Academic performance of opposite-sex and same-sex twins in adolescence

Academic performance of opposite-sex and same-sex twins in adolescence: a Danish national cohort study

Linda Ahrenfeldt[1], Inge Petersen[1], Wendy Johnson[2], and Kaare Christensen[1,3]

1) **The Danish Twin Registry, Epidemiology, Biostatistics and Biodemography, The University of Southern Denmark, 5000 Odense, Denmark**
2) **Department of Psychology, The University of Edinburgh, Scotland Edinburgh, UK**
3) **Department of Clinical Biochemistry and Pharmacology and Department of Clinical Genetics, Odense University Hospital, 5000 Odense, Denmark**

Abstract

Testosterone is an important hormone in the sexual differentiation of the brain, contributing to differences in cognitive abilities between males and females. Effects of prenatal exposure to testosterone can be investigated comparing opposite-sex (OS) and same-sex (SS) twins, and indirect evidence exists of higher testosterone levels *in utero* among OS female compared with SS female twins, as well as some masculinization of perception and cognition. However, differences between OS and SS twins may also exist due to different psycho-social rearing environments. We compared ninth-grade test scores and teacher ratings of OS *(n = 1812)* and SS *(n = 4054)* twins as well as of twins and singletons *(n = 13,900)* in mathematics, physics/chemistry, Danish, and English. Males had significantly higher mathematics test scores than females, whereas females performed better in Danish, English, and neatness. Contrary to expectation, OS females had significantly lower scores than SS females in mathematics, adjusted oral test score difference -.19 (95% CI: -.33 to -.06), written -.22 (95% CI: -.35 to -.08) corresponding to a fifth of a standard deviation. SS female performance was similar to that of female singletons, but OS female performance was slightly lower, especially in mathematics and physics/chemistry. Scores for OS and SS males were similar in all topics. This study did not provide evidence for masculinization of female twins with male co-twins in academic performance in adolescence, but females with male twins showed slightly adverse academic performance, likely due to post-natal environmental influences.

Keywords: Twins; Sex-difference; Testosterone; Behavior; Academic performance; Mathematics; School engagement

Highlights

We examined the twin testosterone transfer hypothesis regarding academic performance
No overall differences in test scores were found between adolescent OS and SS twins
However, females with male twins showed slightly adverse performance in mathematics
No evidence that prenatal testosterone exposure masculinized academic performance

Introduction

Differences between opposite-sex (OS) and same-sex (SS) dizygotic twins can help us to understand how population-level sex differences emerge because members of twin pairs share genetic heritage and many prenatal and rearing environmental conditions, thus providing substantial control of potential sources of sex differences. Differences between female twins with OS/SS co-twins may provide evidence of testosterone exposure *in utero* (Miller, 1994) but of course OS and SS twins may also have different psycho-social environments due to the sexes of their co-twins (Cohen-Bendahan et al., 2005). Identification of the nature and extent of differences is the first step in disentangling these possibilities, and the literature on this remains inconsistent at present.

The literature regarding sex differences in cognitive abilities is, however, reasonably consistent. Differences in quantitative abilities have received most attention because of the large sex

differences in participation in professional careers in science and mathematics favouring males (Halpern et al., 2007). Males tend to outperform females on most measures of visuo-spatial abilities (though distributions overlap considerably, as is the case for all sex differences in cognitive abilities), which may contribute to sex differences in test scores in mathematics and natural science (Halpern et al., 2007). However, the gap appears to increase with age (Bharadwaj et al., 2012; Hyde, 2005). Females tend to do better in primary school, with males gradually catching up and surpassing them in high school, though this pattern has changed in recent years, with females holding their position much longer, even into university (Haworth et al., 2010; Hyde, 2005). It seems that the male advantage tends to emerge as the mathematical concepts being taught require more reasoning, more spatial abilities, and more complex problem-solving (Haworth et al., 2010; Hyde et al., 1990), though recent changes in the patterns suggest that cultural expectations matter too. Conversely, sex-differences favouring females in verbal abilities, such as reading, writing, and language usage, are documented in the literature (Halpern et al., 2007; Hedges and Nowell, 1995). While school achievement measures are not direct measures of abilities, they are generally strongly correlated with them.

The prenatal period is a critical time for sex hormones to shape the brain (Bull et al., 2010; Manson, 2008). Testosterone is an important hormone in the sexual differentiation of the brain, contributing to differences in cognitive abilities between males and females (Durdiakova et al., 2011). One explanation for the cognitive differences such as those in spatial abilities is the twin testosterone transfer (TTT) hypothesis. This reflects the possibility that human sex hormones can transfer from one twin to the other, most likely by diffusing across fetal membranes (Even and vom Saal, 1992). The amniotic fluid can permeate the fetal skin and the placenta until 18 weeks of gestation. Testosterone production in males begins at 7-8 weeks gestation and runs continuously high until 24 weeks of gestation (Tapp et al., 2011). Thus, the female of an OS twin pair may be exposed to higher levels of testosterone during prenatal development than a female member of an SS twin pair, and this in turn could render her more male-typical (masculinized). In addition, the male of an OS twin pair might be exposed to lower levels of testosterone than an SS male, thus rendering him less male-typical (de-masculinized) (Cohen-Bendahan et al., 2005). However, because males are already exposed to high levels of testosterone, the effect of testosterone exposure from male co-twins may be more pronounced in females than in males (Tapp et al., 2011).

Evidence for the TTT hypothesis comes from studies in litter-bearing mammals such as rats and mice demonstrating that exposure to sex hormones is influenced by the intrauterine fetal position (Ryan and Vandenbergh, 2002). Male foetuses have greater concentrations of testosterone than females, and females produce higher amounts of estradiol than males (vom Saal, 1989), but regardless of sex, a foetus located between two males has a higher concentration of testosterone than a foetus of the same sex located between females (Ryan and Vandenbergh, 2002; vom Saal, 1989). This phenomenon causes females to appear masculinized in several anatomical, physiological and behavioural traits such as aggressive behaviour and reproductive organs (Ryan and Vandenbergh, 2002). Likewise, female foetuses developing between females show more feminized traits as adults, for example earlier vaginal opening (Ryan and Vandenbergh, 2002). Thus, intrauterine position and the possibility of steroid transfer of especially testosterone from one foetus to another during fetal life have strong effects in animals commonly used as human models (Ryan and Vandenbergh, 2002; vom Saal, 1989).

In humans, masculinizing effects from prenatal testosterone exposure can be studied by comparing OS with SS twin pairs (Miller, 1994). There has been more evidence for masculinization of females than for de-masculinization of males (Cohen-Bendahan et al., 2005). OS females have been reported to be masculinized in physiological traits, such as tooth size (Dempsey et al., 1999; Ribeiro et al., 2013), otoacoustic emissions (sounds produced by the inner ear) (McFadden, 1993),

second-to-fourth-finger-length ratio (van Anders et al., 2006), and maternal fitness (Lummaa et al., 2007), but other studies have reported negative findings (Gaist et al., 2000; Medland et al., 2008a; Medland et al., 2008b). It has been shown that some aspects of sensation seeking are increased in males relative to females and that this could be influenced by testosterone (Zuckerman et al., 1978). In addition, two studies found greater sensation-seeking, including experience-seeking, in OS females compared with SS females (Resnick et al., 1993; Slutske et al., 2011). These findings suggest effects of hormone exposures on later behavioural development, though psycho-social explanations cannot be excluded (Resnick et al., 1993).

Apart from possible differences in prenatal hormone exposure, OS and SS twins may also be different due to different psycho-social rearing environments (Cohen-Bendahan et al., 2005). It is possible that OS females are raised in more male-typical environments than are SS females, and that this may affect their behaviour. Similarly, the behaviour of OS males may be influenced by being raised with twin sisters (Cohen-Bendahan et al., 2005). Most studies failed to detect de-masculinization in males (Cohen-Bendahan et al., 2005), but a few studies have suggested that OS males might be de-masculinized in sex-typical behaviour (Cohen-Bendahan et al., 2005; Elizabeth and Green, 1984). In general, there is sparse evidence regarding twin interactions; however, indications of differences in socialisation in OS compared with SS twins are present (Pulkkinen et al., 2003). A Finnish study of peer-assessed socio-emotional behaviour of 1874 twins and their 23,200 non-twin classmates found that classmates rated twins higher than sex-matched singletons in adaptive behaviour and that this especially was the case among OS twins (Pulkkinen et al., 2003).

Support for the TTT hypothesis has also been found in studies of perceptual and cognitive differences (Tapp et al., 2011) including mental rotation (MRT) performance (the process of visualizing the representation of an object turning around). In general, males outperform females in MRT and prenatal exposure to testosterone has been suggested as a cause (Heil et al., 2011). Higher mental rotation performance in OS compared with SS females and singleton females raised with older brothers has been observed and it may indicate effects of prenatal testosterone on brain functions (Heil et al., 2011; Vuoksimaa et al., 2010). Nonetheless, studies comparing OS and SS twins regarding academic performance are scarce. Only one recent study of 13,368 twins and 837,752 singletons born during 1973-1981 in Sweden has investigated differences between OS and SS twins in grade point average (GPA) (Hjern et al., 2012). They found that SS twins of both sexes had slightly higher average scores in ninth grade and more often attained university education by age 27-35 than did OS twins. However, the differences in GPA were small and non-significant, and the authors did not discuss possible reasons for their results, which did not support the TTT. However, many factors may influence test scores. A previous Danish study comparing academic performance between twins and singletons born 1986-1988, has identified potential confounding factors, including parental age and education (Christensen et al., 2006). Evidence suggests that parental ages may influence cognitive ability of offspring (Edwards and Roff, 2010; Malaspina et al., 2005; Saha et al., 2009) and that parental education, a surrogate of socioeconomic status, is associated with offspring academic performance as well (Sirin, 2005).

The recent Danish study comparing ninth-grade test scores in twins and singletons concluded that twins have similar academic performance to singletons in adolescence (Christensen et al., 2006). In this study comparing OS and SS twins it was also important to investigate differences between singletons and OS and/or SS twins, because these comparisons can reveal whether the OS/SS twin groups differed in performance from the general population. If, for instance, SS female twins had better performance than OS females in some aspect, there are three possibilities: Firstly, SS twins had better than the average singleton score (a positive influence of having a twin sister) while the OS twins were average (no impact of having a twin brother). Secondly, SS twins were average and the OS twins below average (an inverse influence of having a twin brother). Thirdly,

the SS twins were above average and the OS twins below average, suggesting presence of both kinds of twin effects.

Our primary objective was to compare academic performance in mathematics, physics/chemistry, Danish and English in female and male members of OS and SS twin pairs. The other objective was to determine whether the twin groups differed in performance from the general population. This was achieved by comparing the OS twins and the SS twins, born in Denmark during 1986-1990, for each sex, with a 5% random sample of singletons born in Denmark during the same period and surviving until January 1, 2003. We made use of data on all twins whose co-twins had survived to the age of at least 17 to make sure that most of the twins in the present study had grown up together in Denmark as members of intact twin pairs. Individuals who had lived abroad for more than two years during the age range 6-14 and those who had emigrated but returned to Denmark after age 14 were excluded from the analyses. Eighty-four twin individuals were excluded from the outset because the co-twin was stillborn (not shown). In addition, 396 individuals were excluded due to death before age 17, including 196 twins (45 twin pairs and 106 twin individuals); 106 twins were excluded due to death of the co-twin, and 508 individuals emigrated during the study period. One twin was excluded because the twin pair was separated by emigration. Thus, the study base consisted of 2941 female and 2925 male twins as well as 6771 female and 7129 male singletons. All analyses for OS and SS twins as well as for twins and singletons were conducted separately for each sex.

Material and methods

Material

This study was based on the Danish Civil Registration System (Pedersen, 2011) which identifies individuals by personal identification numbers (CPR-numbers) linked through Statistics Denmark to population databases or registers. We included information from four registers: the Danish Demographic Database, including information on parental identities, deaths, migrations, and adoptions (JK, 2000); the Integrated Database for Labour Market Research containing information on the highest obtained parental education (Petersson et al., 2011); the register of Compulsory School Completion Assessments and Test Scores compiled by the Danish Ministry of Education (Education, 2000) and the Danish Twin Registry, which comprises more than 85,000 Danish twin pairs born since 1870 (Skytthe et al., 2011), with all twins born after 1973 identified through the Medical Birth Registry (Skytthe et al., 2002). Zygosity determination of same-sex twin pairs was based on four standard questions about similarity of appearance, a method with less than 5% misclassification (Christiansen et al., 2003), though responses to these questions were missing for 27% of the SS twin pairs, who were classified as twins of unknown zygosity (ssUZ).

School achievements were available for the years 2002-2006 for the ninth-grade students aged 15-16 at test, corresponding to the 1986-1990 birth cohorts. Most Danish students in ninth grade completed standardized, nationwide tests of academic achievement in several subject areas, scored on a scale of 0-13. Average performance was rated as 8, higher scores indicating better performance (Christensen et al., 2006). Teacher ratings in ninth grade supplemented the test scores, also scored on the 0-13 scale. These ratings were subjective evaluations of the students' overall academic performance during the academic year whereas test scores reflected actual student performance rated by both the teachers and external examiners at specific times on specific tests (Petersen et al., 2009). The test scores and teacher ratings covered major domains of academic achievement, such as mathematics and Danish, which were the main topics included in the present study. Oral and written exercises, neatness, and Danish spelling were graded. In addition, physics/chemistry and English oral exercise grades were recorded.

We distinguished between non-attainment and missing scores. Non-attainment refers to when an individual had no grades at all, which could be a serious problem for a young person's further possibilities in the educational system. For instance it was not possible to continue in an upper secondary school without the grades. Reasons for non-attainment were multiple: drop-out or attendance at a specialized school due to disabilities were common reasons. Schools for children with learning disabilities were not required to report test scores and teacher ratings used for statistics to the ministry of Education. In addition, some private schools such as those following the Waldorf pedagogy as a principle did not test at all (Christensen et al., 2006; Petersen et al., 2009). Test scores were considered missing when an individual had some scores but not all. As the ninth grade exams were not mandatory between 1993 and 2006 in Denmark (Education, 1993; Lærerforening, 2014), a small proportion of individual students in public schools opted not to take some of the tests. Finally, there could be missing teacher ratings due to failures in reporting to the ministry of Education.

Statistical methods and confounders

Analyses of differences in categorical baseline characteristics (mortality, emigration, death of a co-twin and parental education) for OS and SS twins and twins and singletons were tested by chi-square tests. Parental educational attainments were coded as categorical variables for the highest obtained education by December 31, 2002. By that date, the median ages were 42.5 years for mothers and 45.1 years for fathers. The educational variables were coded from 0-5, corresponding to the following categories: Basic school 8^{th}-10^{th} grade, vocational school, secondary education, short higher education, medium higher education or Bachelor's degree, and higher academic or professional degree. Differences in continuous background variables (parental age) were investigated using t-tests. The statistical analyses showed that parental age and education differed significantly between OS and SS twins (Table 1) due to inclusion of both monozygotic (MZ) and dizygotic (DZ) twins in the SS twin group. This was done because of the rather large proportion (app. 27%) of the sample that was ssUZ, which made it impossible to allocate all the SS to zygosity group. Supporting the inappropriateness of dropping the ssUZ twins, a study consisting of 2,413 Danish twin pairs from birth cohorts 1986-1990 found that ssUZ twins averaged lower school achievement scores than twins of known zygosity (Petersen et al., 2009). Because OS twins are always DZ, these lower-achieving twins would not be dropped from the OS group, so exclusion of the ssUZ twins from the SS group would have biased the results towards lower performance for the OS twins. Considering MZ and DZ twins together should not bias group comparisons.

Effect size differences for co-variates (maternal and paternal age, and education) for twins and singletons with at least one test score were performed applying linear regression models to both the raw data and to the data adjusted for the co-variates maternal and paternal age, and education. Parental age and education were independently associated with test scores (Supplementary Tables 1 and 2). Controlling them in our models did slightly change the effect estimates. Thus, we present all results both with and without adjustments for parental age and education.

We considered birth weight an intermediate factor in the association between OS/SS twins and academic performance. Low birth weight is associated with social disadvantage that may result in reduced brain growth potential that can influence later academic achievement (de Kieviet et al., 2012; Fawke, 2007; Mortensen et al., 2009). However, low birth weight is also associated with low maternal IQ, which is genetically influenced, so it may be this that's contributing to the child's poor school achievement, rather than the conditions surrounding low birth weight, per se (Chaudhari et al., 2005; Voss et al., 2012; Weisglas-Kuperus et al., 2009). Adjustment for birth weight is a controversial subject because adjustment for an intermediate factor requires that no unmeasured confounders act on either the intermediary variable or on the outcome (Cole and Hernan, 2002).

This is unlikely to be true for birth weight, hence interpretation is difficult. In addition, when the goal is to estimate overall effects, adjustment for variables on the causal pathway is unwarranted (Hernandez-Diaz et al., 2006). Because we were interested in the overall effects on academic performance of having a co-twin of the opposite sex compared with having a co-twin of the same sex, we present results without adjustment for birth weight. Moreover, results were similar when birth weight was controlled.

Logistic regression models were used to estimate odds ratios (OR) and 95% confidence intervals (CI) for test score non-attainment. We did this separately for males and females, OS and SS twins, and OS/SS twins and singletons, respectively. Analysis of differences in test scores and teacher rating between females and males, OS and SS twins, and OS/SS twins and singletons was done using linear regression models, with separate models for mathematics (oral, written, and neatness), oral physics/chemistry, Danish (oral, written, spelling, and neatness) and oral English. Sex differences were investigated for the full sample consisting of both twins and singletons. In addition, sex differences were investigated in the sample of all twins and repeated in the sample of only SS twins. Effect sizes were calculated by dividing the academic subject mean differences by their overall sample standard deviations. To account for the intra-pair correlations when estimating standard errors, the cluster option in Stata (release 13) was used.

Results

Table 1 presents the numbers of live births among twins and the 5% random sample of singletons born in Denmark 1986-1990, along with parental age and education information. No significant difference in mortality was found between OS and SS female twins: 2.4% vs 3.5% (p=0.107), but mortality was lower for OS than SS males: 2.0% vs 3.6% (p=0.017). Mortality in the entire period (before age 17) was significantly higher for twins than singletons: 3.1% vs 0.8% for females and 3.1% vs 1.2% for males. The difference was significant in the neonatal period (until 28 days after birth) in both sexes (p<0.001); however, mortality rates in the post-neonatal period (29-365 days after birth) were similar for female twins and singletons: 0.49% vs 0.26% (p=0.065), but for males mortality of twins was still higher than that of singletons: 0.68% vs 0.38% (p=0.042). Mortality of a co-twin was lower among OS than among SS twins of both sexes, but the difference was not significant for males (females: 0.9 vs 1.8%, p=0.047; males: 1.3 vs 2.3%, p=0.065.) The proportions of OS and SS twins: 3.2% vs 2.3% (p=0.050) and twins and singletons: 2.6% vs 2.5% (p=0.551) who emigrated were similar.

Mothers of OS twins were significantly older than mothers of SS twins (29.8 vs 28.9 years, p<0.001). A similar pattern was observed in fathers (32.7 vs 31.6 years, p<0.001). Mothers of OS twins had slightly longer education than mothers of SS twins (p<0.001), as did fathers (p=0.033). Mothers of twins were significantly older at childbirth than mothers of singletons (29.2 vs 28.0 years, p<0.001), as were fathers (32.0 vs 31.0 years, p<0.001). Mothers of twins had slightly longer education than mothers of singletons (p=0.021). Fathers tended to as well, but the difference did not reach significance (p=0.077). All means fall in the category “vocational school”.

Non-attainment

Table 2 presents OR and 95% CI for test score non-attainment, before and after adjustment for parental age and education. Males had significantly higher risk of test score non-attainment than females, and the unadjusted and adjusted ORs were approximately similar (adjusted OR: 1.70 (95% CI: 1.54 to 1.88)). Risks of non-attainment were similar in OS and SS females (adjusted OR: 1.05 (95% CI: .76 to 1.47), unadjusted OR: .98 (95% CI: .74 to 1.29)). A tendency was found towards higher risk of non-attainment for OS females compared with singleton female twins, but the difference was not significant (adjusted OR: 1.25 (95% CI: .96 to 1.63), unadjusted OR: 1.15 (95%

CI: .92 to 1.44)). The same pattern was found for SS and singleton females (adjusted OR: 1.16 (95% CI: .92 to 1.45), unadjusted OR: 1.18 (95% CI: .97 to 1.43)).

Similar risks of non-attainment were found for OS and SS male twins (adjusted OR: 1.10 (95% CI: .83 to 1.46), unadjusted OR: .98 (95% CI: .77 to 1.25)). In addition, no significant difference was found between OS and singleton males (adjusted OR: 1.14 (95% CI: .91 to 1.43), unadjusted OR: .98 (95% CI: .81 to 1.19)) as well as between SS and singleton males (adjusted OR: 1.02 (95% CI: .84 to 1.24), unadjusted OR: 1.00 (95% CI: .85 to 1.18)).

Similar patterns of differences in test scores and teacher ratings for those with missing scores and those with all scores available were found for SS and OS twins as well as for singletons, suggesting that the relative frequencies of missingness did not differ with twin type or singleton birth status (not shown).

Academic performance

Table 1 also presents summary statistics for the academic performance scores. Overall, there were high degrees of similarity between OS and SS twins and between twins and singletons for both sexes. Table 3 and Supplementary Table 3 provide further information regarding the effect sizes of differences, before and after statistical adjustment for potentially confounding variables.

Males vs females

Overall, male twins and singletons showed significantly higher ninth-grade test scores in mathematics than females (Table 3). There were adjusted differences of .10 (95% CI: .05 to .14) in oral and .23 (95% CI: .18 to .28) in written mathematics, corresponding to .06 and .15 of SD. Conversely, females had slightly higher teacher ratings in mathematics than males. Performances in physics/chemistry were similar for males and females. Unlike mathematics, males had significantly lower scores in Danish than females. All mean differences in Danish oral, written, spelling, and neatness for both test scores and teacher ratings were significant, with p-values below 0.001, corresponding to .33-.60 SD. A similar pattern was found in English, with males having significantly lower scores than females, adjusted test score difference -.36 (95% CI: -.42 to -.30). In addition, males received lower marks for neatness than females in both mathematics and Danish test scores and teacher ratings, adjusted mathematics difference in test score -.68 (95% CI: -.72 to -.65). The patterns of sex differences were the same in singletons as in the full sample consisting of both twins and singletons. In twins, the overall patterns of sex differences were also identical with those of the full sample even when excluding the OS twins (not shown).

OS vs SS twins, and twins vs singletons

Female OS twins obtained significantly lower scores in mathematics than SS twins, adjusted difference in oral test score -.19 (95% CI: -.33 to .06), written -.22 (95% CI: -.35 to -.08), corresponding to .12 and .14 SD respectively (Table 3). Before adjustments, the differences were still present; however, only borderline significant in oral mathematics. There were similar differences in the teacher ratings. Physics/chemistry showed the same pattern, with significantly lower scores for OS than SS females, adjusted difference in test score -.18 (95% CI: -.34 to -.02); however, the unadjusted difference was not significant. SS female twins had similar mathematics scores as singleton females, whereas OS females had significantly lower test scores than singleton females after adjustments in both oral -.20 (95% CI: -.32 to -.09) and written mathematics -.22 (95% CI: -.33 to -.11) as well as in physics/chemistry -.20 (95% CI: -.33 to -.06). The same pattern was found before adjustments; however, differences in mathematics were only borderline significant and not significant in physics/chemistry. Thus, OS twins performed less well on average than SS female twins in mathematics and in physics/chemistry, contrary to the TTT hypothesis.

Performance was similar in OS and SS female twins in all aspects of Danish and English both before and after adjustments; however, there was a slight tendency, though not significant, towards OS females having lower performance than both SS and singleton females in all adjusted analyses, which would be consistent with a slightly lower OS performance overall. OS females had significantly lower adjusted oral Danish test scores -.13 (95% CI: -.24 to -.01) than singleton females, as well as in the adjusted Danish written and spelling test and teacher-rating scores. SS females had similar performance as singleton females except for a slightly lower performance in written Danish.

OS females had slightly lower teacher ratings than SS females in neatness in mathematics, adjusted difference -.12 (95% CI: -.23 to -.01). SS females had the same mean scores as singletons -.00 (95% CI: -.07 to .08) but OS females scored lower than singletons to the same degree -.12 (95% CI: -.21 to -.03).

OS and SS male twins showed no significant differences in any area of performance, either before or after adjustments (Supplementary Table 3). Performance was similar for twins and singletons after adjustments in all aspects of mathematics, physics/chemistry, and Danish, but before adjustments, OS males showed slightly better performance than singleton males in physics/chemistry .14 (95% CI: .00 to .28) as well as in the teacher scores in oral Danish .15 (95% CI: .05 to .26) and English .12 (95% CI: .00 to .23). The former would not be consistent with the TTT hypothesis, but the latter would. However, in English, twin males obtained significantly lower test scores than singleton males, adjusted difference for OS/singleton males -.15 (95% CI: -.28 to -.01) and for SS/singleton males -.15 (95% CI: -.27 to -.04) but there was no difference in teacher ratings after adjustments.

Parents' age and education

The raw associations between the potential confounding variables and the test scores showed significantly higher performance with higher maternal and paternal age and education in both sexes (Supplementary Tables 1 and 2). When all the potential confounders were included in the regression model, the associations among confounders and test scores were attenuated. Most of the effect of paternal age disappeared, except that females having young fathers (below 22 years) still had lower test scores than females with fathers in the reference group (22-29 years). The significant effect of maternal age persisted after adjustments; however, the difference was only borderline significant for females with mothers in the youngest age group (below 20 years). The associations with maternal education were attenuated most by the adjustments for paternal education and vice versa (not shown). However, the effect of parental education on test scores was still highly significant after adjustments for all the other potential confounders.

Discussion

In this nationwide study of Danish adolescent twins and singletons, we found, as expected, that sex differences in academic achievement were dependent on topics. Males had higher test scores in mathematics than females (.06-.15 SD), whereas females performed significantly better in Danish (.33-.49 SD), English (.20 SD), and neatness (.44-.64 SD). Contrary to hypotheses involving prenatal exposure to testosterone and/or postnatal social exposure, we found slightly lower scores among OS than SS females in mathematics and in physics/chemistry (.11-.14 SD), but no significant differences in Danish and English. In fact, the results suggested slightly adverse influences of growing up with a twin brother for females. The slightly lower scores for OS than SS females in our study were consistent with the results from a recent Swedish study which found higher average scores in ninth grade for SS than OS twins of both sexes (Hjern et al., 2012). No scores for males were dependent on whether they had twin sisters or brothers.

We hypothesized that OS female twins may adopt more boyish patterns of interest and activity that contribute to patterns of academic performance, and vice versa for OS males. The size of the difference in mathematics scores between OS and SS females was comparable with that of the sex difference, but in the opposite direction. SS female performance was not significantly different from that of singletons, but OS performance was lower. In addition, in Danish and English, where females on average performed better than males, no significant difference was found between OS and SS females. Only in teacher ratings in mathematics marks for neatness did OS females resemble males when compared with SS females, but the difference was small (.12 SD). Our findings did not provide evidence of masculinization of female OS twins in academic performance in adolescence; thus, this study implies that the psycho-social effect on an OS female of having a same-aged twin brother may be larger than any effect of prenatal testosterone exposure. Another psycho-social effect could be that parents influence cognitive skills differently depending on child sex. There is evidence that parents estimate the intelligence of their sons to be higher than that of their daughters (Furnham et al., 2002; Furnham and Valgeirsson, 2007), and one study found that this was especially the case regarding logical-mathematical and spatial intelligence (Furnham et al., 2002). In addition, there is evidence that females estimate their own intelligence significantly lower than do males (von Stumm et al., 2009) and that this too is especially the case regarding mathematical and spatial intelligence (Furnham, 2001; Rammstedt and Rammsayer, 2000). As shown by stereotype threat and perceived ability studies, higher confidence in abilities can be associated with better performance, especially in specific testing situations (Ambady et al., 2001; Picho et al., 2013).

Thus, many factors may influence test scores and teacher ratings, but the finding of no differences between OS and SS males was in accordance with other studies investigating perceptual and cognitive traits (McFadden, 1993; McFadden et al., 1996; Vuoksimaa et al., 2010). Even though the results of this study did not provide evidence of masculinization of female OS twins, our findings did not necessarily contradict earlier positive findings of prenatal masculinization in other traits, such as those replicating the male advantage in mental rotation ability in OS females (Heil et al., 2011; Vuoksimaa et al., 2010). The reason is that the tasks on which the students' scores and ratings were based did not rely upon the same ability as the mental rotation scores in any direct way. Even though school achievement tests are strongly correlated with IQ (Bartels et al., 2002), and IQ and GPA are highly heritable (Hicks et al., 2008), evidence suggests that non-cognitive variables including student engagement and school-related motivation are also predictive of academic performance, and that sex differences also exist among these variables (Hicks et al., 2008; Lam et al., 2012). A recent study of 300 seventh- to ninth-grade students from each of 12 countries found that females tended to report higher levels of school engagement than males (Lam et al., 2012). A study of 1353 Austrian students with a mean age of 13.7 years suggested that school-related motivation was more important to school achievement for males than for females, suggesting that females make efforts and remain engaged even when they find the tasks boring (Freudenthaler et al., 2008). If so, females may be more likely to compensate for lower ability by working harder, thus maintaining higher school achievement than they would otherwise. Males' tendencies to more aggressive and disruptive behaviour may also impede their achievement. For example, a large study of 17-year-old SS twin pairs from Minnesota investigating the association between personality and school achievement found that males' higher levels of aggression and its correlated traits, including disruptive behaviour, resulted in significantly lower mean GPA compared with females (Hicks et al., 2008). Moreover, in studies of the effects of interest on reading comprehension of males and females, males' reading comprehension was better when they were interested in the text that they were reading, whereas the performance of females was not related to their expressed interest. This might suggest that the academic performance of males may

be more influenced by their level of interest than is females' (Oakhill and Petrides, 2007). There is evidence that females imitate males much more than males imitate females (Bussey and Perry, 1982) and that tomboyishness is more common and more accepted than the comparable thing in males (Feinman, 1981). It seems possible that twins feed off each other in school engagement, so that SS females support each other in the overall tendency to work harder than males, but their male co-twins' relative lack of motivation undermines that of OS females.

Our results were consistent with those of a recent Swedish study showing similarity between twins and singletons in ninth-grade school achievement (Hjern et al., 2012). Moreover, they were in accordance with our Danish study of the birth cohorts 1986-1988 that showed similar academic performance in adolescence between twins and singletons (Christensen et al., 2006). However, because the present study investigated more birth cohorts and academic performance in specific topics and differences between subgroups of twins compared with singletons, small differences from that study were found.

Strengths of this study were its large nationwide register-base with minimal selection criteria and inclusion of information on important potential confounders for the majority of twins. The observed associations with covariates were as expected based on earlier analyses using the same data sources (Christensen et al., 2006; Hansen et al., 2011), suggesting consistent data quality. Another advantage was the differentiation of academic performance into topics and sub-topics, resulting in more detailed investigation of masculinizing effects among OS females and SS males.

The large group of twins with unknown zygosity (app. 27%) was a limitation of this study because the unknown zygosity of these SS twins made it impossible to exclude the MZ twins and thus make the most valid test of the TTT hypothesis, which is the comparison of only ssDZ with OS twins (Cohen-Bendahan et al., 2005). However, according to a meta-analysis, differences in intelligence between twins and singletons were not influenced by zygosity status (Voracek and Haubner, 2008), and this also applied to academic achievement in a recent study of approximately 10,000 Dutch twins (de Zeeuw et al., 2012).

A further limitation was that control for fertility treatments such as *in vitro* fertilization (IVF) was not possible. A higher proportion of IVF children would be expected in the OS than in our SS group that consisted of both MZ and DZ twins because many multiple births in IVF take place because of implantation of multiple fertilized eggs. These generate DZ twins, and all OS twins are DZ. Parents who undergo IVF treatment tend to have higher socioeconomic position than average (Hjern et al., 2012). This is consistent with the greater average age and educational level of parents of OS twins than parents of SS twins (Table 1), which may favour OS twins regarding academic performance. On the other hand, if IVF treatment influences cognitive outcomes negatively, we would expect lower performance in OS than SS twins. However, natural multiple ovulation in older age is another reason for the increase in the twinning rate since the mid-1980s as females have waited longer to start their families (Blondel and Kaminski, 2002), and it is most likely that only a small proportion of the twins in our sample had been born through in vitro fertilization because the dramatic rise in its use was in the 1990s (Herskind et al., 2005). In addition, no differences were found in studies comparing cognitive development and school performance in IVF and naturally conceived children, controlling for parental education (Mains et al., 2010; Wagenaar et al., 2008).

Missing data was another limitation; however, a similar pattern of differences in test scores for those with missing scores vs those with all scores available was found for SS and OS twins as well as for singletons.

Conclusions

In conclusion, we found no evidence that prenatal testosterone exposure masculinized academic performance in adolescence for either males or females. Instead, for females a slightly adverse

influence on academic performance from a shared environment with a twin brother was indicated. Future twin studies should shed light on twin interactions regarding the psycho-social environment and investigate potential consequences of being raised as a twin.

Acknowledgments
This study was supported by research Grants from the National Institute on Aging (NIA-PO1-AG08761, NIAP01-AG031719) and from the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement no 259679.
No conflicts of interest

References

Ambady, N., Shih, M., Kim, A., Pittinsky, T.L., 2001. Stereotype susceptibility in children: effects of identity activation on quantitative performance. Psychol Sci 12, 385-390.
Bartels, M., Rietveld, M.J., Van Baal, G.C., Boomsma, D.I., 2002. Heritability of educational achievement in 12-year-olds and the overlap with cognitive ability. Twin Res 5, 544-553.
Bharadwaj, P., De Giorgi, G., Hansen, D., Neilson, C., 2012. The Gender Gap in Mathematics: Evidence from Low- and Middle-Income Countries. National Bureau of Economic Research, Cambridge, Mass.
Blondel, B., Kaminski, M., 2002. Trends in the occurrence, determinants, and consequences of multiple births. Semin.Perinatol. 26, 239-249.
Bull, R., Davidson, W.A., Nordmann, E., 2010. Prenatal testosterone, visual-spatial memory, and numerical skills in young children. Learn Individ Differ 20, 246-250.
Bussey, K., Perry, D., 1982. Same-sex imitation: The avoidance of cross-sex models or the acceptance of same-sex models? Sex Roles 8, 773-784.
Chaudhari, S., Otiv, M., Chitale, A., Hoge, M., Pandit, A., Mote, A., 2005. Biology versus environment in low birth weight children. Indian Pediatr 42, 763-770.
Christensen, K., Petersen, I., Skytthe, A., Herskind, A.M., McGue, M., Bingley, P., 2006. Comparison of academic performance of twins and singletons in adolescence: follow-up study. BMJ 333, 1095.
Christiansen, L., Frederiksen, H., Schousboe, K., Skytthe, A., von Wurmb-Schwark, N., Christensen, K., Kyvik, K., 2003. Age- and sex-differences in the validity of questionnaire-based zygosity in twins. Twin.Res. 6, 275-278.
Cohen-Bendahan, C.C., van de Beek, C., Berenbaum, S.A., 2005. Prenatal sex hormone effects on child and adult sex-typed behavior: methods and findings. Neurosci.Biobehav.Rev. 29, 353-384.
Cole, S.R., Hernan, M.A., 2002. Fallibility in estimating direct effects. Int J Epidemiol 31, 163-165.
de Kieviet, J.F., Zoetebier, L., van Elburg, R.M., Vermeulen, R.J., Oosterlaan, J., 2012. Brain development of very preterm and very low-birthweight children in childhood and adolescence: a meta-analysis. Dev Med Child Neurol 54, 313-323.
de Zeeuw, E.L., van Beijsterveldt, C.E., de Geus, E.J., Boomsma, D.I., 2012. Twin specific risk factors in primary school achievements. Twin Res Hum Genet 15, 107-115.
Dempsey, P.J., Townsend, G.C., Richards, L.C., 1999. Increased tooth crown size in females with twin brothers: Evidence for hormonal diffusion between human twins in utero. Am.J.Hum.Biol. 11, 577-586.
Durdiakova, J., Ostatnikova, D., Celec, P., 2011. Testosterone and its metabolites--modulators of brain functions. Acta Neurobiol Exp (Wars) 71, 434-454.
Education, D.M.o., 1993. Lov om Folkeskolen [in Danish].
Education, D.M.o., 2000. Tests, assessment and teaching: an overall evaluation of compulsory school completion assessment and test scores May-June 2004 [in Danish], Copenhagen: Danish Ministry of Education.
Edwards, R.D., Roff, J., 2010. Negative effects of paternal age on children's neurocognitive outcomes can be explained by maternal education and number of siblings. PLoS One 5, e12157.
Elizabeth, P.H., Green, R., 1984. Childhood sex-role behaviors: similarities and differences in twins. Acta Genet.Med.Gemellol.(Roma.) 33, 173-179.
Even, M.D., vom Saal, F.S., 1992. Seminal vesicle and preputial gland response to steroids in adult male mice is influenced by prior intrauterine position. Physiol Behav 51, 11-16.
Fawke, J., 2007. Neurological outcomes following preterm birth. Semin Fetal Neonatal Med 12, 374-382.

Feinman, S., 1981. Why is cross-sex-role behavior more approved for girls than for boys? A status characteristic approach. Sex Roles 7, 289-300.
Freudenthaler, H.H., Spinath, B., Neubauer, A.C., 2008. Predicting school achievement in boys and girls. Eur J Pers 22, 231-245.
Furnham, A., 2001. Self-estimates of intelligence: culture and gender difference in self and other estimates of both general (g) and multiple intelligences. Pers Individ Dif 31, 1381-1405.
Furnham, A., Reeves, E., Budhani, S., 2002. Parents think their sons are brighter than their daughters: sex differences in parental self-estimations and estimations of their children's multiple intelligences. J Genet Psychol 163, 24-39.
Furnham, A., Valgeirsson, H., 2007. Parents' estimations of their own intelligence and that of their children: a comparison between English and Icelandic parents. Scand J Psychol 48, 289-298.
Gaist, D., Bathum, L., Skytthe, A., Jensen, T.K., McGue, M., Vaupel, J.W., Christensen, K., 2000. Strength and anthropometric measures in identical and fraternal twins: no evidence of masculinization of females with male co-twins. Epidemiology 11, 340-343.
Halpern, D.F., Benbow, C.P., Geary, D.C., Gur, R.C., Hyde, J.S., Gernsbacher, M.A., 2007. The Science of Sex Differences in Science and Mathematics. Psychol Sci Public Interest 8, 1-51.
Hansen, T.G., Pedersen, J.K., Henneberg, S.W., Pedersen, D.A., Murray, J.C., Morton, N.S., Christensen, K., 2011. Academic performance in adolescence after inguinal hernia repair in infancy: a nationwide cohort study. Anesthesiology 114, 1076-1085.
Haworth, C.M., Dale, P.S., Plomin, R., 2010. Sex Differences in School Science Performance from Middle Childhood to Early Adolescence. Int J Educ Res 49, 92-101.
Hedges, L.V., Nowell, A., 1995. Sex differences in mental test scores, variability, and numbers of high-scoring individuals. Science 269, 41-45.
Heil, M., Kavsek, M., Rolke, B., Beste, C., Jansen, P., 2011. Mental rotation in female fraternal twins: Evidence for intra-uterine hormone transfer? Biol.Psychol. 86, 90-93.
Hernandez-Diaz, S., Schisterman, E.F., Hernan, M.A., 2006. The birth weight "paradox" uncovered? Am J Epidemiol 164, 1115-1120.
Herskind, A.M., Basso, O., Olsen, J., Skytthe, A., Christensen, K., 2005. Is the natural twinning rate still declining? Epidemiology 16, 591-592.
Hicks, B.M., Johnson, W., Iacono, W.G., McGue, M., 2008. Moderating Effects of Personality on the Genetic and Environmental Influences of School Grades Helps to Explain Sex Differences in Scholastic Achievement. Eur J Pers 22, 247-268.
Hjern, A., Ekeus, C., Rasmussen, F., Lindblad, F., 2012. Educational achievement and vocational career in twins - a Swedish national cohort study. Acta Paediatr 101, 591-596.
Hyde, J.S., 2005. The gender similarities hypothesis. Am Psychol 60, 581-592.
Hyde, J.S., Fennema, E., Lamon, S.J., 1990. Gender differences in mathematics performance: a meta-analysis. Psychol Bull 107, 139-155.
JK, P., 2000. The Danish demographic database: Longitudinal data for advanced demographic methods, Odense, Denmark.
Lærerforening, D., 2014. Undervisning Skolestruktur og lovændringer [in Danish], http://www.dlf.org.
Lam, S.F., Jimerson, S., Kikas, E., Cefai, C., Veiga, F.H., Nelson, B., Hatzichristou, C., Polychroni, F., Basnett, J., Duck, R., Farrell, P., Liu, Y., Negovan, V., Shin, H., Stanculescu, E., Wong, B.P., Yang, H., Zollneritsch, J., 2012. Do girls and boys perceive themselves as equally engaged in school? The results of an international study from 12 countries. J Sch Psychol 50, 77-94.
Lummaa, V., Pettay, J.E., Russell, A.F., 2007. Male twins reduce fitness of female co-twins in humans. Proc.Natl.Acad.Sci.U.S.A 104, 10915-10920.

Mains, L., Zimmerman, M., Blaine, J., Stegmann, B., Sparks, A., Ansley, T., Van Voorhis, B., 2010. Achievement test performance in children conceived by IVF. Hum Reprod 25, 2605-2611.
Malaspina, D., Reichenberg, A., Weiser, M., Fennig, S., Davidson, M., Harlap, S., Wolitzky, R., Rabinowitz, J., Susser, E., Knobler, H.Y., 2005. Paternal age and intelligence: implications for age-related genomic changes in male germ cells. Psychiatr Genet 15, 117-125.
Manson, J.E., 2008. Prenatal exposure to sex steroid hormones and behavioral/cognitive outcomes. Metabolism 57 Suppl 2, S16-21.
McFadden, D., 1993. A masculinizing effect on the auditory systems of human females having male co-twins. Proc.Natl.Acad.Sci.U.S.A 90, 11900-11904.
McFadden, D., Loehlin, J.C., Pasanen, E.G., 1996. Additional findings on heritability and prenatal masculinization of cochlear mechanisms: click-evoked otoacoustic emissions. Hear Res 97, 102-119.
Medland, S.E., Loehlin, J.C., Martin, N.G., 2008a. No effects of prenatal hormone transfer on digit ratio in a large sample of same- and opposite-sex dizygotic twins. Pers Individ Dif 44, 1225-1234.
Medland, S.E., Loehlin, J.C., Willemsen, G., Hatemi, P.K., Keller, M.C., Boomsma, D.I., Eaves, L.J., Martin, N.G., 2008b. Males do not reduce the fitness of their female co-twins in contemporary samples. Twin.Res.Hum.Genet. 11, 481-487.
Miller, E.M., 1994. Prenatal sex hormone transfer: A reason to study opposite-sex twins. Pers Individ Dif 17, 511-529.
Mortensen, L.H., Diderichsen, F., Davey Smith, G., Nybo Andersen, A.M., 2009. Time is on whose side? Time trends in the association between maternal social disadvantage and offspring fetal growth. A study of 1 409 339 births in Denmark, 1981-2004. J Epidemiol Community Health 63, 281-285.
Oakhill, J.V., Petrides, A., 2007. Sex differences in the effects of interest on boys' and girls' reading comprehension. Br J Psychol 98, 223-235.
Pedersen, C.B., 2011. The danish civil registration system. Scand.J.Public Health 39, 22-25.
Petersen, I., Jensen, V.M., McGue, M., Bingley, P., Christensen, K., 2009. No evidence of genetic mediation in the association between birthweight and academic performance in 2,413 Danish adolescent twin pairs. Twin.Res.Hum.Genet. 12, 564-572.
Petersson, F., Baadsgaard, M., Thygesen, L.C., 2011. Danish registers on personal labour market affiliation. Scand J Public Health 39, 95-98.
Picho, K., Rodriguez, A., Finnie, L., 2013. Exploring the Moderating Role of Context on the Mathematics Performance of Females Under Stereotype Threat: A Meta-Analysis. Journal of Social Psychology 153, 299-333.
Pulkkinen, L., Vaalamo, I., Hietala, R., Kaprio, J., Rose, R.J., 2003. Peer reports of adaptive behavior in twins and singletons: is twinship a risk or an advantage? Twin Res 6, 106-118.
Rammstedt, B., Rammsayer, T.H., 2000. Sex differences in self-estimates of different aspects of intelligence. Pers Individ Dif 29, 869-880.
Resnick, S.M., Gottesman, I.I., McGue, M., 1993. Sensation seeking in opposite-sex twins: an effect of prenatal hormones? Behav.Genet. 23, 323-329.
Ribeiro, D.C., Brook, A.H., Hughes, T.E., Sampson, W.J., Townsend, G.C., 2013. Intrauterine hormone effects on tooth dimensions. J Dent Res 92, 425-431.
Ryan, B.C., Vandenbergh, J.G., 2002. Intrauterine position effects. Neurosci.Biobehav.Rev. 26, 665-678.
Saha, S., Barnett, A.G., Foldi, C., Burne, T.H., Eyles, D.W., Buka, S.L., McGrath, J.J., 2009. Advanced paternal age is associated with impaired neurocognitive outcomes during infancy and childhood. PLoS Med 6, e40.

Sirin, S.R., 2005. Socioeconomic Status and Academic Achievement: A Meta-Analytic Review of Research. Review of Educational Research 75, 417-453.
Skytthe, A., Kyvik, K., Holm, N.V., Vaupel, J.W., Christensen, K., 2002. The Danish Twin Registry: 127 birth cohorts of twins. Twin.Res. 5, 352-357.
Skytthe, A., Ohm, K.K., Vilstrup, H.N., Christensen, K., 2011. The danish twin registry. Scand.J.Public Health 39, 75-78.
Slutske, W.S., Bascom, E.N., Meier, M.H., Medland, S.E., Martin, N.G., 2011. Sensation seeking in females from opposite- versus same-sex twin pairs: hormone transfer or sibling imitation? Behav.Genet. 41, 533-542.
Tapp, A., Maybery, M.T., Whitehouse, A.J., 2011. Evaluating the twin testosterone transfer hypothesis: A review of the empirical evidence. Horm.Behav.
van Anders, S.M., Vernon, P.A., Wilbur, C.J., 2006. Finger-length ratios show evidence of prenatal hormone-transfer between opposite-sex twins. Horm.Behav. 49, 315-319.
vom Saal, F.S., 1989. Sexual differentiation in litter-bearing mammals: influence of sex of adjacent fetuses in utero. J.Anim Sci. 67, 1824-1840.
von Stumm, S., Chamorro-Premuzic, T., Furnham, A., 2009. Decomposing self-estimates of intelligence: structure and sex differences across 12 nations. Br J Psychol 100, 429-442.
Voracek, M., Haubner, T., 2008. Twin-singleton differences in intelligence: a meta-analysis. Psychol.Rep. 102, 951-962.
Voss, W., Jungmann, T., Wachtendorf, M., Neubauer, A.P., 2012. Long-term cognitive outcomes of extremely low-birth-weight infants: the influence of the maternal educational background. Acta Paediatr 101, 569-573.
Vuoksimaa, E., Kaprio, J., Kremen, W.S., Hokkanen, L., Viken, R.J., Tuulio-Henriksson, A., Rose, R.J., 2010. Having a male co-twin masculinizes mental rotation performance in females. Psychol.Sci. 21, 1069-1071.
Wagenaar, K., Ceelen, M., van Weissenbruch, M.M., Knol, D.L., Delemarre-van de Waal, H.A., Huisman, J., 2008. School functioning in 8- to 18-year-old children born after in vitro fertilization. Eur J Pediatr 167, 1289-1295.
Weisglas-Kuperus, N., Hille, E.T., Duivenvoorden, H.J., Finken, M.J., Wit, J.M., van Buuren, S., van Goudoever, J.B., Verloove-Vanhorick, S.P., 2009. Intelligence of very preterm or very low birthweight infants in young adulthood. Arch Dis Child Fetal Neonatal Ed 94, F196-200.
Zuckerman, M., Eysenck, S., Eysenck, H.J., 1978. Sensation seeking in England and America: cross-cultural, age, and sex comparisons. J Consult Clin Psychol 46, 139-149.