



Sex differences in non-verbal and verbal abilities in childhood and adolescence



Teemu Toivainen^a, Kostas A. Papageorgiou^b, Maria G. Tosto^{c,d}, Yulia Kovas^{a,d,e,*}

^a Department of Psychology, Goldsmiths, University of London, UK

^b Faculty of Life Sciences and Computing, School of Psychology, Queen's University Belfast, UK

^c Department of Politics and International Studies, SOAS, University of London, UK

^d Tomsk State University, Tomsk, Russia

^e Social, Developmental and Genetic Psychiatry Centre, King's College, London, UK

ARTICLE INFO

Keywords:

Sex differences

Twins

Testosterone transmission hypothesis

Non-verbal abilities

Verbal abilities

ABSTRACT

Twin research has shown that females with male co-twins perform better than females with female co-twins on mental rotation. This beneficial effect of having a male sibling on spatial ability could be due to in-uterine transmission of testosterone from males to females (the Twin Testosterone Transfer hypothesis, TTT). The present study explored sex differences and the TTT in non-verbal and verbal abilities in a large sample of twins assessed longitudinally at 2, 3, 4, 7, 9, 10, 12, 14 and 16 years of age. Females scored significantly higher than males on both verbal and non-verbal abilities at ages 2, 3 and 4. Males scored significantly higher than females on verbal ability at ages 10 and 12. The effect sizes of all differences were very small. No sex differences in non-verbal or verbal abilities were found at 7, 9, 14 and 16 years of age. No support for the TTT was found at any age. The findings indicate that the twin testosterone transfer effect occurs only for specific cognitive abilities, such as mental rotation.

1. Introduction

Research findings have traditionally indicated that sex differences favoring males appear in non-verbal abilities (e.g. Voyer, Voyer, & Bryden, 1995), while sex differences favoring females appear in verbal abilities (Hyde & Linn, 1988). However, the distinction is more nuanced, which is reflected in the more recent literature on sex differences. For example, within verbal abilities males often do better on verbal analogies (Colom, Contreras, Arend, Leal, & Santacreu, 2004), whereas females outperform males on natural language competencies, reading and writing (Geary, 2010; Stoet & Geary, 2013). Non-verbal abilities refer to 'the skill in representing, transforming, generating, and recalling symbolic, non-linguistic information' (Linn & Petersen, 1985). Verbal abilities refer to measures of language usage, such as grammar, spelling, reading, writing, verbal analogies, vocabulary and oral comprehension (Halpern, 2000). In this paper, we adopt these definitions for verbal and non-verbal skills for convenience.

Although research suggests that sexual differentiation in certain cognitive abilities exists, sex differences in general cognitive ability are overall small, if not negligible (Aluja-Fabregat, Colom, Abad, & Juan-Espinosa, 2000; Colom & Garcia-Lopez, 2002; Colom, Juan-Espinosa, Abad, & Garcia, 2000). Cognitive sex differences also change over time

(Miller & Halpern, 2014). Environmental changes, such as changes in education policies, are likely to have cohort-specific effects (Miller & Halpern, 2014). There is currently mixed evidence on sex differences in verbal and non-verbal abilities. Some studies have shown small to moderate sex differences in non-verbal and verbal ability in childhood and adolescence (e.g. Feldman et al., 2000; Galsworthy, Dionne, Dale, & Plomin, 2000; Goldbeck, Daseking, Hellwig-Brida, Waldmann, & Petermann, 2010; Jensen & Reynolds, 1983; Lynn & Irwing, 2004; Pezzuti & Orsini, 2016). For example, males outperformed females in vocabulary tasks in German and Italian samples of children and adolescents ($d = 0.25$, Goldbeck et al., 2010; $d = 0.10$, Pezzuti & Orsini, 2016). Another study, based on a large sample of pupils aged 11–12 years in the UK, showed that females performed better than males in both verbal ($d = 0.15$) and non-verbal ($d = 0.03$) reasoning (Strand, Deary, & Smith, 2006).

A meta-analysis of studies on sex differences in the Raven's Progressive Matrices showed no significant differences between ages 6 and 14 years; however, males outperformed females at 15 years onwards (Lynn & Irwing, 2004). Additionally, a meta-analysis of studies on sex differences in the Colored Progressive Matrices reported that males outperformed females ($d = 0.21$) in a sample of children age 5 to 11 years (Lynn & Irwing, 2004). Another, more recent, study reported

* Corresponding author at: Department of Psychology, Goldsmiths, University of London, UK.
E-mail address: y.kovas@gold.ac.uk (Y. Kovas).

that males had significantly higher scores in a written mathematics test than females ($d = 0.15$), whereas females performed better in the written Danish ($d = 0.49$) and oral English tests ($d = 0.20$; Ahrenfeldt, Petersen, Johnson, & Christensen, 2015).

A recent meta-analysis of the National Assessment of Educational Progress assessments in the USA reported that males outperformed females in mathematics ($d = 0.10$) and science achievement ($d = 0.13$) in the period of 1990–2011 (Reilly, Neumann, & Andrews, 2015). Furthermore, large sex differences (close to one standard deviation), favoring males, have been consistently documented in some aspects of spatial cognition, such as mental rotation (Hyde, 2005; Voyer et al., 1995). Research has shown that sex differences in mental rotation may emerge from three months of age (Frick & Möhring, 2013; Moore & Johnson, 2011; Quinn & Liben, 2013). This male advantage has attracted much research interest due to its potential link with male proficiency in mathematics (Bull, Davidson, & Nordmann, 2010; Bull, Espy, & Wiebe, 2008) and with under-representation of women in the science, technological, engineering and mathematical (STEM) industries (Ceci, Williams, & Barnett, 2009; Wai, Lubinski, & Benbow, 2009). However, several studies did not report significant sex differences in many verbal or non-verbal ability tasks across development (e.g. Goldbeck et al., 2010; Pezzuti & Orsini, 2016).

Research into the biological factors that may contribute to sex differences in cognition suggests that sex hormones, such as testosterone, can influence cognitive development (Kung, Constantinescu, Browne, Noorderhaven, & Hines, 2016). For example, a recent study reported that testosterone, measured in saliva samples collected at 1–3 months of age, negatively predicted parent-reported expressive vocabulary size at 18–30 months of age in boys and in girls (Kung et al., 2016). The study further showed that postnatal testosterone contributed to sexual differentiation by mediating the effect of sex on expressive vocabulary (Kung et al., 2016). Extraneous administration of testosterone was found to have temporary positive effects on cognition: exogenous administration of testosterone improved spatial ability and verbal memory in older men (Cherrier et al., 2001); as well as spatial ability in female to male transsexuals (Slabbekoorn, van Goozen, Megens, Gooren, & Cohen-Kettenis, 1999). However, a study has shown that naturally occurring fluctuations in testosterone levels could not explain differences in performance on spatial ability tasks, within or between sexes, in a sample of young adults (Puts et al., 2010).

Prenatal exposure to testosterone is argued to have more permanent, organizational effects on brain development in comparison to the effects of postnatal exposure to testosterone (Brizendine, 2007). Research on clinical populations reported higher performance on mental rotation tasks in women with congenital adrenal hyperplasia (CAH)—who are exposed to high levels of prenatal androgens in utero—in comparison to healthy women (Berenbaum, Bryk, & Beltz, 2012). Furthermore, a meta-analysis of studies on the association between CAH and spatial ability found that females with CAH perform better on spatial tasks in comparison to controls (Puts, McDaniel, Jordan, & Breedlove, 2008). Conversely, another study has shown that higher levels of prenatal testosterone in amniotic fluid were negatively correlated with vocabulary size at ages 12 and 24 months (Lutchmaya, Baron-Cohen, & Raggatt, 2001). It is likely that prenatal testosterone exposure affects the development of natural language competencies (e.g., mean length utterance and other measures in which girls often show a modest advantage over boys) rather than other verbal abilities (e.g. vocabulary or grammar - measured in this study). However, it is also possible that prenatal testosterone may affect the brain in a way that subsequent learning is affected, therefore extending its influence to different measures of verbal ability. Other studies have also reported that prenatal testosterone levels are associated with later behavior, physiology and cognition (Berenbaum & Beltz, 2011; Cohen-Bendahan, van de Beek, & Berenbaum, 2005; Hines, 2010).

According to an evolutionary account, prenatal hormonal environment is one biological mechanism influencing sex differences in certain

cognitive abilities based on sexual selection (Geary, 2010). For example, the better developed basic language competencies in females, beneficial for intra-sex competition, could be partly explained by prenatal hormonal effects (Geary, 2010). Similarly, prenatal testosterone could explain to some extent males' better performance in certain spatial abilities, specifically in 3D mental rotation. It is argued that the elaboration of some neurocognitive systems, that have evolved for navigating and tracking movement in the 3-dimensional universe, is more beneficial for males than for females (Geary, 1995). Based on the evolutionary processes, the influence of prenatally transferred testosterone may be presented only for those higher order abilities that are dependent on more basic, prenatally organized abilities. The influence could be seen, for example, if mathematical tasks require more basic visuo-spatial processing ability, or if language processing tasks involve phonetic decoding (Geary, 2014).

Research using twin samples has made a contribution towards understanding the potential effect of prenatal testosterone exposure to sex differences in cognition (Tapp, Maybery, & Whitehouse, 2011). Specifically, two studies have shown that females with twin brothers have an advantage in mental rotation performance over females with twin sisters ($d = 0.40$, Heil, Kavšek, Rolke, Beste, & Jansen, 2011; $d = 0.30$, Vuoksimaa et al., 2010). One explanation for this phenomenon is that females with male co-twins are exposed to higher concentrations of testosterone in utero. Alternatively, the advantage may stem from socialization with a male co-twin that may include activities important for spatial development. One recent study did not support the socialization explanation, but instead provided indirect evidence for the TTT. The study used a sample of non-twin siblings and reported that females with brothers did not outperform females with sisters on the mental rotation test (Frenken et al., 2016). However, a recent twin study did not find evidence for the TTT (Ahrenfeldt et al., 2015). The study found small sex differences in mathematics, English and Danish ($d = 0.15$ – 0.49) in an adolescent sample, but the sex differences were not explained by sex of the co-twin.

To summarize, to date research is inconsistent regarding the role of testosterone transmission in the observed sex differences in cognitive abilities. It is possible that the effects are only present for some abilities and/or at certain time in development. Alternatively, the effects could be very small and therefore require large samples to be detected. The present study uses a large longitudinal twin sample to estimate sex differences in non-verbal and verbal abilities over time, using a variety of measures. The study also investigates the influence of prenatal testosterone on these differences by comparing females with male co-twins to females with female co-twins. Evidence for the TTT hypothesis indicates that the effect may be particularly prominent in visuo-spatial abilities (Tapp et al., 2011). Two previous studies using adult samples have demonstrated that women with twin brothers outperform women with twin sisters in a non-verbal task of mental rotation. It can therefore be expected that evidence for TTT can be found in non-verbal abilities. Previous research with singletons found a negative correlation between prenatal testosterone levels and expressive vocabulary in the early childhood (Kung et al., 2016). We therefore expect that females with male twin brothers will perform worse on verbal ability tasks than females with female twin sisters.

The study addresses three main research questions:

- (i) Are there sex differences in non-verbal and verbal abilities at 2, 3, 4, 7, 9, 10, 12, 14 and 16 years of age?
- (ii) Do females with male co-twins outperform females with female co-twins on non-verbal abilities at 2, 3, 4, 7, 9, 10, 12, 14 or 16 years of age?
- (iii) Do females with male co-twins perform worse than females with female co-twins on verbal abilities at 2, 3, 4, 7, 9, 10, 12, 14 or 16 years of age?

2. Method

2.1. Sample

The Twin Early Development Study (TEDS) sample was used. TEDS is an ongoing longitudinal study that has recruited over 16,000 twin pairs born in England and Wales between 1994 and 1996. Over the years, the sample has shown to be representative of the UK population. Rich behavioral and cognitive data have been collected over many years, including measures of verbal and non-verbal abilities at 2, 3, 4, 7, 9, 10, 12, 14 and 16 years of age (Haworth, Davis, & Plomin, 2013; Kovas, Haworth, Dale, & Plomin, 2007).

In the current study, the following exclusion criteria were applied: participants with major medical or psychiatric conditions; with severe perinatal complications; and who did not have English as their first language. The sample size used in this study varied from 14,187 participants at age 4 years to 4959 participants at age 16.

The TTT hypothesis can be investigated by comparing females with male co-twins to females with female co-twins. Previous research has found small differences in average performances between different sex and zygosity groups (Heil et al., 2011; Vuoksima et al., 2010). Preliminary analyses in our study revealed small, but significant differences in verbal ability between monozygotic males and dizygotic same sex males at ages 2 and 3, and between monozygotic females and dizygotic same sex females at ages 2, 3 and 4. Differences between monozygotic and dizygotic same sex females were also found in non-verbal abilities at ages 2 and 4 (see Tables S1 and S2 in Supplementary material). Due to these zygosity differences, analyses were conducted dividing the sample into six groups based on the participants' sex, zygosity and co-twin's sex. Monozygotic twins were separated into two groups: males (MZm) and females (MZf). Dizygotic twins were separated into four groups: dizygotic males with male co-twins (DZsm), dizygotic males with female co-twins (DZosm), dizygotic females with female co-twins (DZssf) and dizygotic females with male co-twins (DZosf). Information on the sample size for each age group and twin group (e.g. number of dizygotic males with female co-twins at 2 years of age) is presented in Table S3 in the Supplementary material.

2.2. Non-verbal and verbal ability measure

Across development, a variety of age-appropriate non-verbal and verbal measures were administered to the twins taking part in TEDS. Twins were assessed on non-verbal and verbal ability at ages 2, 3, 4, 7, 9, 10, 12, 14 and 16. Two non-verbal and two verbal tests were used at ages 2, 3, 4, 7, 9, 10 and 12. The scores from the two tests were standardized and then averaged to create a non-verbal and verbal composite scores (see Table S4 in the Supplementary material). At ages 14 and 16 both non-verbal and verbal ability were assessed using only one test. The following analyses were run on one randomly selected twin per pair to ensure participants' independence. Additionally, randomly selecting only one twin per pair created a second sample that was used as a replication sample. Effects were considered significant only if they replicated in both halves of the twin sample.

2.2.1. Measures at 2, 3, and 4 years of age

Non-verbal ability at age 2, 3 and 4 was assessed using a version of the Parent Report of Children's Abilities (PARCA) test. PARCA is an hour-long test that measures number, shape, size, conceptual grouping and orientation skills (Fenson et al., 2000; Oliver & Plomin, 2002; Saudino et al., 1998). The PARCA test has been validated in an independent sample (Saudino et al., 1998) as well as in the TEDS sample (Oliver & Plomin, 2002).

Verbal ability at ages 2, 3 and 4 was assessed using the age-appropriate expressive vocabulary and grammar tests based on the MacArthur Communicative Development Inventories (MCDI; Fenson et al., 2000). The MCDI has good internal consistency and test-retest

reliability (Fenson et al., 2000).

2.2.2. Measures at 7 years of age

Assessments were conducted over the phone after parents were sent a booklet containing testing instructions.

Non-verbal ability at age 7 was assessed using the Picture Completion test from the Wechsler's Intelligence Scale for Children (WISC-III-UK; Wechsler, 1992); and the test of Conceptual Grouping from the McCarthy Scales of Children's Abilities (MSCA; McCarthy, 1972). Verbal ability at age 7 was assessed using the similarities and the vocabulary tests that derive from the WISC-III-UK (Wechsler, 1992).

2.2.3. Measures at 9 years of age

Participants filled a booklet containing the tasks under the supervision of the parents (Davis et al., 2008).

Non-verbal ability at age 9 was assessed using the puzzle and shapes test from the Cognitive Abilities Test 3 (CAT3; Smith, Fernandez, & Strand, 2001). Verbal ability at age 9 was assessed using the vocabulary and general knowledge tests (WISC-III-UK; Wechsler, 1992).

2.2.4. Measures at 10 and 12 years of age

Data collection was performed using web-based test batteries.

Non-verbal ability at ages 10 and 12 was assessed using the Picture Completion task from the WISC-III-UK (Wechsler, 1992) and the Raven's Standard Progressive Matrices (Raven, Court, & Raven, 1996). Verbal ability at ages 10 and 12 was assessed using age-appropriate versions of the general knowledge and vocabulary tests that derived from WISC-III-PI (Kaplan, Fein, Kramer, Delis, & Morris, 1998; Wechsler, 1992).

2.2.5. Measures at 14 years of age

Only one of each non-verbal and verbal ability measures were collected at age 14 using web-based tests. Non-verbal ability was measured using the Raven's Progressive Matrices (Raven et al., 1996) and verbal ability with the vocabulary multiple choice test from the WISC-III-PI (Kaplan et al., 1998; Wechsler, 1992).

2.2.6. Measures at 16 years of age

Non-verbal ability at age 16 was assessed using a web-based adaptation of the Raven's Standard and Advanced Progressive Matrices (Raven, Court, & Raven, 1998). Verbal ability at age 16 was assessed using the Mill-Hill vocabulary scale (Raven et al., 1996).

2.3. Statistical analyses

Data were screened for normality and outliers. Outliers were defined as data falling ± 3 standard deviations away from the mean. Removing the outliers did not change the pattern of results.

Means and standard deviations in non-verbal and verbal ability for males and females are presented in Tables S5 and S6 in Supplemental material. For the visual presentation of sex differences in non-verbal and verbal abilities, the means are also plotted in Figs. 1 and 2. Standardized means and standard deviations for six twin groups based on the sex of the co-twin and zygosity are presented in Tables 1 and 2 and plotted in Figs. 3 and 4.

One-way ANCOVAs were used to establish the significant group differences, either between sexes or between sex-by-zygosity twin groups. In all analyses age was used as a covariate to account for the possible effect of age differences (in months). Interactions between the covariate and independent variables were checked for homogeneity of regression slopes. To maintain independence of data, all analyses were conducted on the sample that consisted of one randomly selected twin from each pair. To increase the confidence of the obtained results, all analyses were repeated in the second half of the twin sample. To be considered significant, differences needed to be replicated in the second

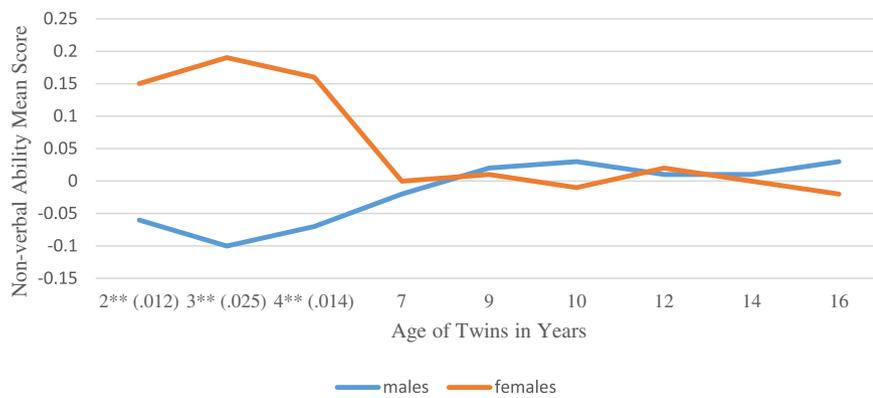


Fig. 1. Standardized mean scores for non-verbal ability for males and females. Effect sizes for the significant sex differences (** $p < 0.01$) are presented in parentheses. The raw scores for the whole sample were standardized, separately for each age cohort. The comparisons between males and females were conducted after randomly selecting one member from each twin pair. Random selection of one twin per pair created two similar singleton samples. Effects were considered significant only if they replicated in both halves of the twin sample.

half of the sample. All analyses were conducted using SPSS for Windows Version 22.

3. Results

3.1. Are there sex differences in non-verbal and verbal abilities across development?

Females scored significantly higher than males in non-verbal ability at ages: 2 ($F(1, 5262) = 59.76, p < 0.01, \eta^2 = 0.012$); 3 ($F(1, 5113) = 125.77, p < 0.01, \eta^2 = 0.025$) and 4 ($F(1, 6766) = 84.02, p < 0.01, \eta^2 = 0.014$). There were no significant sex differences in non-verbal ability at 7, 9, 10, 12, 14 and 16 years of age. Fig. 1 presents standardized means for non-verbal ability, for each age group, for males and females, separately.

Females scored significantly higher than males in verbal ability at ages: 2 ($F(1, 5103) = 135.3, p < 0.01, \eta^2 = 0.032$); 3 ($F(1, 4304) = 86.4, p < 0.01, \eta^2 = 0.019$) and 4 ($F(1, 6315) = 28.57, p < 0.01, \eta^2 = 0.005$). No significant sex differences in verbal ability were found at 7 and 9 years of age. Males scored significantly higher than females in verbal ability at 10 ($F(1, 2462) = 34.42, p < 0.01, \eta^2 = 0.013$) and 12 years of age ($F(1, 4184) = 53.01, p < 0.01, \eta^2 = 0.012$). No significant sex differences in verbal ability were found at 14 and 16 years of age. Fig. 2 presents standardized means for verbal ability, for each age group, for males and females, separately.

3.2. Do females with male co-twins outperform females with female co-twins on non-verbal abilities at 2, 3, 4, 7, 9, 10, 12, 14 or 16 years of age?

To test the effect of TTT on non-verbal ability, comparisons were made between the six sex-by-zygosity groups (MZm, MZf, DZssm, DZssf, DZosm, DZosf). Means and standard deviations for the six twin groups in non-verbal ability at 2, 3, 4, 7, 9, 10, 12, 14 and 16 years of age are

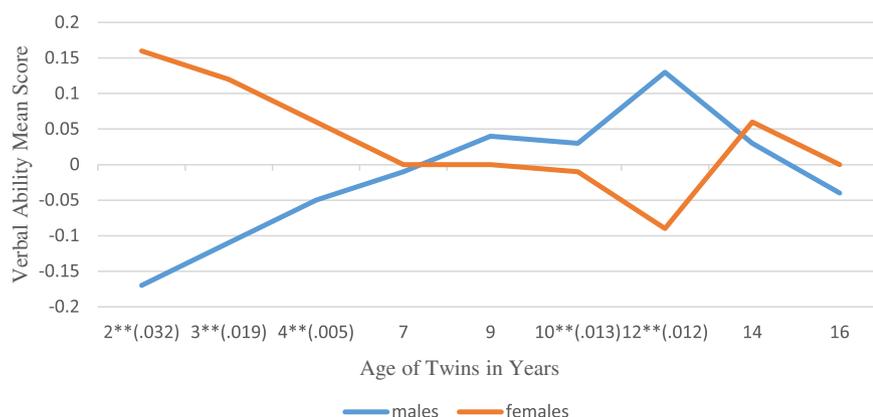


Fig. 2. Standardized mean scores for verbal ability for males and females. Effect sizes for the significant results (** $p < 0.01$) are presented in parentheses. The raw scores for the whole sample were standardized, separately for each age cohort. The comparisons between males and females were conducted after randomly selecting one member from each twin pair. Random selection of one twin per pair created two similar singleton samples. Effects were considered significant only if they replicated in both halves of the twin sample.

presented in Table 1 and plotted in Fig. 3.

Significant differences between the six twin groups in non-verbal ability emerged at ages 2 ($F(5, 5262) = 13.90, p < 0.001, \eta^2 = 0.012$), 3 ($F(5, 5113) = 25.80, p < 0.001, \eta^2 = 0.025$) and 4 ($F(5, 6766) = 26.71, p < 0.001, \eta^2 = 0.014$).

Among the female groups (see Table S7), post hoc tests with Bonferroni corrections showed that at age 4 DZosf (95% CI [0.21, 0.32]) outperformed MZf (95% CI [0.02, 0.12]) in non-verbal ability. However, this difference did not emerge between DZosf and DZssf. The difference between DZosf and MZf is likely to reflect the well documented small advantage of DZ twins over MZ twins, as MZ twins are likely to suffer from more birth complications (Prescott, Johnson, & McArdle, 1999). The absence of differences between DZosf and DZssf suggest that having a twin brother does not lead to advantage in non-verbal ability for females.

There were no twin group differences among males due to the zygosity and the sex of the co-twin. Further group differences were only detected between the sexes. MZm scored significantly lower than MZf and DZssf at ages 2, 3 and 4. Also, DZssm performed significantly worse than MZf and DZssf at ages 2, 3 and 4. Additionally, DZosm scored significantly lower than DZosf at ages 3 and 4. MZm also scored lower than DZosm and DZosf at age 4. The non-verbal ability scores for the six twin groups are plotted in Fig. 3.

3.3. Do females with male co-twins perform worse than females with female co-twins on verbal abilities across development?

To examine group differences in verbal ability, comparisons were made between the six sex-by-zygosity groups (MZm, MZf, DZssm, DZssf, DZosm, DZosf). Means and standard deviations for the six sex-by-zygosity groups in verbal ability at 2, 3, 4, 7, 9, 10, 12, 14 and 16 years of age are presented in Table 2 and plotted in Fig. 4.

Significant differences were found at ages 2 ($F(5, 5350) = 42.90,$

Table 1

Non-verbal ability mean scores and standard deviations for females and males from the mono- and dizygotic same-sex, and dizygotic opposite-sex twin pairs.

Age groups	Females			Males		
	MZf	DZssf	DZosf	MZm	DZssm	DZosm
2-Years	0.12 (0.03)	0.21 (0.03)	0.16 (0.03)	- 0.09 (0.03)	- 0.02 (0.03)	- 0.04 (0.03)
3-Years	0.18 (0.03)	0.25 (0.03)	0.17 (0.03)	- 0.10 (0.03)	- 0.12 (0.03)	- 0.06 (0.03)
4-Years	0.07 (0.03)	0.15 (0.03)	0.27 (0.03)	- 0.14 (0.03)	- 0.05 (0.03)	- 00 (0.03)
7-Years	- 0.01 (0.03)	0.07 (0.03)	0.03 (0.03)	- 0.05 (0.03)	- 0.04 (0.03)	0.02 (0.03)
9-Years	0.01 (0.04)	- 0.04 (0.04)	0.07 (0.04)	- 0.04 (0.04)	0.10 (0.04)	0.00 (0.04)
10-Years	- 0.05 (0.04)	0.01 (0.04)	0.03 (0.04)	0.03 (0.04)	0.13 (0.04)	- 0.05 (0.04)
12-Years	- 0.02 (0.04)	0.04 (0.04)	0.08 (0.04)	- 0.01 (0.03)	0.11 (0.04)	- 0.02 (0.04)
14-Years	- 0.05 (0.05)	0.01 (0.05)	0.08 (0.05)	0.02 (0.04)	0.01 (0.04)	0.02 (0.05)
16-Years	- 0.02 (0.06)	- 0.03 (0.06)	0.01 (0.06)	0.02 (0.04)	0.02 (0.05)	0.04 (0.05)

Note. The means for each age group are based on one randomly selected member from each twin pair. MZm = males from monozygotic twin pair; MZf = females from monozygotic twin pair; DZssm = males from dizygotic same-sex twin pair; DZssf = females from dizygotic same-sex twin pair; DZosm = males from dizygotic opposite-sex twin pair; DZosf = females from dizygotic opposite-sex twin pair.

$p < 0.001$, $\eta^2 = 0.032$); 3 ($F(5, 4634) = 23.04$, $p < 0.001$, $\eta^2 = 0.019$); 4 ($F(5, 6708) = 11.33$, $p < 0.001$, $\eta^2 = 0.005$); 10 ($F(5, 2462) = 8.65$, $p < 0.001$, $\eta^2 = 0.017$); 12 ($F(5, 4184) = 13.11$, $p < 0.001$, $\eta^2 = 0.012$); and 16 ($F(5, 2452) = 2.35$, $p = 0.039$, $\eta^2 = 0.005$).

Among females (see Table S8), the results suggested that verbal ability is unlikely to be related to the co-twins' sex. Post hoc tests with Bonferroni correction showed that at age 12 DZosf (95% CI [- 0.07, 0.08]) scored higher than MZf (95% CI [- 0.24, - 0.11]) in verbal ability. This was contrary to the TTT hypothesis, according to which prenatal testosterone transmitted from a male has a negative impact on his sister's verbal development, as indicated by previous research. However, no difference between DZosf and DZssf was found, suggesting that the DZosf advantage over MZf is likely to reflect the well documented small advantage of DZ twins over MZ twins, described earlier. DZssf at age 4 scored higher than MZf, providing further support for this explanation. We also analyzed separately the tasks for grammar and vocabulary at ages 2, 3 and 4 (see Table S9 for details). The results were similar as for the verbal composite. Post hoc tests showed no significant differences between DZosf and both same-sex female groups and therefore did not give evidence for TTT.

Post hoc comparisons with Bonferroni corrections showed no differences in verbal ability between the three male twin groups, they only emerged between twin groups of different sexes. MZm scored significantly lower in verbal ability at ages 2 and 3 in comparison to all three female groups (MZf, DZssf, DZosf). Also, DZssm scored lower than both MZf and DZssf at ages 2 and 3. Twin group differences were also found between twins from opposite-sex pairs: DZosf performed better than DZosm at ages 2 and 3. At age 4, MZm scored lower than DZssf and DZosf, and DZssm performed worse than DZssf. At age 10, there was

only one group difference: DZssm performed better than MZf. At age 12, MZm performed better than MZf. Additionally, at age 12, DZssm performed better than both MZf and DZssf. Even if overall ANCOVA at age 16 indicated some differences between the twin groups, these differences were so small that they were not statistically significant after the corrections of the alpha levels for multiple comparisons. The verbal ability scores for the six twin groups are plotted in Fig. 4.

4. Discussion

Accumulating evidence suggests that sex differences in general intelligence are negligible (Aluja-Fabregat et al., 2000; Colom & Garcia-Lopez, 2002; Colom et al., 2000). However, sex differences persist in certain, specific cognitive domains such as in visuo-spatial ability (see for example, Frenken et al., 2016; Miller & Halpern, 2014). The present study provides new insights into sex differences in cognition by exploring non-verbal and verbal abilities across development in a large UK representative sample. The twin sample also allowed to test the effect of prenatal twin testosterone transfer on verbal and non-verbal abilities from males to their female co-twins.

4.1. Sex differences

Females scored higher than males on both verbal and non-verbal abilities at 2, 3 and 4 years of age. These findings indicate that females have an advantage over males in the first 4-years of life in most cognitive domains. These sex differences in pre-pubertal cognitive abilities could be due to girls' overall faster development at this stage, potentially leading to advantage in a number of traits. For example, sex differences have been reported in brain maturation processes in a

Table 2

Verbal ability mean scores and standard deviations for females and males from the mono- and dizygotic same-sex, and dizygotic opposite-sex twin pairs.

Age groups	Females			Males		
	MZf	DZssf	DZosf	MZm	DZssm	DZosm
2-Years	0.14 (0.03)	0.25 (0.03)	0.12 (0.03)	- 0.30 (0.03)	- 0.15 (0.03)	- 0.05 (0.03)
3-Years	0.07 (0.03)	0.19 (0.03)	0.12 (0.03)	- 0.21 (0.03)	- 0.11 (0.03)	- 0.02 (0.03)
4-Years	- 0.02 (0.02)	0.10 (0.02)	0.12 (0.03)	- 0.11 (0.03)	- 0.03 (0.03)	- 0.01 (0.03)
7-Years	- 0.04 (0.03)	0.01 (0.03)	0.06 (0.03)	- 0.05 (0.03)	- 0.05 (0.03)	0.07 (0.03)
9-Years	- 0.03 (0.03)	- 0.01 (0.04)	0.05 (0.04)	- 0.02 (0.04)	0.10 (0.04)	0.06 (0.04)
10-Years	- 0.14 (0.04)	- 0.07 (0.04)	- 0.02 (0.04)	0.06 (0.04)	- 0.20 (0.04)	0.08 (0.04)
12-Years	- 0.17 (0.03)	- 0.04 (0.04)	0.01 (0.04)	0.11 (0.04)	0.16 (0.04)	0.14 (0.04)
14-Years	0.01 (0.04)	0.11 (0.04)	0.10 (0.04)	0.01 (0.05)	0.06 (0.05)	0.05 (0.05)
16-Years	- 0.08 (0.04)	- 0.01 (0.04)	0.12 (0.05)	- 0.07 (0.05)	- 0.05 (0.05)	- 0.01 (0.05)

Note. The means for each age group are based on one randomly selected member from each twin pair. MZm = males from monozygotic twin pair; MZf = females from monozygotic twin pair; DZssm = males from dizygotic same-sex twin pair; DZssf = females from dizygotic same-sex twin pair; DZosm = males from dizygotic opposite-sex twin pair; DZosf = females from dizygotic opposite-sex twin pair.

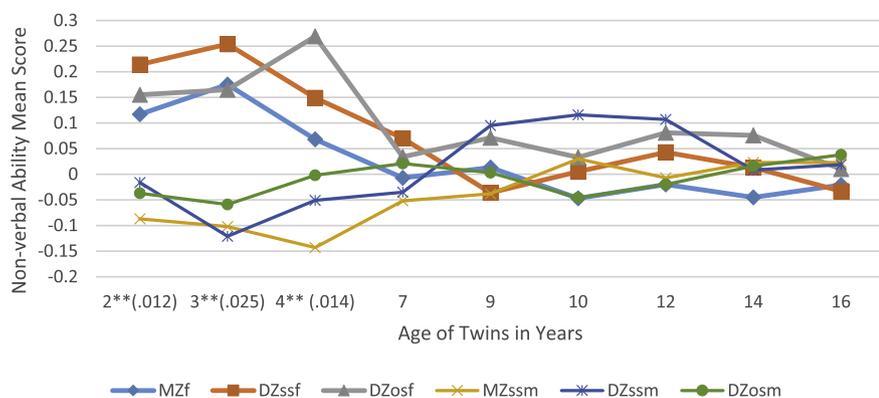


Fig. 3. Non-verbal ability mean scores for males and females from monozygotic same-sex, dizygotic same-sex and dizygotic opposite-sex twin pairs. MZm = males from monozygotic twin pair; MZf = females from monozygotic twin pair; DZssm = males from dizygotic same-sex twin pair; DZssf = females from dizygotic same-sex twin pair; DZosm = males from dizygotic opposite-sex twin pair; DZosf = females from dizygotic opposite-sex twin pair. The effect sizes are presented in parentheses. The means for each age cohort are based on one randomly selected member from each twin pair. Effects were considered significant only if they replicated in both halves of the twin sample. $**p < 0.01$.

sample of pre-pubertal and adolescent participants using a cross-sectional design (De Bellis et al., 2001). Differential brain development could therefore influence sex differences in non-verbal and verbal abilities (Galsworthy et al., 2000), but no conclusion on the causality can be made without replicating the findings in a longitudinal sample. At age 7, there were no statistically significant sex differences; and at ages 10 and 12, males scored higher than females on verbal ability. These results showed that from 4-years onwards males are able to “catch up” with their female peers and to outperform them on verbal ability from age 10 to 12. However, at 14 and 16 years of age, no statistically significant sex differences in non-verbal and verbal abilities were found. Sex differences in verbal ability are not clear-cut. For example, it has been reported that males perform better than females on verbal analogies (Colom et al., 2004); these findings might explain the higher performance of males in verbal ability measures at ages 10 and 12 in the present study.

Our findings contradict some previous studies that reported significant sex differences in some of the measures (see for example a meta-analysis by Lynn & Irwing, 2004) that were employed in the current study. However, there are differences between the samples. For example, the meta-analysis on Progressive Matrices, that found significant sex differences at age 16, was based on international data, collected between the years 1939–2002 (Lynn & Irwing, 2004). During this period, several sociodemographic changes, such as women's improved access to education, have occurred, likely influencing cognitive sex differences (Halpern, 2014). The findings of this study could reflect differences in the characteristics of the samples between our study and previous research (e.g. our study used a large representative sample to explore sex differences in cognitive abilities); or that sex differences in certain cognitive abilities (e.g. general cognitive ability) are now negligible. More longitudinal research is necessary in order to understand the mechanisms underlying the observed dynamics of sex differences across development. For example, why do girls show on average better performance in verbal ability in the early years whereas at ages 10 and

12 boys perform slightly better?

Overall, the results of this study suggest that sex differences in most cognitive abilities are small or non-existent. This is consistent with behavioral genetic research that consistently finds negligible sex differences in the genetic and environmental etiology of individual differences in cognitive abilities and academic achievement (e.g., Kovas et al., 2007). In combination with previous research that finds sizeable sex differences for visuo-spatial ability, such as mental rotation (Voyer et al., 1995), and in early reading comprehension that relies on phonetic decoding (Hyde & Linn, 1988), the results suggest that sex differences are limited to these specific skills (due to evolutionary pressures). Some small and inconsistent sex differences observed in academic achievement and other abilities may partly reflect the ‘washed out’ effects of these evolutionary processes, to the extent that they contribute to other abilities. In our future research, we plan to test the TTT further employing measures that have shown moderate to large sex differences in previous studies (see for example, Voyer et al., 1995) and examining differences in genetic and environmental etiologies for males and females on these traits.

4.2. The Twin Testosterone Transfer hypothesis

Two previous studies have shown that females with twin brothers have an advantage in mental rotation performance over females with twin sisters ($d = 0.40$, Heil et al., 2011; $d = 0.30$, Vuoksimaa et al., 2010). However, a recent twin study that explored the testosterone transfer effect in Mathematics, English and Danish found no support for the TTT hypothesis in an adolescent sample (Ahrenfeldt et al., 2015). It is possible that prenatal exposure to testosterone influences females' performance on mental rotation, but not on other abilities. In line with this, we found no evidence for the TTT on verbal and non-verbal cognitive abilities from age 2 to 16. Females with male co-twins did not show greater performance in non-verbal ability or weaker performance in verbal ability than females with DZ female co-twins at any age. The

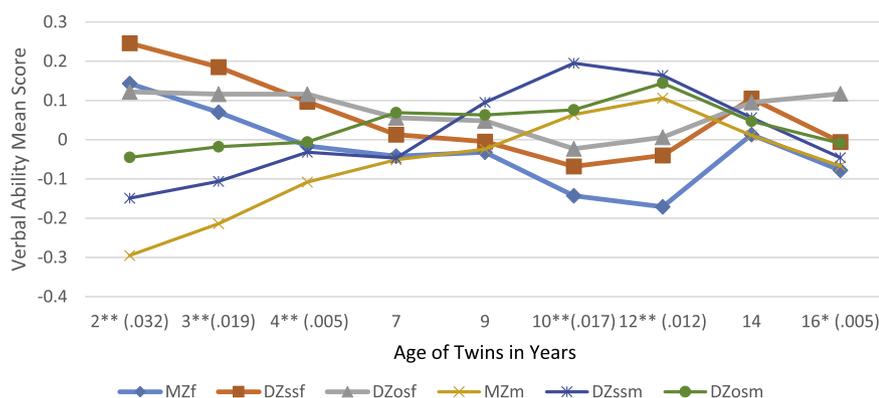


Fig. 4. Verbal ability mean scores for males and females from monozygotic same-sex, dizygotic same-sex and dizygotic opposite-sex twin pairs. MZm = males from monozygotic twin pair; MZf = females from monozygotic twin pair; DZssm = males from dizygotic same-sex twin pair; DZssf = females from dizygotic same-sex twin pair; DZosm = males from dizygotic opposite-sex twin pair; DZosf = females from dizygotic opposite-sex twin pair. The effect sizes are presented in parentheses. The means for each age cohort are based on one randomly selected member from each twin pair. Effects were considered significant only if they replicated in both halves of the twin sample. $**p < 0.01$; $*p < 0.05$.

results from this study are in line with an evolutionary account, according to which the biologically influenced sex differences would appear in abilities that would have been evolutionarily involved in intra-sex competition (Geary, 2010). The influence of prenatal testosterone may only be present for higher order abilities that rely on more basic, prenatally organized abilities, such as visuo-spatial abilities or basic language competencies (Geary, 2010; Geary, 2014).

4.3. Limitations

The tests used in this study varied in content across development and potentially tapped into (partially) different aspects of non-verbal and verbal abilities. While the measures were age-appropriate, it is possible that the observed sex differences could reflect test-specific effects, or differences between cohorts. It is also possible that other tests (e.g., mental rotation) show sex differences to which TTT makes a contribution. The present study utilized the existing data on non-verbal and verbal abilities, collected from a large, longitudinal twin sample. It was therefore not possible to include additional measures, leading to the limitation of not evaluating measures that have shown in the past moderate to large sex differences, such as the mental rotation test (e.g. Voyer et al., 1995). We plan to investigate this in our future work. Nevertheless, this is the first study to explore longitudinally (as opposed to using adult samples assessed once) sex differences and the TTT in various (as opposed to one) cognitive abilities in a large representative sample.

4.4. Conclusion

The results showed negligible sex differences in non-verbal and verbal ability across development. At most, sex explained 3% of the variation in non-verbal, and 2% of the variation in verbal ability. No support for the Twin Testosterone Transfer hypothesis was found. The results indicate that the testosterone transfer may only be relevant to tests that show large and robust sex difference, such as mental rotation, and are thus more likely to be sensitive to androgen. However, before such conclusion can be reached, more research is needed to test whether the effects are also present for other aspects of spatial cognition.

Acknowledgements

We gratefully acknowledge the on-going contribution of the participants in the Twins Early Development Study (TEDS) and their families. TEDS is supported by a program grant from the UK Medical Research Council (G0901245; previously G0500079). YK's research has been supported by the grant from the Russian Science Foundation (project 14-48-00043) to Tomsk State University. TT is supported by the Economic and Social Research Council [grant number ES/J500124/1].

Author note

Development of study design was performed by T. Toivainen, K. A. Papageorgiou, M. Grazia Tosto and Y. Kovas. Data collection was part of the longitudinal TEDS, directed by R. Plomin. T. Toivainen, M. Grazia Tosto conducted data analyses. All authors worked on the manuscript write-up. The final version of the manuscript has been approved for submission by all authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.intell.2017.07.007>.

References

- Ahrenfeldt, L., Petersen, I., Johnson, W., & Christensen, K. (2015). Academic performance of opposite-sex and same-sex twins in adolescence: A Danish national cohort study. *Hormones and Behavior*, *69*, 123–131.
- Aluja-Fabregat, A., Colom, R., Abad, F., & Juan-Espinosa, M. (2000). Sex differences in general intelligence defined as *g* among young adolescents. *Personality and Individual Differences*, *28*, 813–820.
- Berenbaum, S. A., & Beltz, A. M. (2011). Sexual differentiation of human behaviour: Effects of prenatal and pubertal organizational hormones. *Frontiers in Neuroendocrinology*, *32*(2), 183–200.
- Berenbaum, S. A., Bryk, K. L. K., & Beltz, A. M. (2012). Early androgen effects on spatial and mechanical abilities: Evidence from congenital adrenal hyperplasia. *Behavioral Neuroscience*, *126*(1), 86–96.
- Brizendine, L. (2007). *The female brain*. London: Bantam Books.
- Bull, R., Espy, K. A., & Wiebe, S. A. (2008). Short-term memory, working memory, and executive functioning in preschoolers: Longitudinal predictors of mathematical achievement at age 7 years. *Developmental Neuropsychology*, *33*(3), 205–228.
- Bull, R., Davidson, W. A., & Nordmann, E. (2010). Prenatal testosterone, visual-spatial memory, and numerical skills in young children. *Learning and Individual Differences*, *20*(3), 246–250.
- Ceci, S. J., Williams, W. M., & Barnett, S. M. (2009). Women's underrepresentation in science: Sociocultural and biological considerations. *Psychological Bulletin*, *135*(2), 218.
- Cherrier, M. M., Asthana, S., Plymate, S., Baker, L., Matsumoto, A. M., Peskind, E., ... Craft, S. (2001). Testosterone supplementation improves spatial and verbal memory in healthy older men. *Neurology*, *57*(1), 80–88.
- Cohen-Bendahan, C. C. C., van de Beek, C., & Berenbaum, S. A. (2005). Prenatal sex hormones effects on child and adult sex-typed behaviour: Methods and findings. *Neuroscience and Biobehavioral Reviews*, *29*, 353–384.
- Colom, R., & Garcia-Lopez, O. (2002). Sex differences in fluid intelligence among high school graduates. *Personality and Individual Differences*, *32*, 445–451.
- Colom, R., Contreras, M. J., Arend, I., Leal, O. G., & Santacreu, J. (2004). Sex differences in verbal reasoning are mediated by sex differences in spatial ability. *The Psychological Record*, *54*(3), 365–372.
- Colom, R., Juan-Espinosa, M., Abad, F., & Garcia, L. F. (2000). Negligible sex differences in general intelligence. *Intelligence*, *28*(1), 57–68.
- Davis, O. S. P., Kovas, Y., Harlaar, N., Busfield, P., McMillan, A., Frances, J., ... Plomin, R. (2008). Generalist genes and the internet generation: Etiology of learning abilities by web testing at age 10. *Genes, Brain and Behavior*, *7*, 455–462.
- De Bellis, M. D., Keshavan, M. S., Beers, S. R., Hall, J., Frustaci, K., Masalehdan, A., ... Boring, A. M. (2001). Sex differences in brain maturation during childhood and adolescence. *Cerebral Cortex*, *11*(6), 552–557.
- Feldman, H. M., Dollaghan, C. A., Campbell, T. F., Kurs-Lasky, M., Janosky, J. E., & Paradise, J. L. (2000). Measurement properties of the MacArthur Communicative Development Inventories at ages one and two years. *Child Development*, *71*(2), 310–322.
- Fenson, L., Bates, E., Dale, P., Goodman, J., Reznick, J. S., & Thal, D. (2000). Measuring variability in early child language: Don't shoot the messenger. *Child Development*, *71*(2), 323–328.
- Frenken, H., Papageorgiou, K. A., Tikhomirova, T., Malykh, S., Tosto, M. G., & Kovas, Y. (2016). Siblings' sex is linked to mental rotation performance in males but not females. *Intelligence*, *55*, 38–43.
- Frick, A., & Möhring, W. (2013). Mental object rotation and motor development in 8- and 10-month-old infants. *Journal of Experimental Child Psychology*, *115*(4), 708–720.
- Galsworthy, M. J., Dionne, G., Dale, P. S., & Plomin, R. (2000). Sex differences in early verbal and non-verbal cognitive development. *Developmental Science*, *3*(2), 206–215.
- Geary, D. C. (1995). Sexual selection and sex differences in spatial cognition. *Learning and Individual Differences*, *7*(4), 289–301.
- Geary, D. C. (2010). *Male, female: The evolution of human sex differences*. Washington, DC: American Psychological Association.
- Geary, D. C. (2014). Evolved sex differences in modern context. *New Male Studies: An International Journal*, *3*(5), 20.
- Goldbeck, L., Daseking, M., Hellwig-Brida, S., Waldmann, H. C., & Petermann, F. (2010). Sex differences on the German Wechsler intelligence test for children (WISC-IV). *Journal of Individual Differences*, *31*, 22–28.
- Halpern, D. F. (2000). *Sex differences in cognitive abilities* (3rd ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Haworth, C. M. A., Davis, O. S. P., & Plomin, R. (2013). Twins Early Development Study (TEDS): A genetically sensitive investigation of cognitive and behavioural development from childhood to young adulthood. *Twin Research and Human Genetics*, *16*(1), 117–125.
- Heil, M., Kavšek, M., Rolke, B., Beste, C., & Jansen, P. (2011). Mental rotation in female fraternal twins: Evidence for intra-uterine hormone transfer? *Biological Psychology*, *86*, 90–93.
- Hines, M. (2010). Sex-related variation in human behaviour and the brain. *Trends in Cognitive Sciences*, *14*(10), 448–456.
- Hyde, J. S. (2005). The gender similarities hypothesis. *American Psychologist*, *60*(6), 581.
- Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin*, *104*(1), 53–69.
- Jensen, A. R., & Reynolds, C. R. (1983). Sex differences on the WISC-R. *Personality and Individual Differences*, *4*(2), 223–226.
- Kaplan, E., Fein, D., Kramer, J., Delis, D., & Morris, R. (1998). WISC-III as a process. *Instrument (WISC-III-PI)*. New York: The Psychological Corporation.
- Kovas, Y., Haworth, C. M. A., Dale, P. S., & Plomin, R. (2007). The genetic and

- environmental origins of learning abilities and disabilities in the early school years. *Monographs of the Society for Research in Child Development*, 72, vii–160.
- Kung, K. T., Constantinescu, M., Browne, W. V., Noorderhaven, R. M., & Hines, M. (2016). No relationship between early postnatal testosterone concentrations and autistic traits in 18 to 30-month-old children. *Molecular Autism*, 7(1), 1.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterisation of sex differences in spatial ability: A meta-analysis. *Child Development*, 56(6), 1479–1498.
- Lutchmaya, S., Baron-Cohen, S., & Raggatt, P. (2001). Foetal testosterone and vocabulary size in 18- and 24-month-old infants. *Infant Behavior and Development*, 24(4), 418–424.
- Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, 32(5), 481–498.
- McCarthy, D. (1972). *McCarthy scales of children's abilities*. New York: The Psychological Corporation.
- Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in cognitive sciences*, 18(1), 37–45.
- Moore, D. S., & Johnson, S. P. (2011). Mental rotation of dynamic, three-dimensional stimuli by 3-month-old infants. *Infancy*, 16(4), 435–445.
- Oliver, B. R., & Plomin, R. (2002). Twins' Early Development Study (TEDS): A multivariate, longitudinal genetic investigation of language, cognition and behaviour problems from childhood through adolescence. *Twin Research and Human Genetics*, 10(1), 96–105.
- Pezzuti, L., & Orsini, A. (2016). Are there sex differences in the Wechsler intelligence scale for children—fourth edition? *Learning and Individual Differences*, 45, 307–312.
- Prescott, C. A., Johnson, R. C., & McArdle, J. J. (1999). Chorion type as a possible influence on the results and interpretation of twin study data. *Twin Research*, 2(04), 244–249.
- Puts, D. A., McDaniel, M. A., Jordan, C. L., & Breedlove, S. M. (2008). Spatial ability and prenatal androgens: Meta-analyses of congenital adrenal hyperplasia and digit ratio (2D:4D) studies. *Archives of Sexual Behavior*, 37(1), 100–111.
- Puts, D. A., Cárdenas, R. A., Bailey, D. H., Burriss, R. P., Jordan, C. L., & Breedlove, S. M. (2010). Salivary testosterone does not predict mental rotation performance in men or women. *Hormones and Behavior*, 58(2), 282–289.
- Quinn, P. C., & Liben, L. S. (2013). A sex difference in mental rotation in infants: Convergent evidence. *Infancy*, 19(1), 103–116.
- Raven, J. C., Court, J. H., & Raven, J. (1996). *Manual for Raven's progressive matrices and vocabulary scales*. Oxford: Oxford University Press.
- Raven, J. C., Court, J. H., & Raven, J. (1998). *Manual for Raven's advanced progressive matrices*. Oxford: Oxford Psychologists Press.
- Reilly, D., Neumann, D. L., & Andrews, G. (2015). Sex differences in mathematics and science achievement: A meta-analysis of National Assessment of Educational Progress assessments. *Journal of Educational Psychology*, 107(3), 645.
- Saudino, K. J., Dale, P. S., Oliver, B., Petrill, S. A., Richardson, V., Rutter, M., ... Plomin, R. (1998). The validity of parent-based assessment of the cognitive abilities of 2-year-olds. *British Journal of Developmental Psychology*, 16, 349–363.
- Slabbekoorn, D., van Goozen, S. H. M., Megens, J., Gooren, L. J. G., & Cohen-Kettenis, P. T. (1999). Activating effects of cross-sex hormones on cognitive functioning: A study of short-term and long-term hormone effects in transsexuals. *Psychoneuroendocrinology*, 24(4), 423–447.
- Smith, P., Fernandez, C., & Strand, S. (2001). *Cognitive abilities test 3 (CAT3)*. Windsor: nferNELSON.
- Stoet, G., & Geary, D. C. (2013). Sex differences in mathematics and reading achievement are inversely related: Within-and across-nation assessment of 10 years of PISA data. *PLoS One*, 8(3), e57988.
- Strand, S., Deary, I. J., & Smith, P. (2006). Sex differences in cognitive abilities test scores: A UK national picture. *British Journal of Educational Psychology*, 76(3), 463–480.
- Tapp, A. L., Maybery, M. T., & Whitehouse, A. J. O. (2011). Evaluating the twin testosterone hypothesis: A review of the empirical evidence. *Hormones and Behavior*, 60, 713–722.
- Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, 117(2), 250–270.
- 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. *Psychological Science*, 21(8), 1069–1071.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817.
- Wechsler, D. (1992). *Wechsler intelligence scale for children. U.K. (WISC-III-UK) manual* (3rd ed.). London: The Psychological Corporation.