

Genetic and environmental influences on spatial reasoning: A meta-analysis of twin studies[☆]

Michael J. King^{*}, David P. Katz, Lee A. Thompson, Brooke N. Macnamara

Case Western Reserve University, Cleveland, OH, USA

ARTICLE INFO

Keywords:

Spatial reasoning
Behavioral genetics
Meta-analysis
Cognitive development
Intelligence

ABSTRACT

Behavioral genetic approaches, such as comparing monozygotic and dizygotic twins, are often used to evaluate the extent to which variations in human abilities are the result of genetic (heritable), shared environmental, and non-shared environmental factors. We conducted a meta-analysis on the twin study literature—comparing monozygotic and dizygotic twins—to provide clarity and a general consensus regarding the extent to which genetic and environmental factors contribute to variation in spatial ability. Consistent with previous work, we found that spatial ability is largely heritable (meta-analytic $\bar{a} = .61$; 95% CI [.55, .66]), with non-shared environmental factors having a substantial impact (meta-analytic $\bar{e} = .43$; 95% CI [.38, .49]), and shared environmental factors having very little impact (meta-analytic $\bar{c} = .07$; 95% CI [.05, .10]). Moderator analyses were performed to establish if spatial ability type, sex, or age impacted the explanatory power of genetics or environmental factors. These effects did not differ significantly by sex or spatial ability type. However, the influence of shared environments did significantly differ depending on age. This result was driven by the youngest age group (ages 4–15) demonstrating relatively high amounts of shared environmental influence ($c = .15$, 95% CI [.10, .20]) compared with the other age groups ($cs = .00$ – $.07$).

Spatial ability is a category of reasoning skills that refers to the capacity to think about objects in both two and three dimensions, and to draw conclusions about those objects from limited information. Spatial ability is an important area of study because of its association with the functioning of other high-level cognitive skills, such as logical reasoning, memory retrieval, verbal skills, mathematical skills, problem-solving, and reading (Lohman, 1996; Spearman & Jones, 1950). For example, in a study investigating relationships between students' spatial ability, logical thinking, and mathematical achievement, Bektasli (2006) found that logical thinking and mathematical achievement were significantly correlated with spatial ability. Spatial ability is also associated with mathematical thinking (i.e., the ability to use sophisticated strategies effectively on mathematical problems) ($rs = .43$ – $.74$; see e.g., Casey, Nuttall, Pezaris, & Benbow, 1995; Kyttälä & Lehto, 2008; Laski, et al., 2013; Mix et al., 2016; Reuhkala, 2001), mathematic fluency (i.e., the ability to answer calculation problems under a time limit) ($r = .74$; Hart, Petrill, Thompson, & Plomin, 2009), and non-numerical math abilities (e.g., the number line task) ($r = .60$; Tosto et al., 2014). Spatial ability has also been shown to predict geographical skills, specifically on performance in map-use tasks involving route-panning, visual search, symbol identification, and left/right

orientation ($\eta_p^2 = 0.42$; Gilmartin & Patton, 1984).

Spatial ability is also a vital part of success in many disciplines. Fields of study such as engineering, chemistry, and physics all use spatial visualization to help individuals complete tasks and solve problems. For example, chemistry students use 3D modeling kits to create physical replicas of molecular models and chemical reactions that they are retrieving from a memory of a 2D image (Pribyl & Bodner, 1987). Similarly, dentistry students develop spatial mental models of the 3D structure of teeth, which improves their ability to mentally maintain and manipulate representations of these specific structures (Hegarty, Keehner, Khooshabeh, & Montello, 2009). In both of these examples, mental rotation is needed to consistently and accurately complete the tasks.

Tests of spatial ability have also been shown to predict initial performance in physics coursework ($\Phi = .55$ – $.76$); Kozhevnikov, Motes, & Hegarty, 2007). Here, spatial visualization skills were used to solve kinematics problems by predicting the two-dimensional motion of an object, translating from one frame of reference to another, and interpreting kinematics graphs (Kozhevnikov et al., 2007). As another example, Hambrick et al. (2011) displayed that tests of spatial ability predicted performance among novice geologists on a bedrock-mapping

[☆] This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors

^{*} Corresponding author at: Department of Psychological Sciences, 11220 Bellflower Road, Case Western Reserve University, Cleveland, OH 44106, United States.

E-mail addresses: mjk217@case.edu (M.J. King), dpk44@case.edu (D.P. Katz), lat@case.edu (L.A. Thompson), bnm24@case.edu (B.N. Macnamara).

task ($r = .28$). Across all of the science, technology, engineering, and mathematics (STEM) fields, spatial ability plays a critical role in developing expertise. As such, it has been suggested that spatial ability be included in modern talent searches to better identify adolescents with potential for STEM who are currently being missed (Wai et al., 2009). Importantly, the likelihood of earning an advanced degree in STEM areas has been shown to increase as a function of spatial ability (< 10% of those holding STEM PhDs were below the top quartile in spatial ability during adolescence), even while controlling for intelligence (Wai et al., 2009).

The positive associations between spatial abilities and cognitive and academic skills are well-known. However less well known are the sources of variation in spatial abilities. Investigating the extent to which genetic and environmental factors contribute to variation in spatial ability provides insight into the sources of these individual differences. Twin studies allow us to examine these two contributions. However, the twin study literature on spatial reasoning abilities reports highly variable estimates of genetic and environmental influences on spatial abilities. For example, many studies have found additive genetic influences accounting for more variability in spatial reasoning ability as compared to environmental influences ($a^2 = .58-.94$) (Chow, Epp, Lieblich, Barha, & Galea, 2013; Johnson & Bouchard, 2007; Johnson et al., 2007; Mosing, Madison, Pedersen, Kuja-Halkola, & Ullén, 2014; Reynolds, Finkel, Gatz, & Pedersen, 2002; Shakeshaft & Plomin, 2015; Tucker-Drob, Briley, Engelhardt, Mann, & Harden, 2016). There is also a considerable amount of research that has found additive genetic influences, while considerable, account for less variability in spatial reasoning ability as compared to environmental influences ($a^2 = .12-.32$) (Malanchini et al., 2016; McClearn et al., 1997; Mosing et al., 2012; Petrill et al., 1998; Tosto et al., 2014; Tucker-Drob, Reynolds, Finkel, & Pedersen, 2014). Thus, a meta-analysis is needed to synthesize the twin study findings and examine potential sources of heterogeneity.

The primary goal of the current study is to provide clearer estimates of genetic and environmental contributions toward the etiology of individual differences in spatial ability. The second goal is to examine whether variation in heritability and environmental influence on spatial ability is partly systematic, varying depending on other factors. Specifically, we examine whether genetic and environmental contributions differ depending on the type of spatial ability under examination, sex, and developmental stage (i.e., age of individuals). We do this by conducting moderator analyses.

1. Types of spatial abilities

Spatial ability is defined, conceptualized, and studied, not as one single ability, but as a set of abilities that each represent an aspect of spatial skill (Thurstone, 1950; Lohman, 1979; Linn and Petersen, 1985; Hegarty & Waller, 2004, 2005). One primary spatial factor is mental rotation. Mental rotation is defined as the ability to manipulate mental representations of two-dimensional and three-dimensional objects by turning or spinning the representation (Shepard & Metzler, 1971). An example of mental rotation is being able to determine whether an original figure matches a set of similar figures that have been rotated in various orientations.

Spatial visualization is another primary spatial ability. Spatial visualization is conceptualized as the ability to mentally represent visual appearances of an object and the spatial relations between the locations of the objects or movements (Linn & Petersen, 1985). An example of spatial visualization is determining how to fit together blocks with different shapes to make a specific design.

Matrix reasoning ability is a third primary spatial ability. Matrix reasoning is an abstract spatial reasoning skill designed to tap one's

ability to find visuo-spatial patterns. Matrix reasoning is typically measured using a Raven's Progressive Matrices test, which is also used as a non-verbal estimate of fluid intelligence.

Studies examining the etiology of spatial ability typically do not include measures representing all 3 of these spatial factors; therefore, measurement invariance must be considered when comparing results across studies. It is also important to consider how genes and the environment might differentially influence individual variation in each of the spatial factors. In the current study, we test type of spatial ability as a potential moderator, to understand whether the contributions between gene and environment differ between mental rotation, spatial visualization, and matrix reasoning.

2. Sex differences

Variability in spatial ability has been evident in research for more than a century. Most notably, group differences in males compared to females often acknowledge a male advantage in spatial skills (Battista, 1990; Casey, Nuttall, & Pezaris, 2001; Geary, Hamson, & Hoard, 2000; Keig et al., 1993; Voyer et al., 1995). Mental rotation ability specifically has been consistently shown to produce one of the largest and most consistent sex differences, in favor of males, in the cognitive literature (Parsons et al., 2004; Linn & Petersen, 1985; Vandenberg and Kuse, 1979; Voyer et al., 1995). For example, performance on the Mental Rotation Task has resulted in significant mean sex differences ($d = 0.90, p = .04$) (Parsons et al., 2004). Additionally, Voyer et al. (1995) found a large effect size for sex differences in mental rotation tasks ($d = 0.56, p < .05$). Though this difference does not typically appear in matrix reasoning tasks (Eysenck, 1981; Court, 1983; Lynn, 1994; Mackintosh, 1996; Jensen, 1998; Lynn, 1998, 1999).

The etiology of sex differences has been a persistent topic of discussion in psychological research. Some of the early findings of the male advantage were explained in terms of underlying biological (genetic) factors (Bock & Vandenberg, 1968; McGee, 1979), suggesting that spatial ability is highly heritable (Wilson & Vandenberg, 1978). Specifically, Stafford (1961) proposed that an X-linked recessive gene explained sex differences in spatial abilities. That is, that the X-linked gene enhanced spatial abilities, and, because it was recessive rather than dominant, affected more males than females. However, later studies did not support this hypothesis (see McGee, 1979 for a discussion). More recent research has suggested that heightened spatial ability may be driven by testosterone levels, as evidenced by the Twin Testosterone Transfer hypothesis (TTT) (Vuoksimaa et al., 2010). According to TTT, this advantage is due to in-uterine transmission of testosterone from males to females. However, a recent study (Toivainen et al., 2018) found that prenatal testosterone does not explain sex differences in spatial ability.

Alternatively, sex differences have also been explained in terms of environmental influence (e.g., Bing, 1963). For example, traditional gender roles that give rise to social biases may result in child rearing practices that support spatial ability development more in males than in females (Bing, 1963) and in the types of spatial development toys and games marketed to boys versus girls (Cherney & London, 2006; Feld, Grofman, Hartly, Kilgour, & Miller, 1987; Sherman, 1967). Additionally, females may perform below their true ability on measures of spatial ability due to stereotype threat. Stereotype threat is the experience of being in a situation where one faces judgment based on negative societal stereotypes about one's group (Spencer, Steele, & Quinn, 1999). For example, stereotype threat may result in low standardized test scores of students from stigmatized social groups due to negative performance expectations about these groups (McGlone & Aronson, 2006). Stereotype threat research has suggested that poor performance on tests including measures of spatial ability may stem

from preconceived negative expectations about female performance (McGlone & Aronson, 2006).

These contradictory views and findings make unclear the origins of group differences in spatial ability and whether these group differences may relate to individual differences. Additionally, genetic and environmental influences are not mutually exclusive. Environments provided and/or sought out may differ between males and females. That is, males and females may be exposed to different environments, which could differentially effect the degree of gene-environment interplay on spatial ability.

Previous research has suggested a prominent environmental influence on female spatial ability as compared to male (Lynn & Irwing, 2004; Linn & Peterson, 1985; Harris, 1978). In the current study, we test sex as a potential moderator, to understand whether the contributions of genetic and environmental factors differ between males and females. We did not test for group differences in performance, but examined whether there were differences in the genetic and environmental contributions to individual differences in spatial ability based on sex. Sex was hypothesized to moderate the levels of environmental influence on spatial ability, with a larger amount of variation due to shared and non-shared environmental influences in females. This was hypothesized due to research described above suggesting that stereotype threats negatively influence female performance on spatial ability tasks. Recent work has also shown that children as young as preschool are susceptible to gender stereotypes and that anxiety about spatial activities is more prevalent among girls than boys in first and second grades (Casey et al., 1995). This influence may result in female spatial abilities having a larger impact from environmental factors such as educational environment, and peer negative expectations.

3. Developmental stages

Cognitive development occurs rapidly during the early years of life, with different abilities emerging during specific periods. This developmental growth reaches its peak at some point during late adolescence and then slowly deteriorates into old age (Li et al., 2004). With spatial abilities, the primary period of development occurs between 7 and 11 years of age (Hart & Moore, 1973). Fluid intelligence performance, a construct highly correlated with spatial reasoning, has been shown to peak in individuals by their mid-20s, with age-related declines already visible by the mid-30s (Li et al., 2004).

Research has shown that heritability estimates vary as a function of development, a phenomenon that we know has been observed for intelligence (e.g. Haworth et al., 2010; Tucker-Drob, Briley, & Harden, 2013). The influence of genetic contributions are known to increase with age because of individuals' increased control over their environment (Haworth et al., 2010). Additionally, correlations between biological parents' and child's spatial ability factor scores increase from ages 2 to 16, but does not increase for adopted children (Plomin, Fulker, Corley, & DeFries, 1997). However, the current behavioral genetic literature has not adequately explored potential changes in genetic and environmental contributions specific to spatial abilities during different developmental stages across the life span. In the current study, we test developmental stage as a potential moderator by grouping studies depending on the age range of participants. We hypothesized that genetic influence would be higher among adults (participants aged 26–65) as compared to children (participants aged ages 4–15) and late adolescents (participants aged 16–25) because genetic influences on cognitive traits increase with age (Haworth et al., 2010).

4. The current study

While we know spatial ability is highly heritable, it is important to understand if and how much variation may be due to environmental influences, and arrive at consensus estimates for both heritability and

shared environment influences. This is important because it can provide insight on whether spatial abilities might be amenable to training. High heritability does not imply immutability. However, if the shared environmental influence is minimal, this means that current differences between environments (e.g., households, schools) explain little variance, suggesting that it may be difficult (though not impossible) to develop interventions or training regimens that yield strong effects (Rimfeld, Kovas, Dale, & Plomin, 2016). The potential moderators examined also allow us to investigate whether spatial ability training may be more promising (i.e., if shared environment contributions are larger) for certain types of spatial ability, for one gender over another, or during certain developmental stages.

5. Methods

All hypothesis, methods and analysis plan were pre-registered and can be found at <https://osf.io/58rf2/register/5730e99a9ad5a-102c5745a8a>.

We designed the meta-analysis and report the results in accordance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Tetzlaff, Altman, & The PRISMA Group, 2009).

5.1. Inclusion criteria and literature search

To be included in the meta-analysis, the study needed to:

- be a twin study design;
- include data for same sex twin pairs;
- include a measure of spatial ability;
- measure heritability;
- report an effect size reflecting the genetic influence on spatial reasoning or enough information to compute this effect size;
- report the methods and results in English; and
- include only human participants without known disorders or disabilities.

To identify studies meeting these criteria, we systematically searched for relevant published and unpublished studies, regardless of discipline or date of publication. This was done by searching electronic databases (PsychInfo, PubMed, ProQuest) and Google Scholar, using combinations of the following search terms: twin; genetic; heritability; spatial reasoning; spatial ability. Reference sections of twin studies were scanned for relevant articles. We also e-mailed authors of relevant studies (29 authors were contacted) and requested unpublished data. 35% responded to our request, however this did not lead to additional data sets.

The literature search was conducted from April 2016 – November 2016. The initial search yielded 1883 potentially relevant articles. After removing duplicates, this number decreased to 1841 potential articles. From the records collected during the search, we used the inclusion criteria to screen titles and abstracts, which resulted in the evaluation of 148 full text articles. Thirty-seven studies met all of the inclusion criteria. (See Fig. 1.)

These studies included a total sample size of 41,623 same-sex twin pairs, 18,296 of which were monozygotic (MZ) twins, and 23,327 were dizygotic (DZ) twins. See Table 1 for a list of these studies. The mean age across all samples was 30.8, with a range of 3–98. The majority of the twin pairs from the selected studies were sampled from the Swedish Twin Registry and the Twins Early Development Study (TEDS) (see Table 2). An open data spread sheet including detailed information from each study included in the meta-analysis is available at <https://osf.io/v73sg/>.

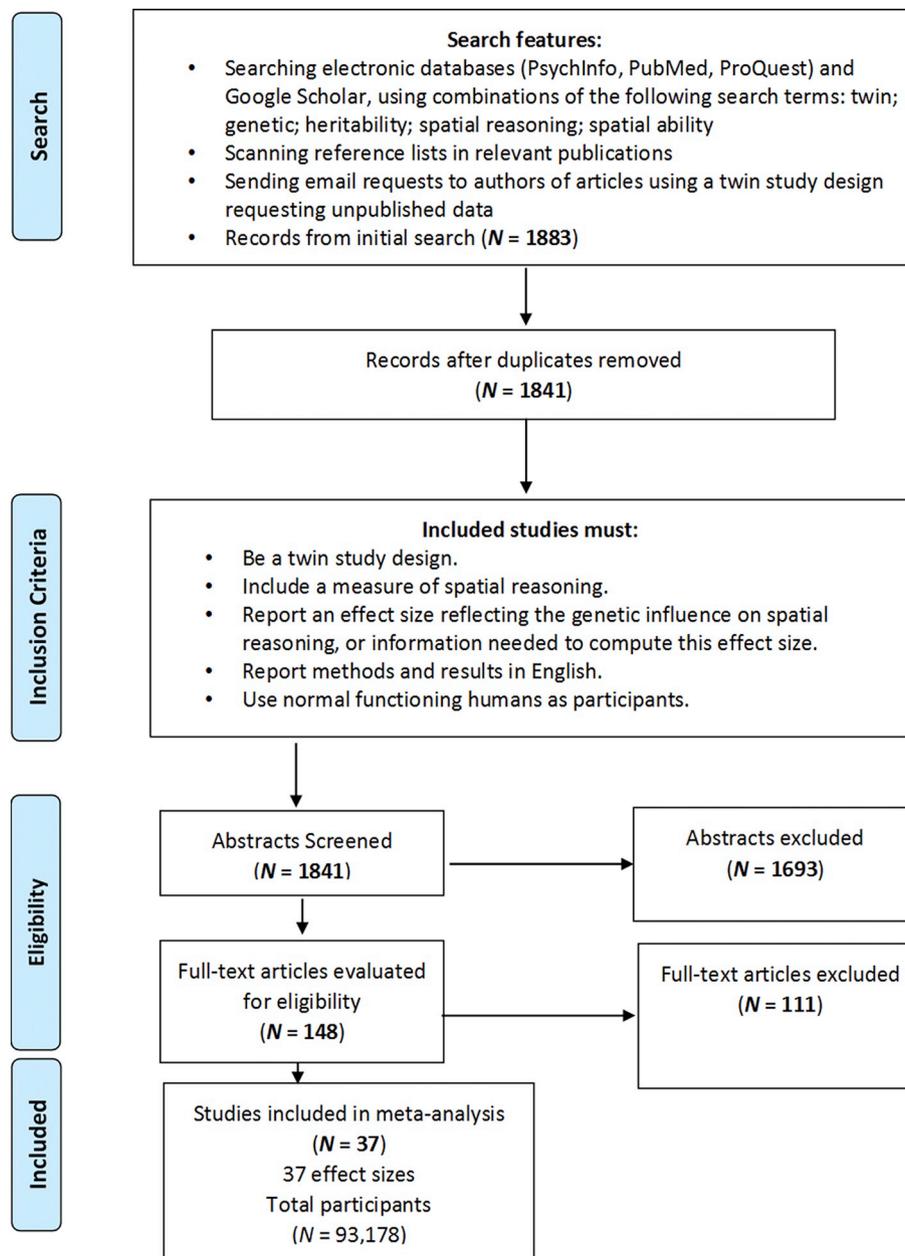


Fig. 1. Flow diagram of the literature search.

5.2. Measures of heritability and environmental contributions

5.2.1. ACE variance components

The “ACE model”¹ is a type of structural equation model specialized for examining twin data. “A” represents the estimate of additive genetic contributions (i.e., heritability) to individual differences in a trait. Monozygotic (MZ, i.e., identical) twins share 100% of their genes while dizygotic (DZ, i.e., fraternal) twins share on average 50% of their genes. Thus, higher rates of similarity between MZ twins compared to DZ twins are attributable to heritability. (Twin studies provide estimates of broad sense heritability which includes both additive and nonadditive

¹ If MZ twin resemblance is roughly twice the size of the DZ resemblance, then genetic influences are assumed to be primarily additive. However, if the ratio of the MZ resemblance is > 2:1 when compared to the DZ resemblance, nonadditive genetic influences may be present and should be tested for using an alternate “ADE” model. We had no such cases and therefore no ADE models were included in the meta-analysis.

(dominance) genetic influences. “C” represents an estimate of the effect of common (shared) environmental influences. This includes variation due to factors and events that are shared between twin pairs who were reared in the same environment. If MZ twin pair resemblance is greater than the heritability of the trait, then shared environmental influences are evident (Zyphur, et al., 2013). “E” represents an estimate of non-shared environmental influences. This includes variation due to unique factors and events that occur in one twin but not the other. The extent to which MZ twin resemblance is < 1 indicates nonshared-environmental effects. “E” also captures measurement error. Using this method, researchers can test both MZ and DZ twin pairs on a given measure, and then calculate estimates of the proportion of variance due to genetic and environmental influences. For each study included in the meta-analysis, we recorded the reported estimates from the ACE model analysis, if such an analysis was produced from the study.

5.2.2. Twin correlations

The ACE model analysis in essence is based on comparing the trait

Table 1
Studies Included in the Meta-analysis.

| Citation |
|---|
| Ando, Ono, and Wright (2001) |
| Bartels et al. (2012) |
| Calvin et al. (2012) |
| Chow et al. (2013) |
| Derks, Dolan, and Boomsma (2006) |
| Engelhardt et al. (2016) |
| Finkel and Pedersen (2000) |
| Giubilei et al. (2008) |
| Hutson-Khalid (2008) |
| Johnson and Bouchard Jr (2007) |
| Johnson et al. (2007) |
| Madison, Mosing, Verweij, Pedersen, and Ullén (2016) |
| Malanchini et al. (2016) |
| McClearn et al. (1997) |
| Mosing et al. (2014) |
| Mosing et al. (2012) |
| Neubauer, Spinath, Riemann, Borkenau, and Angleitner (2000) |
| Panizzon et al. (2014) |
| Petrill et al. (1998) |
| Read et al. (2006) |
| Reynolds et al. (2002) |
| Rijsdijk, Vernon, and Boomsma (1998) |
| Shakeshaft et al. (2016) |
| Schermer, Johnson, Jang, and Vernon (2015) |
| Shikishima et al. (2009) |
| Singer, MacGregor, Cherkas, and Spector (2006) |
| Svedberg, Gatz, and Pedersen (2009) |
| Swagerman et al. (2016) |
| Tosto et al. (2014) |
| Trzaskowski, Shakeshaft, and Plomin (2013) |
| Tucker-Drob et al. (2014) |
| Tucker-Drob et al. (2016) |
| Vasilopoulos et al. (2012) |
| Vasilopoulos et al. (2010) |
| Vuoksimaa et al. (2010) |
| Wallace et al. (2010) |
| Woodley and Madison (2015) |

resemblance of MZ and DZ twin pairs. If a study did not use the ACE modeling approach, correlations between scores on spatial ability tests for each member of a twin pair (when reported separately for both MZ and DZ twin pairs) can also be used to estimate heritability, shared, and nonshared environmental influences. For each study included in the

Table 2
Twin registry descriptive statistics.

| Twin Registry | Country | K | n pairs (MZ) | n pairs (DZ) | Age Range |
|--|------------------|---|--------------|--------------|-------------|
| Swedish Twin Registry | Sweden | 8 | 4393 | 4879 | 27–100 |
| Twins Early Development Study (TEDS) | England & Wales | 5 | 6530 | 9890 | 4.5–13 |
| Netherlands Twin Register (NTR) | Netherlands | 3 | 2902 | 3098 | 10–86 |
| Vietnam Era Twin Study of Aging (VETSA) | United States | 2 | 1735 | 1335 | 51–60 |
| Keio Twin Project | Japan | 2 | 750 | 340 | 16–36 |
| Minnesota Study of Twins | United States | 2 | 300 | 178 | 16–79 |
| Texas Twin Project | United States | 2 | 423 | 441 | 7.8–15.25 |
| Brisbane Memory, Attention and Problem Solving (MAPS) twin study | Australia | 2 | 450 | 550 | 15.42–18.16 |
| German Observational Study of Adult Twins (GOSAT) | Germany | 1 | 507 | 393 | 18–70 |
| FinnTwin12 study | Finland | 1 | 143 | 121 | 21–24 |
| Dutch twin pairs | Netherlands | 1 | 160 | 216 | 16–18 |
| Via advertisements in London, Ontario, Vancouver, and British Columbia | England & Canada | 1 | 77 | 102 | 6 |
| “Typically developing Chinese twins” | China | 1 | 228 | 84 | 3–11 |
| UK Government’s Department for Children, Schools and Families | United Kingdom | 1 | 765 | 1347 | 11 |
| Brisbane Longitudinal Twin Studies | Australia | 1 | 57 | 74 | 12 |
| Child Psychiatry Branch of the National Institute of Mental Health | United States | 1 | 224 | 119 | 5.3–19.5 |
| Western Reserve Twin Project | United States | 1 | 87 | 75 | 6–15 |
| The Italian twin registry | Italy | 1 | 35 | 58 | 62–80 |
| St Thomas’ Adult UK Twin Registry | United Kingdom | 1 | 108 | 170 | 18–76 |

Note. *k* = number of effect sizes contributing to full sample; *n* pairs = number of twin pairs contributing to the meta-analysis; Age Range = age range in years of participants from studies contributing to the meta-analysis by registry; single numbers in this column reflect that all participants in the study or studies included in the meta-analysis from the registry in that row were that age.

meta-analysis, we recorded the reported twin correlations, and the number of MZ and DZ pairs that contributed to this data. This produces all of the information necessary to calculate heritability, and environmental influence using Falconer’s Equation so that they could be included in the meta-analysis. Using this equation, heritability can be expressed as twice the difference of MZ and DZ correlations [$2 \times (r_{MZ} - r_{DZ})$] (Falconer, Mackay, & Frankham, 1996). The values are then transformed back to correlations for interpretation. There were no cases where the MZ:DZ *r* ratio was $> 2:1$, therefore no further adjustments were necessary.

5.2.3. Moderator variables

When a study collected multiple measures relevant to our moderator sub-groups, the effect sizes associated with these measures were entered in the moderator analysis (even if they had been combined and adjusted for dependent samples in the main analysis available). For example, if a study administered two types of spatial ability measures, only one effect size is entered into the main analysis (adjusted for dependent effects). However, for the spatial ability type moderator analysis, these two effects are separated and entered into the moderator analysis, one in each sub-group, in order to examine potential differences in spatial ability measure. When individual effect sizes could not be extracted based on moderator level, they were not entered into the analysis. For example, if a study only reported an effect size for all participants, not separate effect sizes for their male and female participants, we could not include it in the sex differences moderator analysis. Thus, the number of effect sizes for each moderator analysis could be greater or smaller than the number of effect sizes entered into the models examining the overall effects.

5.2.4. Spatial ability type

When available, effect sizes for each spatial ability type were entered into the analysis separately based on the three distinct categories described below. That is, when studies reported multiple measures of spatial ability using the same sample, each estimate was entered as independent in order to test this moderator. Effect sizes based on composite measures that spanned multiple categories were not included in this moderator analysis.

5.2.5. Mental rotation

Mental rotation refers to the ability to mentally rotate

representations of two-dimensional and three-dimensional objects (Shepard & Metzler, 1971). Tests of mental rotation often involve a comparison of two or more 3-D or 2-D figures that have been rotated in some degree, and a judgment of whether or not the images are the same figures. Types of mental rotation tasks include the Mental Rotation Test (MRT), card rotation, and cube rotation.

5.2.6. Spatial visualization

Spatial visualization is the ability to mentally represent visual appearances of an object and the spatial relations between the locations of the objects or movements (Linn & Petersen, 1985). Tests of spatial visualization tap into ones' ability to mentally manipulate 3-D and 2-D figures, in ways other than mental rotation. For example, the paper folding task presents a 2-D folded piece of paper with distinct holes punched through it. Participants must choose from a set of unfolded pieces of paper, which one corresponds to the completed folded one. Other types of spatial visualization tasks include the Object Learning Test, line orientation, box folding, jigsaw, hidden shapes, form board, surface development, figure logic, and block design.

5.2.7. Matrix reasoning

Matrix reasoning is an abstract spatial reasoning ability that is often used as a non-verbal estimate of fluid intelligence. Matrix reasoning typically involves a set of figures that are associated with a specific pattern, with one figure in the series left blank. The participant is tasked with selecting the figure from multiple options that fits the missing piece. The tasks used to measure matrix reasoning ability are typically one of the variations of Raven's Progressive Matrices Test. This includes the Standard Progressive Matrices test (SPM), the Advanced Progressive Matrices test (APM), and the Colored Progressive Matrices. Other types of matrix reasoning tests include the Wechsler Abbreviated Scale of Intelligence (WASI) Matrix Reasoning sub-test, and the Wiener Matrizen Test.

5.2.8. Sex

When ACE value estimates were provided separately for males and females, they were entered separately in the meta-analysis. Likewise, when available, twin correlations for males and females were entered separately resulting in four groups: monozygotic male (MZM), monozygotic female (MZF), dizygotic male (DZM), and dizygotic female (DZF) pairs. In these cases, or when the sample only included one sex, we included these effect sizes in this moderator analysis.

5.2.9. Age

Age was classified into four groups on the basis of the age range of the included samples. Due to the limited number of effect sizes with a mean sample for a typical childhood cohort, we set the range from ages 4–15, extending into early adolescence. This range also includes the primary period for spatial development (childhood ages 7–11) (Hart & Moore, 1973). Maximum fluid intelligence performance, a construct highly correlated with spatial reasoning, has been shown to occur by individuals in their mid-20s (Li et al., 2004). We capture the late adolescence age range (ages 16–25) before this peak while cognitive development is still progressing. The adulthood cohort was set at ages 26–65, during a period where cognitive ability progressively declines. The older adulthood cohort was set at age 65 and older (Erikson, 1964). Only studies whose entire sample fell within one of our age ranges were included in this moderator analysis.

5.3. Meta-analytic procedure

The meta-analyses involved five steps. The first step was to obtain effect sizes. These were taken directly from the reported ACE values. Effect sizes were also calculated from reported twin correlations using Falconer's equations (Falconer et al., 1996). Using these equations, estimates of heritability, shared environment, and non-shared

environment can be generated based solely on the MZ and DZ twin correlations. For example, heritability can be expressed as twice the difference of MZ and DZ correlations [$2 \times (r_{MZ} - r_{DZ})$]. The values were transformed back to correlations for interpretation.

The second step was to conduct the DerSimonian-Laird (DSL) random-effects meta-analysis as described by Schulze (2004) using the R package metafor. The DSL method transforms correlations into Fisher z -values before the meta-analysis is conducted in order to reduce the risk of larger correlations being assigned too much weight due to small standard errors. The meta-analytic method ensures that the weight assigned to each effect size is based solely on sample size instead of the standard error of the reported correlation, which is a better practice to follow when sample sizes are very large as is often the case with twin studies (Schulze, 2004). When studies used multiple measures from a single sample, we adjusted the weight for these effects using Cheung and Chan (2008) method.

Two separate DSL random-effects meta-analysis were run and compared. The first meta-analysis included studies that reported ACE model-fitting values, whereas the second meta-analysis included studies that reported twin correlations and ACE values. A similar side-by-side comparison was made by Polderman et al. (2015) who found that reported estimates of variance components from model-fitting resulted in lower values of heritability, compared with twin correlation heritability estimates.

The third step was to conduct mixed effect moderator analyses using the R package metafor. Three moderator analyses were conducted, one for spatial ability type, one for sex, and one for age each for heritability, shared environment, and non-shared environment.

The fourth step was to conduct a sensitivity analysis of the large twin registries. Due to the nature of twin research, there are several studies in this meta-analysis that use data from the same twin registries. It is unknown how much of the samples across studies might overlap when recruiting from the same pool. To the best of our knowledge, past meta-analyses of twin studies have always treated the samples produced from each study as independent (Polderman et al., 2015). In most cases it is near impossible to know if the same or an overlapping sample of twins may have been used across multiple studies. In order to address this potential issue of dependent samples within registries, two analyses were conducted in which all of the effect sizes from one registry were removed and the meta-analysis was re-conducted to see if the average estimates changed. We did this for each of the two registries that provided a large number of effect sizes. The first "leave-one-out" analysis removed all effect sizes from the Swedish Twin Registry and the second analysis removed all the effect sizes that used data from the TEDS registry.

The final step was to conduct publication bias analyses. We examined funnel plots and conducted Duval's trim and fill analyses (Duval & Tweedie, 2000). Duval's trim and fill analysis uses funnel plots to estimate the number of studies missing from the meta-analysis and estimates the overall meta-analytic average effects for if these studies were included.

6. Results

The results are summarized in Figs. 2–5 and Table 3. The effect sizes are plotted in ascending order in separate forest plots for additive genetic contributions (Fig. 2), shared environment contributions (Fig. 3), and non-shared environment contributions (Fig. 4). These separate components were analyzed independent of one another, and across multiple studies. For this reason, the meta-analytic average estimates for each separate contribution do not sum to exactly 1.0, as would be the case in a single study (Polderman et al., 2015; Tucker-Drob & Briley, 2014). The results from the DSL random effects meta-analysis on the aggregated data set that included 42 effect sizes yielded an average estimate of $a^2 = .61$, 95% CI [.55, .66] for additive genetics, $c^2 = .07$, 95% CI [.05, .10] for shared environment, and $e^2 = .43$, 95% CI [.38,

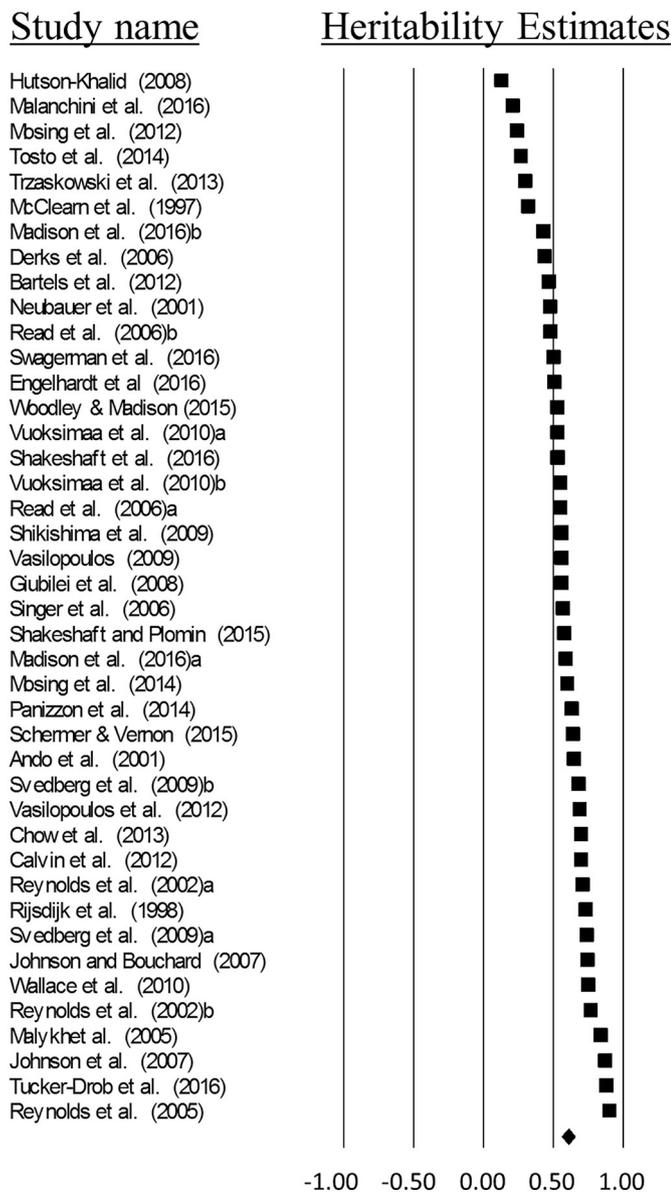


Fig. 2. Forest plot of the heritability estimates (black rectangles). All confidence intervals are smaller than the width of the rectangles representing the estimates. This represents the model including all reported ACE values. Documents with more than one sample have a's and b's at the end of the citation.

.49] for non-shared environment. The I^2 statistic, which describes the percentage of variation across studies that is due to heterogeneity rather than chance (Higgins & Thompson, 2002; Higgins et al., 2003), was high for additive genetics (98.23), shared environment (79.15), and non-shared environment (96.14).

An analysis was also conducted using only the estimates derived from twin correlations, rather than the ACE values (see Table 4). For this analysis the meta-analytic average for heritability was $a^2 = .61$, 95% CI [.54, .67]. The averages for estimates of environmental influence were $c^2 = .06$, 95% CI [.02, .11] for shared environment, and $e^2 = .48$, 95% CI [.41, .55] for non-shared environment.

These results indicate that the effect sizes based on variance components, i.e. ACE values, produced a near identical estimate of heritability as the effect sizes based on twin correlations. This is not in line with previous findings in a meta-analysis on the heritability of multiple human traits which found that reported estimates of variance components from model-fitting results in lower values of heritability, when

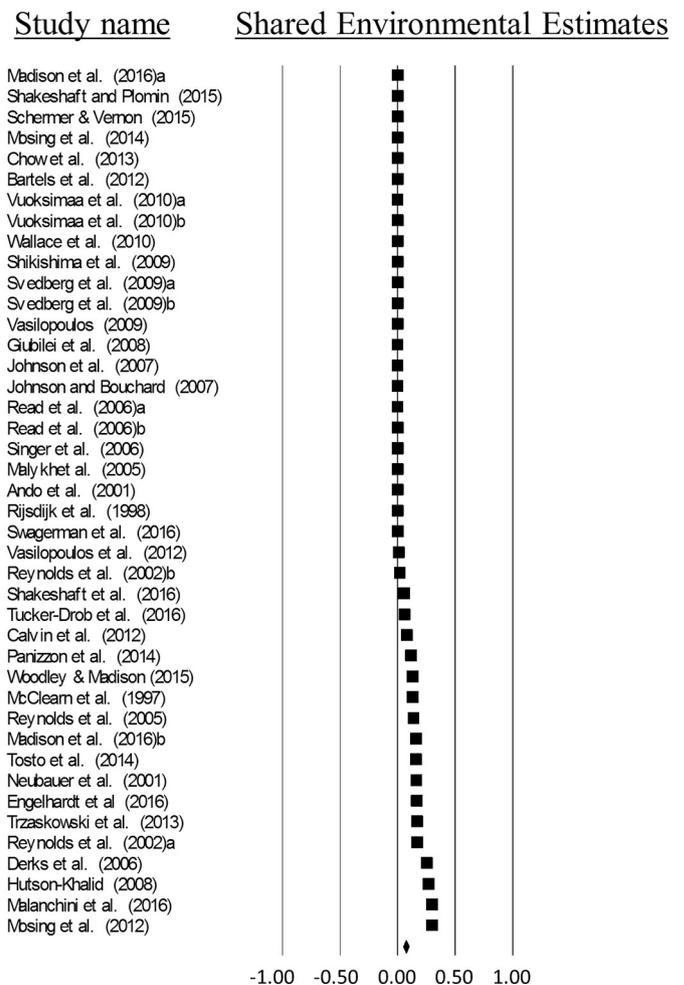


Fig. 3. Forest plot of the shared environment estimates (black rectangles). All confidence intervals are smaller than the width of the rectangles representing the estimates. This represents the model including all reported ACE values.

compared with heritability based on twin correlations (Polderman et al., 2015).

6.1. Moderator analyses

6.1.1. Spatial ability type

The moderator analysis for spatial ability type included 50 effect sizes. Of the 41 effect sizes, 20 were spatial visualization measures, 14 were matrix-reasoning tests, and the remaining 7 were mental rotation measures. The model was not significant for differences in heritability ($Q = 1.00, p = .687$), shared environment ($Q = 0.45, p = .800$), or non-shared environment ($Q = 0.33, p = .847$). See Tables 11–13 for ACE values for each spatial ability type.

6.1.2. Sex

Only a subset of studies reported separate effect sizes for male and female participants. Consequently, this moderator analysis only included 11 effect sizes, 6 consisting of all males and 5 consisting of all females. The moderator analysis of sex was not significant for heritability ($Q = 0.29, p = .588$), shared environment ($Q \leq .01, p = .946$), or non-shared environment ($Q = 0.390, p = .530$). Estimates of additive genetics and non-shared environment were significant for both females and males, $p < .001$. The effect of shared environment was significant for males but not for females; although, the effect sizes were similar. See Tables 5 and 6 for ACE values.

Study name **Non-shared Environmental Estimates**

- Tucker-Drob et al. (2016)
- Reynolds et al. (2002)a
- Malykh et al. (2005)
- Reynolds et al. (2005)
- Reynolds et al. (2002)b
- Calvin et al. (2012)
- Wallace et al. (2010)
- Svedberg et al. (2009)a
- Johnson et al. (2007)
- Chow et al. (2013)
- Vasilopoulos et al. (2012)
- Svedberg et al. (2009)b
- Woodley & Madison (2015)
- Schermer & Vernon (2015)
- Mbsing et al. (2014)
- Johnson and Bouchard (2007)
- Madison et al. (2016)a
- Madison et al. (2016)b
- Shakeshaft and Plomin (2015)
- Rijsdijk et al. (1998)
- Singer et al. (2006)
- Vasilopoulos (2009)
- Giubilei et al. (2008)
- Vuoksima et al. (2010)b
- Read et al. (2006)a
- McClearn et al. (1997)
- Derks et al. (2006)
- Shikishima et al. (2009)
- Mbsing et al. (2012)
- Vuoksima et al. (2010)a
- Panizzon et al. (2014)
- Malanchini et al. (2016)
- Read et al. (2006)b
- Trzaskowski et al. (2013)
- Ando et al. (2001)
- Tosto et al. (2014)
- Neubauer et al. (2001)
- Hutson-Khalid (2008)
- Engelhardt et al. (2016)
- Bartels et al. (2012)
- Shakeshaft et al. (2016)
- Swageman et al. (2016)

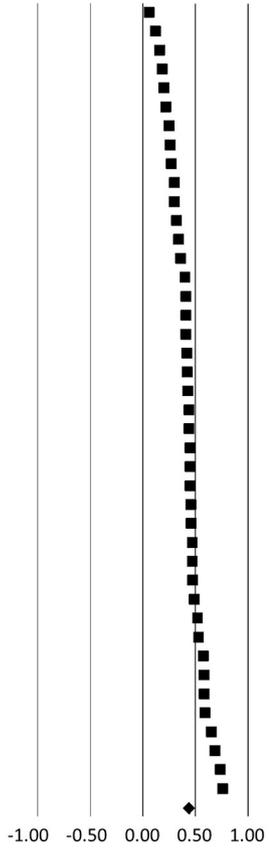


Fig. 4. Forest plot of the non-shared environment estimates (black rectangles). All confidence intervals are smaller than the width of the rectangles representing the estimates. This represents the model including all reported ACE values.

6.1.3. Developmental stage

This analysis included 27 effect sizes. Of the 27 total effect sizes, 10 fell into the children age group, 5 in the late adolescence group, 6 in the

Table 3

Summary estimates based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | I ² |
|--------------|----------|----------|----------------|
| A | .61 | .55–.66 | 98.23% |
| C | .07 | .05–.10 | 79.15% |
| E | .43 | .38–.49 | 96.19% |

Note. 42 effect sizes included in analysis.

Table 4

Summary estimates of ACE values based on twin correlations with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Derived ACE | Estimate (C.I.) | 95% C.I. | I ² |
|-------------|-----------------|----------|----------------|
| A | .61 | .54–.67 | 98.74% |
| C | .06 | .02–.11 | 93.0% |
| E | .48 | .41–.55 | 96.19% |

Note. 25 effect sizes included in analysis.

Table 5

Summary estimates for Spatial Visualization based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .66 | .55–.77 | < .001 |
| C | .08 | .04–.13 | < .001 |
| E | .46 | .37–.56 | < .001 |

Table 6

Summary estimates for Matrix Reasoning based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .58 | .45–.71 | < .001 |
| C | .08 | .03–.13 | .003 |
| E | .51 | .39–.62 | < .001 |

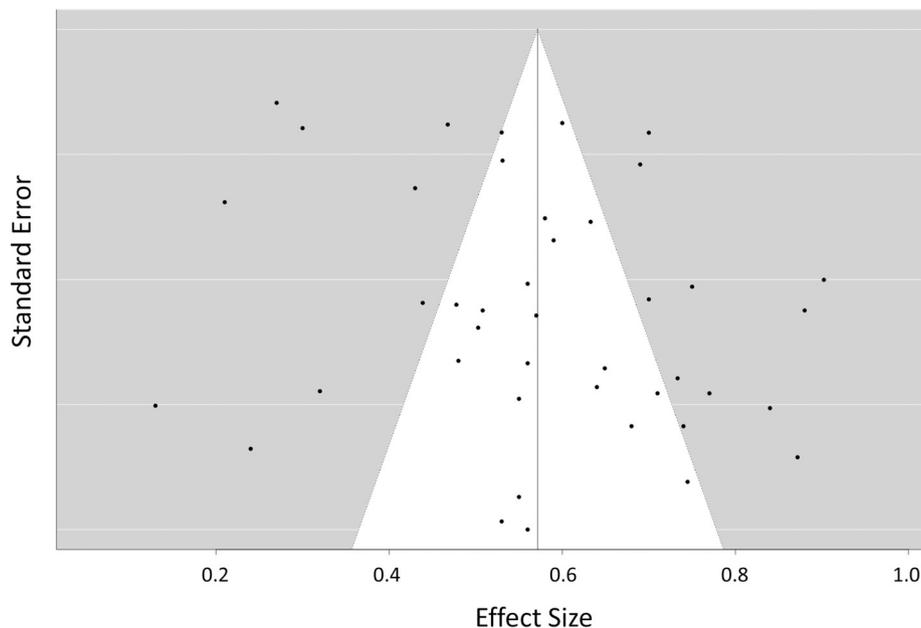


Fig. 5. Funnel plot of additive genetic estimates for the aggregated data.

Table 7

Summary estimates for Mental Rotation based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .65 | .46–.84 | < .001 |
| C | .05 | –.03–.13 | .203 |
| E | .50 | .34–.66 | < .001 |

Table 8

Summary estimates for females based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .57 | .40–.75 | < .001 |
| C | .07 | –.02–.16 | .147 |
| E | .48 | .38–.58 | < .001 |

Table 9

Summary estimates for males based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .50 | .31–.69 | < .001 |
| C | .07 | .03–.18 | .153 |
| E | .52 | .41–.63 | < .001 |

Table 10

Summary estimates for Childhood based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .56 | .35–.77 | < .001 |
| C | .15 | .10–.20 | < .001 |
| E | .47 | .33–.61 | < .001 |

Table 11

Summary estimates for Late Adolescence based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .58 | .28–.88 | .001 |
| C | .00 | –.08–.08 | 1.00 |
| E | .66 | .46–.85 | < .001 |

Table 12

Summary estimates for Adulthood based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .69 | .42–.96 | < .001 |
| C | .07 | .01–.13 | .074 |
| E | .35 | .17–.53 | < .001 |

adulthood age group, and 6 in the older age group. We were unable to treat age as a continuous variable due to constraints from the information provided. Studies provided information on the sample's mean or median age, and the age range. Studies included in each age category

Table 13

Summary estimates for Older Adulthood based on ACE values with confidence intervals for additive genetics (A), shared environment (C), and non-shared environment (E).

| Reported ACE | Estimate | 95% C.I. | p-value |
|--------------|----------|----------|---------|
| A | .81 | .54–1.08 | < .001 |
| C | .05 | –.03–.13 | .201 |
| E | .37 | .19–.55 | < .001 |

described above were done so only if their age range fit into the specified category. Studies omitted from the age moderator analysis included participants whose ages spanned more than one of the developmental stages selected for this study. The only moderator analysis that yielded a significant result was the effect of developmental stage on the impact of shared environment $Q(3) = 11.86, p = .008$. This result is driven by the youngest age group (4–15 years old) where the impact of the shared environment was significant and differed from the other age groups. The impact of age on heritability ($Q = 2.33, p = .507$) and non-shared environment were non-significant ($Q = 6.28, p = .099$). See Tables 7–10 for ACE values for each age group.

6.2. Additional analyses

6.2.1. Registry sensitivity analysis

Due to the nature of twin research, there are several studies in this meta-analysis that use data from the same twin registries. It is unknown how much of the samples across studies might overlap when recruiting from the same pool. Two registries—the Swedish Twin Registry and TEDS—contributed substantially more effect sizes than the other registries. Given that twins may have been resampled across studies from these registries, we conducted sensitivity analyses to test whether either of these registries unduly influenced our results. When removing all of the effect sizes from the Swedish Twin Registry, the meta-analytic estimate of heritability was $a^2 = .59$ (compare with $a^2 = .61$), shared environment was $c^2 = .08$ (compare with $c^2 = .07$), and non-shared environment was $e^2 = .47$ (compare with $e^2 = .43$). None of these results were significantly different from the main model (all $ps > .05$). When removing all of the effect sizes from TEDS, the meta-analytic estimate of heritability was $a^2 = .63$ (compare with $a^2 = .61$), shared environment was $c^2 = .06$ (compare with $c^2 = .07$), and non-shared environment was $e^2 = .41$ (compare with $e^2 = .43$). Again, none of these results were significantly different from the main model (all $ps > .05$).

6.2.2. Publication bias analysis

Funnel plots (Figs. 5, 6, and 7), which depict the relationship between sampling error and effect size, appear to be relatively symmetrical. Symmetrical funnel plots suggest that small studies with lower than average effect sizes are not missing from our meta-analysis.

7. Discussion

Spatial reasoning ability has been extensively researched and recognized as a cognitive ability similar to fluid reasoning and a component of intelligence (Cattell, 1963). The benefits and impact of spatial ability have been connected to the functioning of other high-level cognitive skills (e.g., logical reasoning, memory retrieval, verbal skills), development of expertise in STEM related skills, as well as academic and occupational achievement in STEM fields later in life. The literature examining the etiology of individual differences in spatial ability has reported highly variable results in terms of the magnitude of genetic and environmental influences with estimates ranging from .21 (Malanchini et al., 2016) to .94 (Tucker-Drob et al., 2016) for heritability effects, and from .16 (Malykh et al., 2005) to .79 (Malanchini

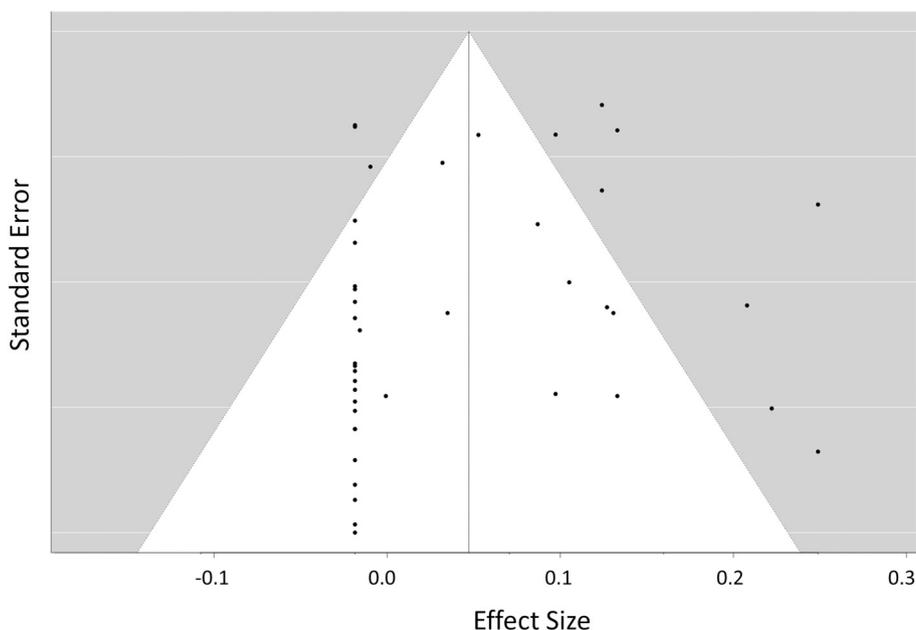


Fig. 6. Funnel plot of shared environment estimates for the aggregated data.

et al., 2016) for the combined environmental effect. The reason for this high variability is unclear.

We conducted a meta-analysis on the spatial reasoning twin study literature to provide a consensus on the magnitude of genetic and environmental influences on spatial reasoning ability. We found that spatial reasoning ability is largely heritable, with non-shared environmental factors having a moderate impact, and shared environment having very little impact. We also tested whether genetic and environmental influences were moderated by differences in spatial ability type, sex, or age. Mental rotation, spatial visualization, and matrix reasoning are often considered to be separate and unique skills that collectively contribute to spatial reasoning ability (Thurstone, 1950; Lohman, 1979; Linn and Petersen, 1985; Hegarty & Waller, 2004, 2005). Moderator analyses revealed that there are no significant differences in the relative contributions of genetic, shared environmental, or non-shared environmental influences for tests of spatial

visualization, mental rotation, or matrix reasoning. This suggests that individual variation in these 3 types of spatial abilities are similarly determined by the same proportions of genetic and environmental influences.

Moderator analyses did not reveal any significant differences in genetic, shared environmental, or non-shared environmental influences on spatial ability between males and females. However, the number of effect sizes that contributed to this moderator analysis were severely limited. As the literature in this area expands, future meta-analyses may find differences in genetic and/or environmental influences by sex that we were unable to detect.

Research on spatial ability across age has attempted to provide an explanation for individual differences possibly stemming from early developmental factors. Moderator analyses revealed that the influence of shared environment differed significantly across the age groups. Shared environmental influence was significantly different from zero

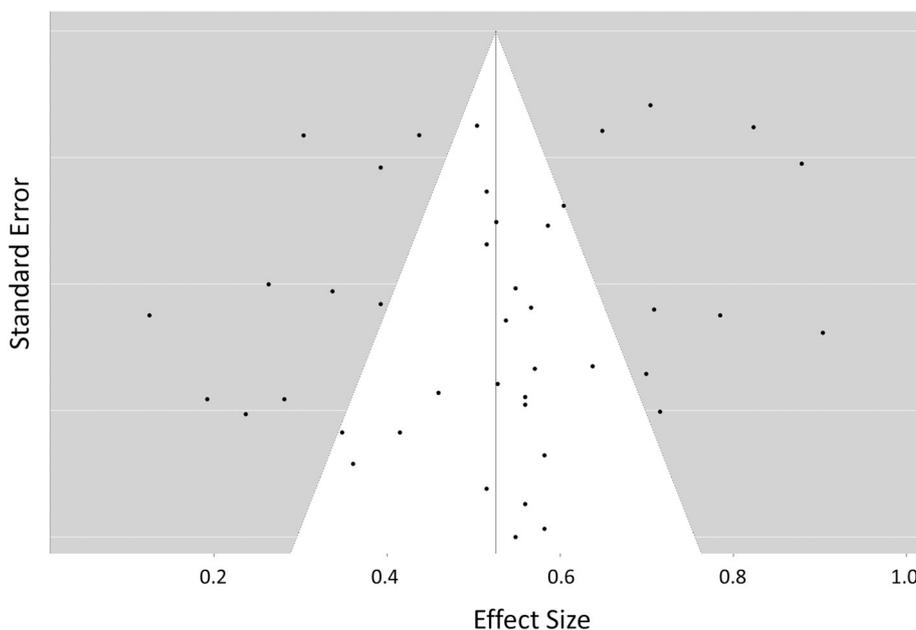


Fig. 7. Funnel plot of non-shared environment estimates for the aggregated data.

for the childhood group (ages 4–15) but not significant for the late adolescence (ages 16–25), adulthood (age 26–65), or older adulthood (age 65+) groups. We found no significant differences in the magnitude of genetic influence or in the magnitude of non-shared environmental influence across age groups. According to previous research, the impact of additive genetics tends to grow over the course of the lifespan (suggesting environmental influence is more important earlier on in life; Tucker-Drob & Briley, 2014). Thus, our finding that shared environment is only significant at the youngest age group is consistent with this finding. However, again, relatively few effect sizes contributed to these sub-groups and the majority came from two registries so these results should be interpreted with caution.

7.1. Limitations

One limitation of the current study is the use of meta-analysis. Meta-analyses are limited by the amount and quality of the studies included. The current study made efforts to include every study possible on this topic, including studies that were not published. However, it is difficult to be sure that the studies included in the current meta-analysis is completely exhaustive.

Twin studies assume that MZ and DZ twin pairs share to the same extent those aspects of the environment relevant for spatial ability (see e.g., Plomin et al., 1997; Rijdsdijk et al., 1998). While past attempts to examine the validity of the equal environments assumption have found that it is a valid assumption, to our knowledge, no one has specifically tested the assumption for spatial ability. Twin studies also assume that findings can be generalized to non-twin populations.

A limitation of our moderator analyses is that, despite large sample sizes contributing to each effect size, there were not a large number of effect sizes for certain sub-groups because not all effect sizes available for the main analyses were available for each of our moderator analyses. Thus, a limited number of effect sizes were available to investigate sex, and we could only coarsely examine age. Thus, our lack of findings with sex, and our significant finding with age should both be interpreted with caution.

Furthermore, we did not have enough statistical power to examine interactions. For example, we could not assess whether environmental factors played more of a role for females in their formative years but less so for adult females. As the number of twin studies examining spatial abilities increases in the literature, future researchers should re-examine these moderators and meta-analytically examine interactions among sex, age, and spatial ability type. Examining how moderating factors might vary across levels of another moderating factor could provide a clearer understanding of the genetic and environmental contributions of spatial abilities.

Additionally, there are a number of other possible moderating variables outside the scope of the present meta-analysis that may explain why twin studies have been so variable in estimating the effect of genes and shared environment. Other potential moderators we were not able to address include socioeconomic status (SES), country of origin, and fluctuations of twin registries over time. Further research is needed to determine the influence of other factors on the origins of spatial ability.

8. Conclusion

Spatial reasoning ability is an important cognitive skill that has far reaching impacts such as success in STEM-related academic and occupational endeavors (Wai et al., 2009). Individual differences in spatial ability have been researched and discussed, with mixed findings regarding the sources of these differences. We conducted a meta-analysis on twin studies of spatial reasoning ability to provide a consensus on this issue. Our findings supported the hypothesis that spatial reasoning is largely heritable, with 52% of its variation due to genetic factors. We also investigated whether genetic or environmental contributions

differed by sex (male, female), age (childhood, late adolescence, adulthood, and older adulthood), or spatial ability type (mental rotation, matrix reasoning, spatial visualization). The only significant moderator variable was the effect of age on shared environmental influences. This finding emphasizes the importance of one's formative years where environmental factors, while not as important as genetic factors, appear to impact individual differences in spatial ability.

Acknowledgments

The authors extend their appreciation to Navi Baskar and Izma Khaliq for their contributions to the literature search.

References[†]

- *Ando, J., Ono, Y., & Wright, M. J. (2001). Genetic structure of spatial and verbal working memory. *Behavior Genetics*, 31, 615–624. <https://doi.org/10.1023/A:1013353613591>.
- *Bartels, M., van Weegen, F. I., van Beijsterveldt, C. E., Carlier, M., Polderman, T. J., Hoekstra, R. A., & Boomsma, D. I. (2012). The five factor model of personality and intelligence: A twin study on the relationship between the two constructs. *Personality and Individual Differences*, 53, 368–373. <https://doi.org/10.1016/j.paid.2012.02.007>.
- Battista, M. T. (1990). Spatial visualization and gender differences in high school geometry. *Journal for Research in Mathematics Education*, 21, 47–60. <https://doi.org/10.2307/749456>.
- Bektasli, B. (2006). *The relationships between spatial ability, logical thinking, mathematics performance and kinematics graph interpretation skills of 12th grade physics students*. Doctoral dissertation The Ohio State University.
- Bing, E. (1963). Effect of childrearing practices on development of differential cognitive abilities. *Child Development*, 34, 631–648. <https://doi.org/10.2307/1126757>.
- Bock, R. D., & Vandenberg, S. G. (1968). Components of heritable variation in mental test scores. In S. G. Vandenberg (Ed.), *Progress in human behavior genetics*. Baltimore, Md: John Hopkins University Press.
- Calvin, C. M., Deary, I. J., Webbink, D., Smith, P., Fernandes, C., Lee, S. H., & Visscher, P. M. (2012). Multivariate genetic analyses of cognition and academic achievement from two population samples of 174,000 and 166,000 school children. *Behavior Genetics*, 42, 699–710. <https://doi.org/10.1007/s10519-012-9549-7>.
- Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education*, 32, 28–57. <https://doi.org/10.2307/749620>.
- Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, 31, 697–705. <https://doi.org/10.1037/0012-1649.31.4.697>.
- Cherney, I. D., & London, K. (2006). Gender-linked differences in the toys, television shows, computer games, and outdoor activities of 5- to 13-year-old children. *Sex Roles*, 54, 717–726. <https://doi.org/10.1007/s11199-006-9037-8>.
- Cheung, S. F., & Chan, D. K. S. (2008). Dependent correlations in meta-analysis: The case of heterogeneous dependence. *Educational and Psychological Measurement*, 68, 760–777. <https://doi.org/10.1177/0013164408315263>.
- *Chow, C., Epp, J. R., Lieblich, S. E., Barha, C. K., & Galea, L. A. (2013). Sex differences in neurogenesis and activation of new neurons in response to spatial learning and memory. *Psychoneuroendocrinology*, 38, 1236–1250.
- Court, J. H. (1983). Sex differences in performance on Raven's Progressive Matrices: a review. *Alberta Journal of Educational Research*, 29(1), 54–74.
- Derks, E. M., Dolan, C. V., & Boomsma, D. I. (2006). A test of the equal environment assumption (EEA) in multivariate twin studies. *Twin Research and Human Genetics*, 9, 403–411. <https://doi.org/10.1375/twin.9.3.403>.
- Duval, S., & Tweedie, R. (2000). Trim and fill: a simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, 56, 455–463. <https://doi.org/10.1111/j.0006-341X.2000.00455.x>.
- *Engelhardt, L. E., Mann, F. D., Briley, D. A., Church, J. A., Harden, K. P., & Tucker-Drob, E. M. (2016). Strong genetic overlap between executive functions and intelligence. *Journal of Experimental Psychology: General*, 145(9), 1141. <https://doi.org/10.1037/xge0000195>.
- Erikson, E. H. (1964). *Insight and responsibility: Lectures on the ethical implications of psychoanalytic insight*. New York, NY, US: W W Norton & Co.
- Eysenck, H. J. (1981). *Intelligence: the battle for the mind*. London: Pan.
- Falconer, D. S., Mackay, T. F., & Frankham, R. (1996). Introduction to quantitative genetics (4th edn). *Trends in Genetics*, 12(7), 280.
- Feld, S. L., Grofman, B., Hartly, R., Kilgour, M., & Miller, N. (1987). The uncovered set in spatial voting games. *Theory and Decision*, 23(2), 129–155. <https://doi.org/10.1007/BF00126302>.
- *Finkel, D., & Pedersen, N. L. (2000). Contribution of age, genes, and environment to the relationship between perceptual speed and cognitive ability. *Psychology and Aging*,

[†] References marked with an asterisk indicate studies included in the meta-analysis.

- 15, 56–64. <https://doi.org/10.1037/0882-7974.15.1.56>.
- Geary, D. C., Hamson, C. O., & Hoard, M. K. (2000). Numerical and arithmetical cognition: A longitudinal study of process and concept deficits in children with learning disability. *Journal of Experimental Child Psychology*, 77(3), 236–263. <https://doi.org/10.1006/jecp.2000.2561>.
- Gilmartin, P. P., & Patton, J. C. (1984). Comparing the sexes on spatial abilities: Map-use skills. *Annals of the Association of American Geographers*, 74, 605–619. <https://doi.org/10.1111/j.1467-8306.1984.tb01477.x>.
- *Giubilei, F., Medda, E., Fagnani, C., Bianchi, V., De Carolis, A., Salvetti, M., & Stazi, M. A. (2008). Heritability of neurocognitive functioning in the elderly: evidence from an Italian twin study. *Age and Ageing*, 37, 640–646. <https://doi.org/10.1093/ageing/afn132>.
- Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., & Callahan, C. (2011). A test of the circumscription-of-limits hypothesis in geological bedrock mapping. *Journal of Experimental Psychology: General*, 14, 397–403. <https://doi.org/10.1037/a0025927>.
- Harris, L. J. (1978). Sex differences in spatial ability: Possible environmental, genetic, and neurological factors. In M. Kinsbourne (Ed.), *Asymmetrical Functions of the Brain* (pp. 405–522). Cambridge: Cambridge University Press.
- Hart, S. A., Petrill, S. A., Thompson, L. A., & Plomin, R. (2009). The ABCs of math: A genetic analysis of mathematics and its links with reading ability and general cognitive ability. *Journal of Educational Psychology*, 101(2), 388–402. <https://doi.org/10.1037/a0015115>.
- Hart, R. A., & Moore, G. T. (1973). The development of spatial cognition: A review. In R. M. Downs, & D. Stea (Eds.), *Image & Environment: Cognitive Mapping and Spatial Behavior* (pp. 246–288). New Brunswick, NJ: Aldine Transaction.
- Haworth, C. M., Wright, M. J., Luciano, M., Martin, N. G., de Geus, E. J., van Beijsterveldt, C. E., ... Kovas, Y. (2010). The heritability of general cognitive ability increases linearly from childhood to young adulthood. *Molecular Psychiatry*, 15(11), 1112–1120. <https://doi.org/10.1038/mp.2009.55>.
- Hegarty, M., & Waller, D. (2004). A dissociation between mental rotation and perspective-taking spatial abilities. *Intelligence*, 32(2), 175–191. <https://doi.org/10.1016/j.intell.2003.12.001>.
- Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. In P. Shah, & A. Miyake (Eds.), *The Cambridge Handbook of Visuospatial Thinking* (pp. 121–169). Cambridge University Press.
- Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences*, 19(1), 61–70. <https://doi.org/10.1016/j.lindif.2008.04.006>.
- *Hutson-Khalid, A. M. (2008). *The Phenotypic and Genetic Structure of Math Ability*. Doctoral dissertation Case Western Reserve University.
- Jensen, A. R. (1998). *The g factor: The science of mental ability*. 648. Westport, CT: Praeger.
- *Johnson, W., & Bouchard, T. J., Jr. (2007). Sex differences in mental abilities: g masks the dimensions on which they lie. *Intelligence*, 35(1), 23–39. <https://doi.org/10.1016/j.intell.2006.03.012>.
- *Johnson, W., Bouchard, T. J., Jr., McGue, M., Segal, N. L., Tellegen, A., Keyes, M., & Gottesman, I. I. (2007). Genetic and environmental influences on the Verbal-Perceptual-Image Rotation (VPR) model of the structure of mental abilities in the Minnesota study of twins reared apart. *Intelligence*, 35(6), 542–562. <https://doi.org/10.1016/j.intell.2006.10.003>.
- Keig, P. F., & Rubba, P. A. (1993). Translation of representations of the structure of matter and its relationship to reasoning, gender, spatial reasoning, and specific prior knowledge. *Journal of Research in Science Teaching*, 30(8), 883–903. <https://doi.org/10.1002/tea.3660300807>.
- Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, 31, 549–579. <https://doi.org/10.1080/15326900701399897>.
- Kyttälä, M., & Lehto, J. E. (2008). Some factors underlying mathematical performance: The role of visuospatial working memory and non-verbal intelligence. *European Journal of Psychology of Education*, 23(1), 77–92. <https://doi.org/10.1007/BF03173141>.
- Laski, E. V., Casey, B. M., Yu, Q., Dulaney, A., Heyman, M., & Dearing, E. (2013). Spatial skills as a predictor of first grade girls' use of higher level arithmetic strategies. *Learning and Individual Differences*, 23, 123–130. <https://doi.org/10.1016/j.lindif.2012.08.001>.
- Li, S. C., Lindenberger, U., Hommel, B., Aschersleben, G., Prinz, W., & Baltes, P. B. (2004). Transformations in the couplings among intellectual abilities and constituent cognitive processes across the life span. *Psychological Science*, 15(3), 155–163. <https://doi.org/10.1111/j.0956-7976.2004.01503003.x>.
- Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56(6), 1479–1498. <https://doi.org/10.2307/1130467>.
- Lohman, D. F. (1979). *Spatial ability: A review and reanalysis of the correlational literature*. Technical Report No. TR-8. Palo Alto, CA: Aptitude Research Project. School of Education, Stanford University.
- Lohman, D. F. (1996). Spatial ability and g. In I. Dennis, & P. Tapsfield (Eds.), *Human Abilities: Their Nature and Measurement* (pp. 97–116). Mahwah, NJ: Erlbaum.
- Lynn, R. (1994). Sex differences in intelligence and brain size: A paradox resolved. *Personality and Individual Differences*, 17(2), 257–271. [https://doi.org/10.1016/0191-8869\(94\)90030-2](https://doi.org/10.1016/0191-8869(94)90030-2).
- Lynn, R. (1998). Sex differences in intelligence: A rejoinder to Mackintosh. *Journal of Biosocial Science*, 30(4), 529–532.
- Lynn, R. (1999). Sex differences in intelligence and brain size: A developmental theory. *Intelligence*, 27(1), 1–12.
- Lynn, R., & Irwing, P. (2004). Sex differences on the progressive matrices: A meta-analysis. *Intelligence*, 32(5), 481–498. <https://doi.org/10.1016/j.intell.2004.06.008>.
- Mackintosh, N. J. (1996). Sex differences and IQ. *Journal of Biosocial Science*, 28(4), 558–571. <https://doi.org/10.1017/S0021932000022586>.
- *Madison, G., Mosing, M. A., Verweij, K. J., Pedersen, N. L., & Ullén, F. (2016). Common genetic influences on intelligence and auditory simple reaction time in a large Swedish sample. *Intelligence*, 59, 157–162. <https://doi.org/10.1016/j.intell.2016.10.001>.
- *Malanchini, M., Tosto, M. G., Garfield, V., Dirik, A., Czerwik, A., Arden, R., & Kovas, Y. (2016). Preschool drawing and school mathematics: the nature of the association. *Child Development*, 87, 929–943. <https://doi.org/10.1111/cdev.12520>.
- Malykh, S. B., Iskoldsky, N. V., & Gindina, E. D. (2005). Genetic analysis of IQ in young adulthood: A Russian twin study. *Personality and Individual Differences*, 38, 1475–1485. <https://doi.org/10.1016/j.paid.2003.06.014>.
- *McClearn, G. E., Johansson, B., Berg, S., Pedersen, N. L., Ahern, F., Petrill, S. A., & Plomin, R. (1997). Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science*, 276, 1560–1563. <https://doi.org/10.1126/science.276.5318.1560>.
- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychological Bulletin*, 86(5), 889–918. <https://doi.org/10.1037/0033-2909.86.5.889>.
- McGlone, M. S., & Aronson, J. (2006). Stereotype threat, identity salience, and spatial reasoning. *Journal of Applied Developmental Psychology*, 27(5), 486–493. <https://doi.org/10.1016/j.appdev.2006.06.003>.
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group (2009). Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA statement. *PLoS Med* 6: Article, e1000097. Available at <http://www.plosmedicine.org/article/info:doi/10.1371/journal.pmed.1000097>.
- *Mosing, M. A., Madison, G., Pedersen, N. L., Kuja-Halkola, R., & Ullén, F. (2014). Practice does not make perfect: No causal effect of music practice on music ability. *Psychological Science*, 25, 1795–1803. <https://doi.org/10.1177/0956797614541990>.
- *Mosing, M. A., Medland, S. E., McRae, A., Landers, J. G., Wright, M. J., & Martin, N. G. (2012). Genetic influences on life span and its relationship to personality: a 16-year follow-up study of a sample of aging twins. *Psychosomatic Medicine*, 74, 16–22. <https://doi.org/10.1097/PSY.0b013e3182385784>.
- *Neubauer, A. C., Spinath, F. M., Riemann, R., Borkenau, P., & Angleitner, A. (2000). Genetic and environmental influences on two measures of speed of information processing and their relation to psychometric intelligence: Evidence from the German observational study of adult twins. *Intelligence*, 28(4), 267–289. [https://doi.org/10.1016/S0160-2896\(00\)00036-2](https://doi.org/10.1016/S0160-2896(00)00036-2).
- *Panizzon, M. S., Vuoksimaa, E., Spoon, K. M., Jacobson, K. C., Lyons, M. J., Franz, C. E., ... Kremen, W. S. (2014). Genetic and environmental influences on general cognitive ability: Is g a valid latent construct? *Intelligence*, 43, 65–76. <https://doi.org/10.1016/j.intell.2014.01.008>.
- Parsons, T. D., Larson, P., Kratz, K., Thiebaut, M., Bluestein, B., Buckwalter, J. G., & Rizzo, A. A. (2004). Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia*, 42(4), 555–562. <https://doi.org/10.1016/j.neuropsychologia.2003.08.014>.
- *Petrill, S. A., Plomin, R., Berg, S., Johansson, B., Pedersen, N. L., Ahern, F., & McClearn, G. E. (1998). The genetic and environmental relationship between general and specific cognitive abilities in twins age 80 and older. *Psychological Science*, 9(3), 183–189. <https://doi.org/10.1111/1467-9280.00035>.
- Plomin, R., Fulker, D. W., Corley, R., & DeFries, J. C. (1997). Nature, nurture, and cognitive development from 1 to 16 years: A parent-offspring adoption study. *Psychological Science*, 8(6), 442–447. <https://doi.org/10.1111/j.1467-9280.1997.tb00458.x>.
- Polderman, T. J., Benyamin, B., De Leeuw, C. A., Sullivan, P. F., Van Bochoven, A., Visscher, P. M., & Posthuma, D. (2015). Meta-analysis of the heritability of human traits based on fifty years of twin studies. *Nature Genetics*, 47, 702–709. <https://doi.org/10.1038/ng.3285>.
- Pribyl, J. R., & Bodner, G. M. (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching*, 24(3), 229–240. <https://doi.org/10.1002/tea.3660240304>.
- *Read, S., Pedersen, N. L., Gatz, M., Berg, S., Vuoksimaa, E., Malmberg, B., ... & McClearn, G. E. (2006). Sex differences after all those years? Heritability of cognitive abilities in old age. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, 61, P137-P143. [10.1093/geronb/61.3.P137](https://doi.org/10.1093/geronb/61.3.P137).
- Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology*, 21(4), 387–399. <https://doi.org/10.1080/01443410120090786>.
- *Reynolds, C. A., Finkel, D., Gatz, M., & Pedersen, N. L. (2002). Sources of influence on rate of cognitive change over time in Swedish twins: An application of latent growth models. *Experimental Aging Research*, 28, 407–433. <https://doi.org/10.1080/03610730290103104>.
- *Rijsdijk, F. V., Vernon, P. A., & Boomsma, D. I. (1998). The genetic basis of the relation between speed-of-information-processing and IQ. *Behavioural Brain Research*, 95, 77–84. [https://doi.org/10.1016/S0166-4328\(97\)00212-X](https://doi.org/10.1016/S0166-4328(97)00212-X).
- Rimfeld, K., Kovas, Y., Dale, P. S., & Plomin, R. (2016). True grit and genetics: Predicting academic achievement from personality. *Journal of Personality and Social Psychology*, 111, 780–789. <https://doi.org/10.1037/pspp0000089>.
- *Schermer, J. A., Johnson, A. M., Jang, K. L., & Vernon, P. A. (2015). Phenotypic, genetic, and environmental relationships between self-reported talents and measured intelligence. *Twin Research and Human Genetics*, 18, 36–42. <https://doi.org/10.1017/thg.2014.80>.
- Schulze, R. (2004). *Meta-analysis: A comparison of approaches*. Ashland, OH, US: Hogrefe & Huber Publishers.
- Shakeshaft, N. G., & Plomin, R. (2015). Genetic specificity of face recognition. *Proceedings of the National Academy of Sciences*, 112, 12887–12892. <https://doi.org/10.1073/pnas.1212887112>.

- pnas.1421881112.
- *Shakeshaft, N. G., Rimfeld, K., Schofield, K. L., Selzam, S., Malanchini, M., Rodic, M., & Plomin, R. (2016). Rotation is visualisation, 3D is 2D: Using a novel measure to investigate the genetics of spatial ability. *Scientific Reports*, 6, 30545–30555. <https://doi.org/10.1038/srep30545>.
- Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, 171(3972), 701–703. <https://doi.org/10.1126/science.171.3972.701>.
- Sherman, J. A. (1967). Problem of sex differences in space perception and aspects of intellectual functioning. *Psychological Review*, 74(4), 290–299. <https://doi.org/10.1037/h0024723>.
- *Shikishima, C., Hiraishi, K., Yamagata, S., Sugimoto, Y., Takemura, R., Ozaki, K., ... Ando, J. (2009). Is g an entity? A Japanese twin study using syllogisms and intelligence tests. *Intelligence*, 37(3), 256–267. <https://doi.org/10.1016/j.intell.2005.11.005>.
- *Singer, J. J., MacGregor, A. J., Cherkas, L. F., & Spector, T. D. (2006). Genetic influences on cognitive function using the Cambridge Neuropsychological Test Automated Battery. *Intelligence*, 34(5), 421–428. <https://doi.org/10.1016/j.intell.2005.11.005>.
- Spearman, C., & Jones, L. L. W. (1950). *Human ability: A continuation of "The abilities of man."* London: Macmillan.
- Spencer, S. J., Steele, C. M., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, 35, 4–28. <https://doi.org/10.1006/jesp.1998.1373>.
- Stafford, R. E. (1961). Sex differences in spatial visualization as evidence of sex-linked inheritance. *Perceptual and Motor Skills*, 13(3), 428. <https://doi.org/10.2466/pms.1961.13.3.428>.
- *Svedberg, P., Gatz, M., & Pedersen, N. L. (2009). Genetic and environmental mediation of the associations between self-rated health and cognitive abilities. *Experimental Aging Research*, 35(2), 178–201. <https://doi.org/10.1080/03610730902720372>.
- *Swagerman, S. C., de Geus, E. J., Kan, K. J., van Bergen, E., Nieuwboer, H. A., Koenis, M. M., ... & Boomsma, D. I. (2016). The Computerized Neurocognitive Battery: Validation, aging effects, and heritability across cognitive domains. *Neuropsychology*, 30, 53–64. <https://doi.org/10.1037/neu0000248>.
- Thurstone, L. L. (1950). Some primary abilities in visual thinking. *Proceedings of the American Philosophical Society*, 94(6), 517–521.
- Toivainen, T., Pannini, G., Papageorgiou, K. A., Malanchini, M., Rimfeld, K., Shakeshaft, N., & Kovas, Y. (2018). Prenatal testosterone does not explain sex differences in spatial ability. *Scientific Reports*, 8, 13653–13661. <https://doi.org/10.1038/s41598-018-31704-y>.
- *Tosto, M. G., Hanscombe, K. B., Haworth, C., Davis, O. S., Petrill, S. A., Dale, P. S., ... Kovas, Y. (2014). Why do spatial abilities predict mathematical performance? *Developmental Science*, 17, 462–470. <https://doi.org/10.1111/desc.12138>.
- *Trzaskowski, M., Shakeshaft, N. G., & Plomin, R. (2013). Intelligence indexes generalist genes for cognitive abilities. *Intelligence*, 41(5), 560–565. <https://doi.org/10.1016/j.intell.2013.07.011>.
- Tucker-Drob, E. M., Briley, D. A., & Harden, K. P. (2013). Genetic and environmental influences on cognition across development and context. *Current Directions in Psychological Science*, 22(5), 349–355. <https://doi.org/10.1177/0963721413485087>.
- Tucker-Drob, E. M., & Briley, D. A. (2014). Continuity of genetic and environmental influences on cognition across the life span: A meta-analysis of longitudinal twin and adoption studies. *Psychological Bulletin*, 140(4), 949–978. <https://doi.org/10.1037/a0037091>.
- *Tucker-Drob, E. M., Reynolds, C. A., Finkel, D., & Pedersen, N. L. (2014). Shared and unique genetic and environmental influences on aging-related changes in multiple cognitive abilities. *Developmental Psychology*, 50, 152–166. <https://doi.org/10.1037/a0032468>.
- *Tucker-Drob, E. M., Briley, D. A., Engelhardt, L. E., Mann, F. D., & Harden, K. P. (2016). Genetically-mediated associations between measures of childhood character and academic achievement. *Journal of Personality and Social Psychology*, 111, 790–815. <https://doi.org/10.1037/pspp0000098>.
- Vandenberg, S. G., & Kuse, A. R. (1979). Spatial ability: A critical review of the sex-linked major gene hypothesis. In M. Wittig, & A. C. Petersen (Eds.). *Sex-Related Differences in Cognitive Functioning: Developmental Issues* (pp. 67–96). New York: Academic Press.
- *Vasilopoulos, T., Kremen, W. S., Kim, K., Panizzon, M. S., Stein, P. K., Xian, H., ... Franz, C. E. (2012). Untreated hypertension decreases heritability of cognition in late middle age. *Behavior Genetics*, 42, 107–120. <https://doi.org/10.1007/s10519-011-9479-9>.
- *Vasilopoulos, T., Mack, H. A., McClearn, G. E., Berg, S., Johansson, B., & Vogler, G. P. (2010). Stability of genetic influences on pulmonary function in a longitudinal study of octogenarian twins. *American Journal of Human Biology*, 22(3), 375–377. <https://doi.org/10.1002/ajhb.21003>.
- 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. <https://doi.org/10.1037/0033-2909.117.2.250>.
- Vuoksima, 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. <https://doi.org/10.1177/0956797610376075>.
- *Vuoksima, E., Viken, R. J., Hokkanen, L., Tuulio-Henriksson, A., Rose, R. J., & Kaprio, J. (2010). Are there sex differences in the genetic and environmental effects on mental rotation ability? *Twin Research and Human Genetics*, 13, 437–441. <https://doi.org/10.1375/twin.13.5.437>.
- 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), 817–831. <https://doi.org/10.1037/a0016127>.
- *Wallace, G. L., Lee, N. R., Prom-Wormley, E. C., Medland, S. E., Lenroot, R. K., Clasen, L. S., ... & Giedd, J. N. (2010). A bivariate twin study of regional brain volumes and verbal and nonverbal intellectual skills during childhood and adolescence. *Behavior Genetics*, 40(2), 125–134. <https://doi.org/10.1007/s10519-009-9329-1>.
- Wilson, J. R., & Vandenberg, S. G. (1978). Sex differences in cognition: evidence from the Hawaii family study. In T. E. McGill (Ed.). *Sex and behavior: stages and prospectus*. Plenum, New York (pp. 317–335).
- *Woodley, M. A., & Madison, G. (2015). The association between g and K in a sample of 4246 Swedish twins: A behavior genetic analysis. *Personality and Individual Differences*, 74, 270–274. <https://doi.org/10.1016/j.paid.2014.10.027>.