .89 and .69 for males and .89 and .75 for females. Doubling the difference in these MZ and DZ correlations suggest heritabilities of 40% for males and 28% for females; shared environment is estimated as 49% for males and 61% for females. Although one might be tempted to conclude that females show less genetic influence and more shared environmental influence, the overlapping confidence intervals for the twin correlations (Table 2) indicate that these quantitative sex differences are not significant. Similarly, the average opposite-sex DZ correlation across the three ages is .64 which is lower than the average same-sex correlation (.72), suggesting qualitative sex differences. However, again the overlapping confidence intervals suggest that these qualitative sex differences are not significant. Both of these interpretations are confirmed by our model-fitting analyses which formally test the significance of quantitative and qualitative sex differences in the liability threshold model, presented in the following section.

Liability Threshold Model Fitting

Sex-limitation model fitting was first used to assess the significance of differences in prevalence between the sexes (Table 3; “Equate Prevalence” models). At 9 years there was a significant difference in prevalence between males and females, with more males present in the high performance group. At 10 and 12 years it was possible to equate the prevalence in males and females without significantly worsening the model fit, indicating no significant sex difference.

The second stage of the analyses was to assess the significance of qualitative sex differences, which would indicate different genetic/environmental influences for males and females. Qualitative sex differences are tested by equating the genetic/environmental correlation between opposite-sex dizygotic twins to that for same-sex dizygotic twins (0.5 and 1 respectively). There was no evidence for significant qualitative sex differences at any age (Table 3; “Fix r_{GO}/r_{CO} ” models).

The third stage of the analyses tests for the significance of quantitative sex differences – that is, different levels of effect in males and females such as different heritabilities. This is tested by equating the male and female parameters. There were no significant quantitative sex differences at any age (Table 3; “Equate male and female ACE” models). Finally the significance of the individual genetic and shared environmental parameters was tested, indicating that all are significant. The parameter estimates and 95% confidence intervals from the best-fitting model are shown in Table 4. These results confirm the results gleaned from the twin correlations in Table 2: moderate heritability (30–45%) and moderate shared environmental influence (41–56%). The 95% confidence intervals (in parentheses in Table 4), indicate that the trend towards decreasing heritability and increasing shared environmental influence is not nearly statistically significant.

Discussion

This is the first study to formally assess the genetic and environmental etiology of the liability to high performance in science. Moreover it is the first to test whether there are sex differences in etiology that could begin to explain the under-representation of women in scientific careers. Our results indicate that both genetic and shared environmental influences are important for high science performance, similar to results seen across the whole distribution of scores (Haworth et al., in press). But the focus of this paper is on the possible presence of sex differences in etiology. Using a twin design that includes both same-sex male and female pairs and opposite-sex pairs allowed us to assess three types of sex difference: different prevalence, different kinds of genes/environments, and different levels of effect.

At 10 and 12 years we were able to equate the thresholds for males and females without worsening the fit of the model to the data, indicating roughly equal proportions of males and females in the high ability groups. At 9 years it was not possible to equate the thresholds; the values indicated a significantly lower threshold for males, which translates to significantly more males in the high-ability group. However, the difference in thresholds is small, and this finding is not replicated at 10 and 12 years.

Even though there are no consistent prevalence differences there could still be differences in etiology between the sexes for high performance. We found no evidence for significant qualitative sex differences in etiology at any age – our opposite-sex dizygotic twins could be equated with our same-sex dizygotic pairs. This result is consistent with the hypothesis that the same genes and environments are important for high performance in boys and girls. Thus, we would expect that when specific genes associated with science performance are identified they will be associated with performance in both sexes. Finally, with respect to

quantitative differences, we were able to equate the genetic and environmental parameters for males and females, indicating that these influence science performance to the same extent in males and females at all three ages. The prediction for molecular genetic research would be that specific genes associated with science will have similar effect sizes in males and females.

It is important to remember that heritability describes ‘what is’ in a population – the genetic and environmental provenances of measured differences among individuals as they exist in a particular population with its particular mix of genetic and environmental influences. Heritability does not predict ‘what could be’, nor does it prescribe ‘what should be’. Furthermore, heritability denotes probabilistic genetic influence for a population, not predeterminism or immutability for an individual. If environments, say for education, were the same for all individuals we would still not expect that every child would achieve the same level of performance, and that is because genetic influences are also important.

In research into high ability we should consider that genes and environment do not just directly affect ability (aptitude); they can also have substantial influences on motivation (appetite and attitudes). Evidence suggests that there is dwindling enthusiasm for school science in early adolescence (Jenkins and Nelson, 2005; Osborne, Simon, and Collins, 2003), an effect that is more marked for females than males (Miller, Slawinski Blessing, and Schwartz, 2006; Osborne et al., 2003; Speering and Rennie, 1996). Eccles’ (1983) expectancy value model considered both aptitudes and personal beliefs to be important in predicting achievement-related choices and in explaining sex differences in course choice and performance. More recent analyses have shown that motivation predicts school subject choice (and therefore available career choices) beyond that predicted by grades (Simpkins, vis-Kean, and Eccles, 2006). An analysis of the effects of attitudes towards science is urgently needed, given the enduring lack of recruitment of women into the physical sciences and engineering, with the goal of identifying the antecedents of any sex differences in course and career selection. Moreover, the etiological links between attitudes and performance in science have not yet been investigated. We aim to investigate this issue in the crucial adolescent period when individuals make important course and career choices.

In addition, unlike other academic subjects such as reading and mathematics, we find that shared environmental influences are increasingly important with age. The obvious candidates for this increase are school experiences, which may become more varied as children progress into secondary education. Research is needed to identify these environmental influences; studying high achievers could aid in the identification of these environments.

Limitations

We cannot rule out the possibility that extreme performance is influenced by *different* genetic and environmental factors responsible for the normal distribution of variation in science performance. Even if the magnitude of genetic and environmental influence is the same for high ability and typical ability, completely different genetic and environmental factors could be responsible for the genetic and environmental influence. One test, DeFries-Fulker extremes analysis (DeFries and Fulker, 1988), goes some way towards testing the

links between the etiology of extreme and typical performance (Plomin and Kovas, 2005). Results from such analyses (available from the first author) provide an indication that the genetic and environmental influences on extreme performance are largely the same as those that influence typical performance. Moreover, our DeFries-Fulker extremes estimates of group heritability (0.52, 0.48 and 0.26 at 9, 10 and 12 years) are similar to the heritability estimates from the liability-threshold models presented here (0.49, 0.45 and 0.30). However, the definitive test of the overlap between typical and high performance will be possible when specific genetic and environmental influences are identified. In addition, different results may be found for more extreme levels of performance. Lykken's emergence model (Lykken, McGue, Tellegen, and Bouchard, Jr., 1992) suggests that very high performance is likely to be due to the epistatic interaction of many genes. This would be indicated by high resemblance in MZ twins and little or no resemblance in DZ twins – a pattern which is not seen in our data for the high-ability group. Nevertheless, we cannot rule out the role of epistatic genetic influence on very high performance. Different results may also be found for different science curricula. The UK National Curriculum provides an excellent opportunity for assessing performance on the same curriculum, but these results may not generalize across curricula. There are currently some excellent international assessments of science performance (e.g., PISA), but these studies are not genetically-sensitive in design and cannot be used to assess sex differences in etiology. Finally, categorical analyses are less powerful than continuous analyses, particularly when investigating sex differences, which requires splitting the sample still further. However, our categorical results for the variance components and the tests of sex differences are highly similar to those from the much more powerful continuous analyses. It is difficult to prove the null hypothesis that there are no sex differences, even with a large sample like TEDS. Replication is certainly needed, as is the assessment of etiological sex differences later in development.

Conclusions

We find no evidence for sex differences in the genetic and environmental etiology of science excellence. Not only do males and females perform equally well in school science, the same genes and environments are involved in both sexes. Differences in etiology could appear later in development, but it is more likely that the under-representation of women in scientific careers depends more on attitude than aptitude. Differences in ability are not keeping women out of science, but perhaps the best females are choosing alternative careers.

Acknowledgments

We gratefully acknowledge the ongoing contribution of the parents and children in the Twins Early Development Study (TEDS). TEDS is supported by a program grant (G0500079) from the U.K. Medical Research Council; our work on school environments and academic achievement is also supported by grants from the US National Institutes of Health (HD44454 and HD46167).

Abbreviations

TEDS	Twins Early Development Study.
NC	National Curriculum.

ACE	Additive genetic (A), Common (shared) environment (C), and non-shared environment (E).
MZ	Monozygotic.
DZ	Dizygotic.

References

- Akaike H. Factor analysis and AIC. *Psychometrika*. 1987; 52:317–332.
- Boomsma D, Busjahn A, Peltonen L. Classical twin studies and beyond. *Nature Reviews Genetics*. 2002; 3:872–882.
- Cohen, J. Statistical power analysis for the behavioral sciences. 2nd ed.. Hillsdale, New Jersey: Lawrence Erlbaum Associates; 1988.
- 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.
- Eccles, JS.; Adler, TF.; Futterman, R.; Goff, SB.; Kaczala, CM.; Meece, JL.; Midgley, C. Expectancies, values, and academic behaviors. In: Spence, JT., editor. *Perspective on achievement and achievement motivation*. San Francisco: Freeman; 1983. p. 75-146.
- Ericsson KA, Roring RW, Nandagopal K. Giftedness and evidence for reproducibly superior performance: an account based on the expert performance model. *High Ability Studies*. 2007; 18:3–56.
- Falconer DS. Inheritance of Liability to Certain Diseases Estimated from Incidence Among Relatives. *Annals of Human Genetics*. 1965; 29:51-&.
- Halpern DF, Benbow CP, Geary DC, Gur RC, Hyde JS, Gernsbacher M. The Science of Sex Differences in Science and Mathematics. *Psychological Science in the Public Interest*. 2007; 8:1–51.
- Haworth CMA, Dale PS, Plomin R. A twin study into the genetic and environmental influences on academic performance in science in 9-year-old boys and girls. *International Journal of Science Education*. 2008; 30:1003–1025. [PubMed: 20111670]
- Haworth CMA, Dale PS, Plomin R. The Etiology of Science Performance: Decreasing Heritability and Increasing Importance of the Shared Environment from 9 to 12 Years of Age. *Child Development*. (in press).
- Haworth CMA, Kovas Y, Dale PS, Plomin R. Science in Elementary School: Generalist Genes and School Environments. *Intelligence*. 2008; 36:694–701.
- Howe MJA, Davidson JW, Sloboda JA. Innate talents: Reality or myth? *Behavioral and Brain Sciences*. 1998:399–442. [PubMed: 10097018]
- Hyde JS. The Gender Similarities Hypothesis. *American Psychologist*. 2005; 60:581–592. [PubMed: 16173891]
- Hyde JS, Linn MC. Gender similarities in mathematics and science. *Science*. 2006; 314:599–600. [PubMed: 17068246]
- Jenkins EW, Nelson NW. Important but not for me: students' attitudes towards secondary school science in England. *Research in Science & Technological Education*. 2005; 23:41–57.
- Kovas Y, Haworth CMA, Dale PS, Plomin R. The genetic and environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*. 2007; 72:vii-160.
- Loehlin, JC.; Nichols, J. *Heredity, environment and personality*. Austin: University of Texas; 1976.
- Lykken DT, McGue M, Tellegen A, Bouchard TJ Jr. Emergenesis: Genetic traits that may not run in families. *American Psychologist*. 1992; 47:1565–1577. [PubMed: 1476327]
- Martin N, Boomsma DI, Machin G. A twin-pronged attack on complex trait. *Nature Genetics*. 1997; 17:387–392. [PubMed: 9398838]

- Miller PH, Slawinski Blessing J, Schwartz S. Gender differences in high-school students' views about science. *International Journal of Science Education*. 2006; 28:363–381.
- Neale, MC.; Boker, SM.; Xie, G.; Maes, H. Mx: Statistical Modeling. 7th ed.. Richmond, VA: Department of Psychiatry; 2006. VCU Box 900126, 23298
- Neale MC, Eaves LJ, Kendler KS. The power of the classical twin study to resolve variation in threshold traits. *Behavior Genetics*. 1994; 24:239–258. [PubMed: 7945154]
- Oliver BR, Plomin R. Twins Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behavior problems from childhood through adolescence. *Twin Research and Human Genetics*. 2007; 10:96–105. [PubMed: 17539369]
- Osborne J, Simon S, Collins S. Attitudes towards science: a review of the literature and its implications. *International Journal of Science Education*. 2003; 25:1049–1079.
- Plomin, R.; DeFries, JC.; McClearn, GE.; McGuffin, P. *Behavioral Genetics*. 5th Edition. New York: Worth; 2008.
- Plomin R, Kovas Y. Generalist genes and learning disabilities. *Psychological Bulletin*. 2005;592–617. [PubMed: 16060804]
- Price TS, Freeman B, Craig IW, Petrill SA, Ebersole L, Plomin R. Infant zygosity can be assigned by parental report questionnaire data. *Twin Research*. 2000; 3:129–133. [PubMed: 11035484]
- Rijsdijk FV, Sham PC. Analytic approaches to twin data using structural equation models. *Briefings in Bioinformatics*. 2002; 3:119–133. [PubMed: 12139432]
- Simpkins SD, vis-Kean PE, Eccles JS. Math and Science Motivation: A Longitudinal Examination of the Links between Choices and Beliefs. *Developmental Psychology*. 2006; 42:14.
- Smith C. Concordance in Twins - Methods and Interpretation. *American Journal of Human Genetics*. 1974; 26:454–466. [PubMed: 4858380]
- Speering W, Rennie L. Students' perceptions about science: The impact of transition from primary to secondary school. *Research in Science Education*. 1996; 26:283–298.
- Spelke ES. Sex differences in intrinsic aptitude for mathematics and science? A critical review. *American Psychologist*. 2005; 60:950–958. [PubMed: 16366817]

Key Points

- Although mean differences in aptitude do not appear to explain the under-representation of women in scientific careers, it is possible that underlying etiology of aptitude differs for boys and girls
- Could there be genetic and environmental factors that differentially influence the performance of males and females in science, especially at the high end?
- We find that not only do males and females perform equally well in school science, the same genes and environments are involved to the same extent in both sexes
- Differences in ability are not keeping women out of science, but perhaps the best females are choosing alternative careers.

Table 1

Means (standard deviations) and t-test results for science performance in boys and girls assessed by teachers at 9, 10 and 12 years

Age	Mean (sd)	Male Mean (sd)	Female Mean (sd)	d	Student's t-test
9	2.98 (0.58) n=2735	3.00 (0.60) n=1291	2.97 (0.55) n=1444	0.05	p=0.122
10	3.37 (0.66) n=2864	3.39 (0.69) n=1350	3.36 (0.64) n=1514	0.05	p=0.305
12	4.46 (0.94) n=3815	4.47 (0.97) n=1797	4.44 (0.91) n=2018	0.03	p=0.258

Note. Analyses with twin data were done on one randomly selected member of each twin pair. We repeated the analyses in the remaining twin pair and found the same pattern of results (details available from first author).

Table 2

Tetrachoric Correlations (95% CIs) by sex and zygosity

Age	MZM	MZF	DZM	DZF	DZOS
9	0.92 (0.85–0.96) N=438	0.89 (0.81–0.94) N=529	0.65 (0.50–0.77) N=423	0.70 (0.56–0.81) N=457	0.63 (0.52–0.73) N=831
10	0.87 (0.78–0.93) N=445	0.93 (0.87–0.96) N=562	0.69 (0.54–0.80) N=436	0.73 (0.60–0.82) N=472	0.65 (0.54–0.73) N=912
12	0.88 (0.81–0.93) N=633	0.85 (0.77–0.91) N=747	0.74 (0.62–0.83) N=567	0.82 (0.73–0.89) N=673	0.63 (0.54–0.72) N=1181

Note. MZM = monozygotic male pairs; DZM = dizygotic male pairs; MZF = monozygotic female pairs; DZF = dizygotic female pairs; DZOS = dizygotic opposite sex pair; CIs = confidence intervals. N = number of complete twin pairs.

Table 3

Fit statistics from the liability threshold model

Age	Model	χ^2	df	p	AIC
9 Years	Full model	6.813	8	0.557	-9.187
	Equate prevalence	9.376	1	0.002	7.376
	Fix r_{GO}/r_{CO} *	0.233	1	0.629	-1.767
	Equate male and female ACE*	1.132	4	0.889	-6.868
	AE*	26.798	5	0.000	16.798
	CE*	34.116	5	0.000	24.116
10 Years	E*	500.225	6	0.000	488.225
	Full model	4.987	8	0.759	-11.013
	Equate prevalence	0.237	1	0.626	-1.763
	Fix r_{GO}/r_{CO}	1.080	2	0.583	-2.920
	Equate male and female ACE	3.249	5	0.662	-6.751
	AE	36.225	6	0.000	24.225
12 Years	CE	34.154	6	0.000	22.154
	E	547.929	7	0.000	533.929
	Full model	13.514	8	0.095	-2.486
	Equate prevalence	0.414	1	0.520	-1.586
	Fix r_{GO}/r_{CO}	0.432	2	0.806	-3.568
	Equate male and female ACE	10.005	5	0.075	0.005
12 Years	AE	73.432	6	0.000	61.432
	CE	26.007	6	0.000	14.007
	E	641.763	7	0.000	627.763

Note. The full model is compared to the saturated model, and all subsequent models are compared to the full model. χ^2 = change in chi-square; df = change in degrees of freedom; AIC = Akaike Information Criterion (lower value indicates better fit); r_{GO}/r_{CO} = genetic/shared environmental correlation between opposite-sex pairs; ACE = additive genetic (A), shared environment (C), and non-shared environment (E).

* Allowing different prevalences.

The full model allows qualitative and quantitative sex differences as well as differences in prevalence between males and females. Fixing the genetic correlation between opposite-sex pairs (rGO) to be equivalent to same-sex dizygotic pairs (0.5) removes qualitative sex differences from the model. Equating male and female ACE parameters removes quantitative differences from the model. The AE model tests the significance of C; the CE model tests the significance of A; and the E model tests the significance of A and C.

Table 4

Parameter estimates from the best-fitting liability threshold model:

Age	A	C	E	Male Threshold	Female Threshold
9 Years	0.49 (0.33–0.65)	0.41 (0.26–0.55)	0.10 (0.06–0.15)	0.97 (0.91–1.04)	1.10 (1.04–1.17)
10 Years	0.45 (0.30–0.60)	0.45 (0.31–0.58)	0.10 (0.06–0.14)	1.04 (0.99–1.09)	1.04 (0.99–1.09)
12 Years	0.30 (0.16–0.45)	0.56 (0.44–0.68)	0.13 (0.09–0.18)	1.03 (0.99–1.07)	1.03 (0.99–1.07)

Note. A = additive genetic; C = shared environment; E = non-shared environment. The significantly lower threshold in males at 9 years indicates that there were more males than females in the high performance group at this age.