American Psychological Association

Psychological Bulletin

0033-2909/19/$12.00

2019, Vol. 145, No. 6, 537–565
http://dx.doi.org/10.1037/bul0000191

The Development of Gender Differences in Spatial Reasoning: A Meta-Analytic Review

Jillian E. Lauer
New York University and Emory University

Eukyung Yhang
Yale University

Stella F. Lourenco
Emory University

Gender differences in spatial aptitude are well established by adulthood, particularly when measured by tasks that require the mental rotation of objects (Linn & Petersen, 1985; Voyer, Voyer, & Bryden, 1995). Although the male advantage in mental rotation performance represents one of the most robust gender differences in adult cognition, the developmental trajectory of this male advantage remains a topic of considerable debate. To address this debate, we meta-analyzed 303 effect sizes pertaining to gender differences in mental rotation performance among 30,613 children and adolescents. We found significant developmental change in the magnitude of the gender difference: A small male advantage in mental rotation performance first emerged during childhood and then subsequently increased with age, reaching a moderate effect size during adolescence. Procedural factors, including task and stimulus characteristics, also accounted for variability in reported gender differences, even when controlling for the effect of age. These results demonstrate that both age and procedural characteristics moderate the magnitude of the gender difference in mental rotation throughout development.

Public Significance Statement
This meta-analysis documents the development of gender differences in spatial reasoning, finding that boys first outperform girls on mental rotation tasks during primary school and that this male advantage subsequently increases with age into adolescence. Because children's spatial skills predict their later science, technology, engineering, and math (STEM) achievement, addressing the gender difference in mental rotation early in childhood may help to minimize later gender disparities in STEM success.

Keywords: gender difference, mental rotation, meta-analysis, spatial development

Supplemental materials: http://dx.doi.org/10.1037/bul0000191.supp

Gender differences in cognition have long captured the interests of the scientific community and general public alike, yet the existence of cognitive gender differences remains the subject of considerable debate (Hyde, 2014). In a seminal review on this topic, Maccoby and Jacklin (1974) concluded that gender differences are indeed present in mathematical, spatial, and verbal reasoning, with measures of spatial competence producing the most profound gender differences in cognitive performance. Subsequent meta-analyses partially supported these early conclusions: whereas gender differences in verbal and mathematical abilities tend to be small or nonexistent (Hyde & Linn, 1988; Lindberg, Hyde, Petersen, & Linn, 2010), men tend to outperform women on some spatial reasoning measures, namely mental rotation tasks that assess one's ability to generate and rotate mental images of objects (Linn & Petersen, 1985; Voyer et al., 1995). In fact, the male advantage in mental rotation performance represents one of the largest gender differences reported in the contemporary cognitive literature (Hyde, 2014; Miller & Halpern, 2014), with meta-analytic findings suggesting that only ~25% of women perform above the mean for men on standard mental rotation measures

This article was published Online First April 11, 2019.

Jillian E. Lauer, Department of Psychology, New York University, and Department of Psychology, Emory University; Eukyung Yhang, Yale Child Study Center, School of Medicine, Yale University; Stella F. Lourenco, Department of Psychology, Emory University.

The authors wish to thank the many researchers who provided additional data on request. The authors also thank Kim Wallen, Scott Lilienfeld, Lynne Nygaard, Patricia Bauer, and members of the Bauer Memory Development Lab, for valuable comments. This research received support from the Emory College of Arts and Sciences (Jillian E. Lauer & Eukyung Yhang) and a Scholars Award from the John Merck Fund (Stella F. Lourenco).

Correspondence concerning this article should be addressed to Jillian E. Lauer, Department of Psychology, New York University, 6 Washington Place, New York, NY 10003. E-mail: jillian.lauer@nyu.edu

(*d*s = 0.57–0.73; Linn & Petersen, 1985; Maeda & Yoon, 2013; Voyer et al., 1995).

Gender differences in mental rotation performance have sparked considerable interest among psychologists not only because of their magnitude but also because of their potential role in the gender gap in science, technology, engineering, and mathematics (STEM) fields. Cross-sectional studies have indicated that spatial skills positively relate to achievement across diverse STEM disciplines in adulthood (e.g., Guillot, Champely, Batier, Thiriet, & Collet, 2007; Jones & Burnett, 2008; Kozhevnikov, Motes, & Hegarty, 2007; Stieff, 2007; Wei, Yuan, Chen, & Zhou, 2012; Wu & Shah, 2004), and concurrent associations between spatial ability and STEM achievement have been documented in childhood and adolescence as well (e.g., Ganley, Vasilyeva, & Dulaney, 2014; Gilligan, Hodgkiss, Thomas, & Farran, 2018; Hodgkiss, Gilligan, Tolmie, Thomas, & Farran, 2018; Tracy, 1990). Longitudinal studies on this topic have provided further evidence that spatial reasoning abilities foster success in male-dominated STEM fields. For example, superior spatial skills in adolescence are associated with greater academic, occupational, and creative success in STEM domains in adulthood (Kell, Lubinski, Benbow, & Steiger, 2013; Wai, Lubinski, & Benbow, 2009; Wai, Lubinski, Benbow, & Steiger, 2010). This longitudinal relation has also been documented during childhood, with a number of studies reporting that children's early spatial skills predict their later math performance (e.g., Casey et al., 2015; Gilligan et al., 2017; Verdine, Golinkoff, Hirsh-Pasek, & Newcombe, 2017). For example, individual differences in mental rotation abilities during the first years of elementary school have been shown to predict children's math achievement when measured in subsequent years (Geer, Quinn, & Ganley, 2018; Gunderson, Ramirez, Beilock, & Levine, 2012). Furthermore, one recent study found that individual differences in infants' performance on a mental rotation task during the first year of life predicted their later mathematical competence when measured at preschool age, even when accounting for general cognitive abilities (Lauer & Lourenco, 2016). Taken together, these findings provide compelling evidence that spatial abilities during childhood and adolescence represent important predictors of later achievement in STEM disciplines.

Importantly, prior research has demonstrated that the gender difference in mental rotation relates to the gender disparity in STEM attainment. For example, Ganley and colleagues (2014) found that a male advantage in middle-school students' mental rotation performance accounted for gender differences in their science achievement as measured by standardized testing in physical sciences, technology, and engineering courses. Likewise, Casey and colleagues (1995) found that a male advantage in mental rotation performance among college-bound students contributed to male students' relatively higher SAT Math scores. Thus, the gender difference in mental rotation is one factor that contributes to gender gaps in students' STEM achievement by early adolescence, a formative time for the development of students' academic interests and professional aspirations in STEM fields (Wang & Degol, 2013). Research that characterizes the developmental origins and trajectory of the gender difference in mental rotation during childhood therefore has the potential to advance our understanding of one gender-related obstacle to STEM success and, in doing so, inform interventions aimed at promoting gender equity in STEM.

Review of Developmental Findings

Meta-analytic reviews of gender differences in spatial reasoning have consistently documented a robust male advantage in mental rotation performance in adulthood, but have left open questions regarding the developmental trajectory of this gender difference. In the most comprehensive meta-analysis on this topic to date, Voyer and colleagues (1995) reported that the gender difference in mental rotation performance was small in magnitude prior to adolescence ($d = 0.33$) and increased across the teenage years to reach a moderate effect size by adulthood ($d = 0.66$; Voyer et al., 1995). These findings introduce the intriguing possibility that the male advantage undergoes a protracted developmental course between childhood and adulthood; yet, a number of other meta-analyses have found no evidence of age-related change in the gender difference, instead reporting that the male advantage is large in magnitude across development ($d = 0.57–0.73$; Linn & Petersen, 1985; Maeda & Yoon, 2013; Voyer, 2011). In addition to these inconsistencies, earlier meta-analyses on gender differences in mental rotation performance included very few samples of children, preventing analysis of the developmental trajectory of the male advantage during early and middle childhood. However, in the two decades since Voyer and colleagues (1995) published their meta-analysis, research on the development of mental rotation abilities has flourished, particularly among school-age children. Hence, a meta-analytic review of the extant literature would have power to provide insight into the developmental emergence and continuity of gender differences in mental rotation skills.

As stated above, there have been considerable efforts in recent years to document gender differences in children's mental rotation performance. For example, Levine and colleagues (1999) designed the Children's Mental Transformation Task (CMTT), a task that requires children to mentally assemble two-dimensional shapes via the translation and rotation of their component parts (see Figure 1A), to examine gender differences in children's mental rotation abilities across the preschool years. The authors found that a small, but significant, male advantage in performance became detectable between age 4 and 5 ($d = 0.32$ in 4.5- to 5-year-olds), suggesting

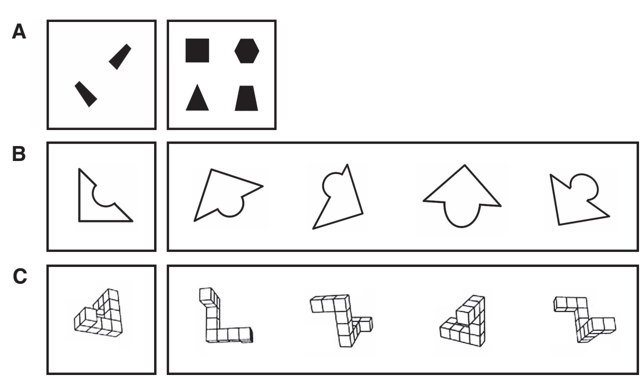

Figure 1. Example mental rotation task items. In each task, children are presented with the target image(s) in the left-hand box and must select from the response image(s) in the right-hand box. Items are similar to those in: (A) the Children's Mental Transformation Task (CMTT; Levine et al., 1999), (B) the Primary Mental Abilities-Space Relations task (PMA-SR; Thurstone & Thurstone, 1943), and (C) the Vandenberg-Kuse Mental Rotation Test (VMRT; Vandenberg & Kuse, 1978).

that the gender difference in spatial reasoning may emerge early in development. Nonetheless, later research administering the CMTT to preschool-aged children presented variable evidence of a gender difference in performance, with some studies replicating the male advantage (Ehrlich, Levine, & Goldin-Meadow, 2006; Levine, Ratliff, Huttenlocher, & Cannon, 2012; Pruden, Levine, & Huttenlocher, 2011), but others reporting no gender difference (Harris, Newcombe, & Hirsh-Pasek, 2013; Hawes, Moss, Caswell, & Poliszczuk, 2015; Lauer & Lourenco, 2016). These mixed findings of gender differences in preschoolers' CMTT performance parallel contradictory results obtained when administering other mental rotation tasks to young children: some studies have reported a significant male advantage (e.g., Frick, Ferrara, & Newcombe, 2013; Rosser, Ensing, & Mazzeo, 1985), but others have reported no gender difference (e.g., Frick, Hansen, & Newcombe, 2013; Klein, Adi-Japha, & Hakak-Benizri, 2010) or even a significant female advantage in performance (Krüger & Krist, 2009; Lehmann, Quaiser-Pohl, & Jansen, 2014). Therefore, the presence of a gender difference during early childhood remains contested.

Studies examining mental rotation abilities during middle childhood have also produced notable inconsistencies, with some studies reporting no gender differences in performance (ds < 0.2; e.g., Lauer, Esposito, & Bauer, 2018; Wei et al., 2012) but others documenting a substantial male advantage (ds > 0.8, e.g., De Lisi & Wolford, 2002; Geiser, Lehmann, & Eid, 2008). Evidence of developmental change during middle childhood has also been mixed. Whereas some findings suggest that there are age-related increases in the magnitude of the male advantage across middle childhood (e.g., Choi, 2000; Hoyek, Collet, Fargier, & Guillot, 2012), others suggest that the effect of gender exhibits stability within this age range (e.g., Amponsah, 2000; Geiser, Lehmann, Corth, & Eid, 2008; Lauer et al., 2018). These discrepant findings, in combination with conflicting evidence on preschool-aged children, have prevented definitive conclusions regarding the presence, magnitude, and stability of the male advantage in mental rotation prior to adolescence.

Inconsistent evidence of gender differences in the spatial development literature has limited our ability to assess the validity of varying theoretical accounts on their origins, resulting in continued debate over the factors that produce and sustain the male advantage in mental rotation performance across development. One prominent perspective proposes that gender-related variation in mental rotation performance results from the enduring influence of prenatal androgens on visuospatial processing (Grimshaw, Sitarenios, & Finegan, 1995; Hampson, Rovet, & Altmann, 1998; Resnick, Berenbaum, Gottesman, & Bouchard, 1986), suggesting that the male advantage in mental rotation could emerge as early as the first year of life and persist throughout development. In contrast, other accounts posit that gender differences in spatial abilities arise in response to environmental inputs whose effects are expected to accumulate with age, such as experience with activities that foster spatial thinking (e.g., play with blocks, sports, videogames; Baenninger & Newcombe, 1989; Serbin, Zelkowitz, Doyle, Gold, & Wheaton, 1990) or exposure to societal gender stereotypes regarding the respective spatial competence of women and men (e.g., Moè, 2012; Titze, Jansen, & Heil, 2010b). Given that these theoretical perspectives proffer differing predictions about the developmental emergence and continuity of the male advantage in mental rotation, research characterizing its presence and magnitude between early childhood and adulthood will be critical to identifying its causal determinants.

Influential Factors in the Gender Difference in Mental Rotation

Although previous meta-analyses have presented conflicting findings regarding the development of gender differences in spatial skills, they have provided clear evidence that procedural variables influence their magnitude, generating new hypotheses about the cognitive and psychosocial factors that underpin the male advantage in spatial reasoning performance. Most notably, previous meta-analyses have demonstrated that the magnitude of the gender difference in mental rotation varies according to the task used to evaluate performance. For example, Linn and Petersen (1985) reported that the Vandenberg-Kuse Mental Rotation Task (VMRT; Vandenberg & Kuse, 1978) produced larger gender differences in performance than the Primary Mental Abilities-Spatial Relations subtest (PMA-SR; Thurstone & Thurstone, 1943), a finding later replicated in the meta-analysis of Voyer and colleagues (1995). Explanations for these task differences have primarily focused on stimulus characteristics: whereas the PMA-SR consists of simple two-dimensional (2D) line-drawings rotated around the picture plane (Figure 1B), the VMRT consists of three-dimensional (3D) cube figures and their mirrored images rotated in depth (Figure 1C). Consequently, it has been proposed that the VMRT produces relatively larger gender differences because its stimuli are 3D, rather than 2D, as in the PMA-SR (cf. Linn & Petersen, 1985). This argument has been supported by primary studies reporting that other standard 2D tasks also produce smaller gender differences relative to the VMRT (e.g., Peters et al., 1995; Sanders, Soares, & D'Aquila, 1982), indicating that stimulus dimensionality may be an important moderator of the male advantage. However, former meta-analyses have not examined the effects of stimulus dimensionality on reported effect sizes, leaving open the possibility that other task-related factors account for the observed differences between the PMA-SR and VMRT.

An alternative account of the task effect documented by Linn and Petersen (1985) is that the PMA-SR and VMRT are differentially amenable to problem-solving strategies that vary in their use between genders. Items on the PMA-SR may more readily lend themselves to analytical problem-solving strategies that involve comparing features of stimuli to identify matching characteristics (e.g., shape; see Figure 1B), whereas items on the VMRT may be more effectively solved via spatial strategies that involve the active visualization of object rotation (see Figure 1C), and females may be relatively more likely to use the former strategy type than the latter. Prior research lends support to this contention. In one study, Geiser and colleagues (2006) utilized multivariate latent-class analyses to categorize adolescents by strategy use on differing VMRT items. One group of adolescents correctly solved items in which the target and distractor figures were structurally distinct (i.e., allowing for the use of feature-based comparison strategies), yet displayed chance performance on items in which distractors were the mirrored image of the target figure (i.e., reducing the utility of feature comparison). Adolescents in this group not only exhibited decreased performance because of their reliance on feature-based strategies, but they were also more likely to be female, resulting in a male advantage in task performance. Geiser et al. (2008) later replicated these results in middle-schoolchildren, suggesting that relying on feature-based comparison strategies may disadvantage girls' mental rotation performance during development. Consistent with this possibility, Ehrlich and colleagues (2006) found that girls gestured about movement less often than boys when solving mental rotation tasks by age 5. However, previous meta-analyses have not assessed whether tasks that allow for feature-based comparison (e.g., tasks with structurally different target and choice stimuli) produce smaller effect sizes than tasks that necessitate spatial strategies (e.g., tasks that require discriminating between rotated mirrored images), so it is not known whether this distinction underlies differences in the size of the male advantage across tasks.

Another potential explanation for task-related variability in the magnitude of the male advantage relates to gender differences in the familiarity of task stimuli. Some have argued that males have a greater advantage over females on the VMRT because boys have greater experience with spatially relevant toys (e.g., blocks, LEGOs) that resemble the abstract cube stimuli often presented during standard mental rotation tasks (e.g., Shepard & Metzler, 1971), and this experience may facilitate their performance. This conjecture has been supported by evidence that greater experience with abstract stimuli during mental rotation tasks leads to performance gains (Bethell-Fox & Shepard, 1988) as well as findings that gender differences in adults' mental rotation abilities are attenuated when familiar animate stimuli, rather than cube figures with similar internal structures, are used to evaluate performance (Alexander & Evardone, 2008). Nevertheless, the effect of stimulus type on gender differences has not previously been examined via meta-analysis and has only rarely been considered in primary studies. Therefore, we do not currently know the extent to which stimulus type influences the magnitude of the gender difference in mental rotation or whether its role may change across development.

In addition to the task and stimulus characteristics outlined above, results of previous meta-analyses suggest that testing conditions may influence the magnitude of gender differences in mental rotation performance. For example, Voyer (2011) reported that tasks administered with a strict time limit produced effect sizes that were, on average, more than double the size of those produced by tasks administered with no time limit ($d = 1.03$ vs. $d = 0.51$; see also Maeda & Yoon, 2013). These findings indicate that women's performance may benefit more than men's performance when additional time is provided to solve mental rotation tasks, potentially because of women's greater reliance on analytic problem-solving strategies that require serial processing (Heil & Jansen-Osmann, 2008a). Conversely, women may experience relatively greater test-related anxiety in response to prohibitive time constraints, resulting in performance decrements when strict time limits are imposed (Maeda & Yoon, 2013; Voyer, 2011), a supposition in line with evidence that female participants experience greater anxiety regarding spatial tasks across development (e.g., Lauer et al., 2018; Lyons et al., 2018). Furthermore, Voyer and colleagues (1995) reported that the male advantage in spatial performance was greater when tasks were administered individually rather than in group testing environments, which the authors argued could reflect the detrimental effects of women's anxiety about their spatial skills being directly evaluated. In addition to time constraints and test settings, preliminary evidence that gender differences are attenuated when mental rotation abilities are evaluated via individualized computer-based testing, rather than canonical paper-and-pencil assessments (Maeda & Yoon, 2013; Monahan, Harke, & Shelley, 2008), is congruent with the notion that testing conditions influence the magnitude of the gender difference in performance. Thus, further examination of the moderating effects of various testing conditions may advance our understanding of the factors that contribute to the male advantage in mental rotation performance and in doing so provide insight into the mechanisms that drive its development.

As outlined above, there is compelling evidence that procedural variables influence the magnitude of reported effect sizes in adulthood. Consequently, age-related variation in the procedures used to evaluate mental rotation abilities may cloud the developmental trajectory of the gender difference in performance. That is, if procedural variables affect the magnitude of the male advantage and the procedures used to assess mental rotation performance vary with age, any analysis aimed at characterizing developmental change in the gender difference must account for potential methodological confounds related to the testing procedure. Prior research has not addressed the respective roles of age and procedural factors in the gender difference in mental rotation during childhood, so we do not yet know the extent to which procedural variables moderate the male advantage in children's task performance or account for age-related change in the magnitude of this gender difference. Thus, exploring the moderating effects of procedural characteristics throughout development not only has the potential to further our understanding of factors that produce and sustain the male advantage but also will be necessary to charting the developmental trajectory of the gender difference in mental rotation skills.

Present Meta-Analysis

In the present meta-analysis, we investigated developmental change in the male advantage in mental rotation performance between the ages of 3 and 17 years and evaluated the moderating effects of procedural factors within this age range. The aims of this meta-analysis were threefold: (a) to estimate the age at which the male advantage in mental rotation emerges, (b) to chart its developmental trajectory throughout childhood and adolescence, and (c) to characterize the influence of procedural moderators on its magnitude across development.

Defining Mental Rotation

For the purpose of this meta-analysis, mental rotation was defined as the process of visualizing the rigid transformation and rotation of objects and/or object parts. Mental rotation tasks therefore included measures that required children to perform object completion through the mental translation and/or rotation of component shapes (e.g., Levine et al., 1999; Thurstone & Thurstone, 1943; Figure 1A & Figure 1B), to discriminate between objects and their rotated mirrored images (e.g., Vandenberg & Kuse, 1978; Figure 1C), and/or to solve spatial analogies by visualizing the rotation of novel objects (e.g., Bodner & Guay, 1997). As in previous meta-analyses (e.g., Linn & Petersen, 1985; Voyer et al., 1995), tasks that require physical construction; mental folding, cutting, or nonrigid object transformations; and/or the identification of shapes embedded within complex spatial configurations were not classified as mental rotation tasks and were not included in the meta-analysis.

Identifying Potential Moderators

As previously discussed, we know relatively little about how procedural factors affect the expression of gender differences in mental rotation and how age-related variation in the procedures used to test mental rotation skills could influence the development of the male advantage. Hence, we selected variables for consideration as potential moderators in our meta-analysis by identifying procedural factors that relate to the magnitude of the gender difference in adulthood and that vary in their implementation across childhood and adolescence. Ultimately, we selected and coded four variables that characterized the mental rotation task and its stimuli and three variables that characterized the testing conditions. The four task characteristics were the name of the mental rotation task (i.e., CMTT, PMA-SR, or VMRT), the dimensionality of its stimuli (i.e., 2D or 3D), the type of discrimination required for successful completion of the task (i.e., whether mirror discrimination was required for successful task completion or not), and the type of stimuli presented (i.e., abstract or animate). The three testing conditions were presentation method (i.e., digital or paper-and-pencil), test setting (i.e., individual or group testing), and time constraints (i.e., whether the task was administered with a time limit or not). See Appendix B for further details of coding procedure.

Meta-Analytic Procedure

Inclusion Criteria

The following criteria were used to identify articles for inclusion in the meta-analysis.

1. Mental rotation measure. The article must have included a study in which a behavioral measure of mental rotation ability was administered and response accuracy data obtained from that measure were reported.

2. Age range of interest. The article must have reported data obtained from at least one sample of children or adolescents with a mean age between 3 and 17 years and an age range of less than 5 years.

3. Population of interest. The article must have focused solely on healthy, typically developing children or adolescents. Articles that examined clinical populations, such as individuals with neuropsychiatric, physical, or learning differences (e.g., Down's syndrome, spina bifida, dyslexia), or individuals whose spatial abilities were expected to deviate from the norm as a result of other differences (e.g., highly gifted youth) were not included.

4. Sufficient statistics. The article must have reported sufficient statistics to compute an effect size for a gender difference in mental rotation performance using data obtained during the first administration of a mental rotation task for a given sample, or the article's author(s) must have provided these statistics upon request. Necessary statistics for effect size computation consist of means and variance estimates for the mental rotation performance of females and males or a test statistic that quantified the gender difference in their performance (i.e., F, t, d, r, or p).

Literature Search

Articles were identified for potential inclusion in four stages (see Figure 2). First, we conducted computerized literature searches of all records contained in the PsycINFO, PubMed, and ERIC databases using the following search term: ("mental rotation" OR "mental transformation" OR "spatial rotation" OR "spatial transformation"). This search term was also used to identify unpublished student works for potential inclusion via the ProQuest Dissertations and Theses repository. Database searches were conducted in July of 2015, and yielded 4,135 hits consisting of 2,760 unique articles. Next, we reviewed the reference lists of five previous meta-analyses that examined various aspects of spatial cognition (Linn & Petersen, 1985; Maeda & Yoon, 2013; Reilly & Neumann, 2013; Uttal et al., 2013; Voyer et al., 1995), identifying 14 additional articles for consideration. We then conducted forward searches of seminal papers on mental rotation in children, including articles that presented novel mental rotation paradigms (e.g., Levine et al., 1999), obtaining 20 more articles. Finally, additional unpublished data were obtained via Internet searches. In total, 2,798 articles were identified for consideration.

Selection Strategy

From the sample of 2,798 articles, we identified articles for inclusion in the meta-analysis in three steps, summarized below.

Step 1. We reviewed the titles and journals of all records obtained during the search process, excluding articles that were not expected to contain original data (e.g., review articles), were not published in English, did not pertain to psychology or education (e.g., examined spatial transformation in urban development), and did not study human populations. Following this precursory exclusion process, 2,533 records remained for further consideration.

Step 2. We reviewed the abstracts of the remaining records, discarding articles that did not meet our four inclusion criteria (outlined above). Records excluded in this step primarily corresponded to articles examining adult or clinical populations. At the conclusion of Step 2, 334 records remained for further consideration.

Step 3. Full-text versions of 301 articles were obtained and reviewed in entirety; full-text versions of the remaining 33 articles were unobtainable or were not available in English. During full-text review, we excluded articles that did not meet the first three inclusion criteria outlined above (73 articles) or did not report original data (22 articles). An additional 96 articles met the first three inclusion criteria, but did not report sufficient statistics for effect size computation (the fourth inclusion criterion outlined above; see Supplemental Material for further details on articles excluded because of insufficient statistics). In cases in which these articles were published in the 10 years preceding our literature search, we contacted authors to request the necessary statistics for inclusion; authors provided sufficient statistics for 18 articles. Thus, 128 articles were included in the meta-analysis.

Data Extraction

Effect size calculation. For each gender difference that met our inclusion criteria, we extracted the necessary statistics for

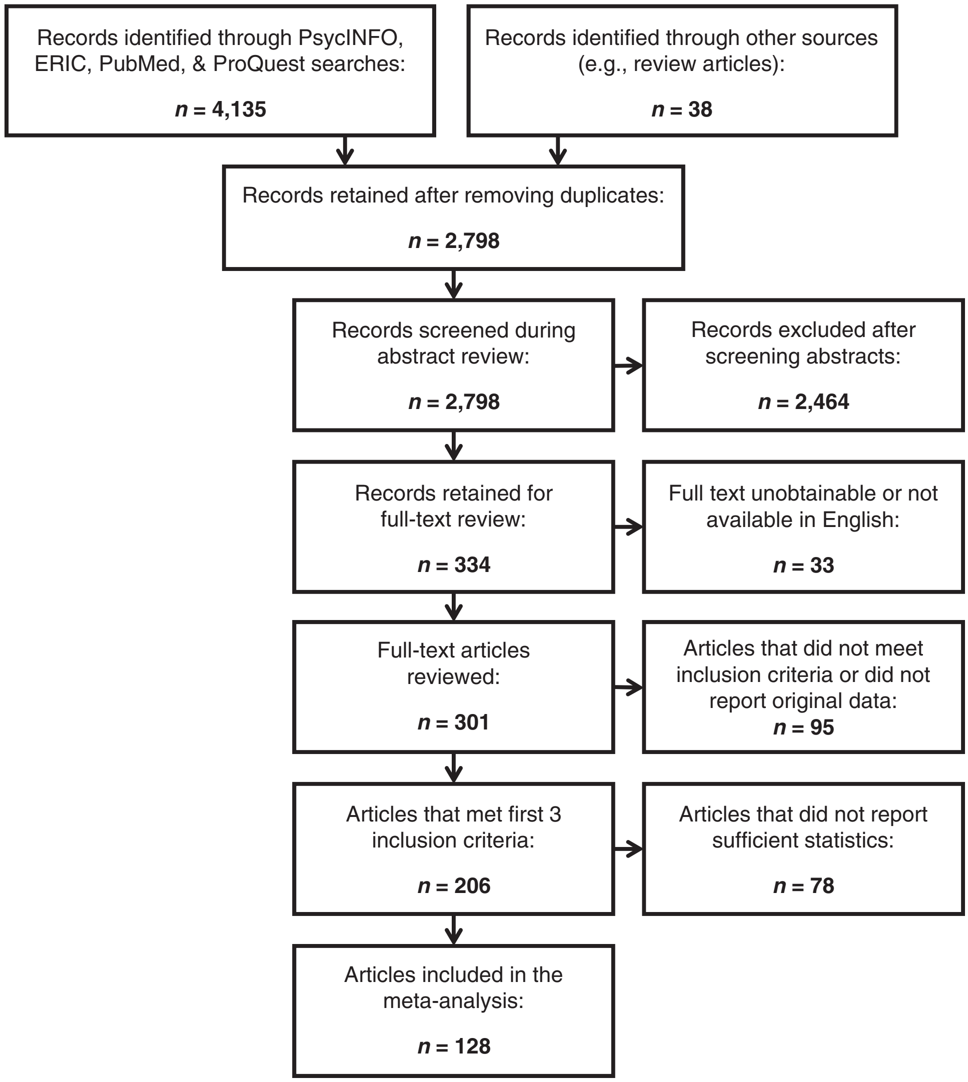

Figure 2. Schematic illustrating the process used to identify articles for inclusion in the meta-analysis and the number of records (n) reviewed at each stage.

effect size computation, including the sample sizes for females and males and descriptive or test statistics that quantified the difference in their mental rotation performance. We then computed Hedges' g, which represents the standardized mean difference between the performance of females and males, to measure the effect size. In these calculations, a g greater than 0 indicated a male advantage in task performance whereas a g less than 0 indicated a female advantage. The equations used to calculate Hedges' g can be found in the Supplemental Material.

To eliminate any potential influence of training or retesting effects on reported gender differences, effect sizes were only included in the meta-analysis if they were calculated using data obtained during the initial task administration for a given sample (see fourth inclusion criterion above). Thus, in longitudinal studies in which children were administered the same mental rotation task at multiple time points (e.g., at the beginning and end of a school semester), effect sizes were calculated using data obtained at the first time point. Similarly, when children participated in a training intervention aimed at developing spatial reasoning, effect sizes were calculated using data collected prior to the intervention.

The 128 articles included in the meta-analysis ultimately produced 303 effect sizes derived from 251 unique samples composed of 30,613 individuals (see Appendix A).

Coding. We coded article-level, sample-level, and measure-level characteristics to describe the meta-analytic sample (see Table 1), to examine the effects of age and procedural moderators of interest (see Table 2), and to assess publication bias in the mental rotation literature. See Appendix A for the meta-analytic dataset and Appendix B for details on the coding procedure for all coded variables.

Article-level variables. We coded year of publication and publication status. Year of publication ranged from 1947 to 2016, and the majority of articles (70%) were published after the year 2000 (see

Table 1
Summary of Study-Level and Sample-Level Characteristics, Including the Mean Weighted Effect Size (g), the Number of Articles (m), the Number of Effect Sizes (k), and the Mean and Range of the Mean Sample Ages (in Years) Corresponding to Each Characteristic

Characteristic	*g* [95% CI of *g*]	*m*	*k*	*M* age (range)
Mean sample age				
3 to 7 years	.20 [.12, .28]	46	112	6.21 (3.50–7.96)
8 to 12 years	.40 [.29, .50]	41	90	10.20 (8.00–12.50)
13 to 17 years	.54 [.45, .63]	53	101	15.57 (13.00–17.99)
Publication status				
Published	.42 [.34, .50]	103	245	10.26 (3.50–17.99)
Unpublished	.28 [.17, .39]	25	58	11.59 (4.54–17.00)
Publication year				
Before 1990	.50 [.22, .78]	15	31	12.17 (3.96–17.50)
1990s	.47 [.32, .62]	23	44	11.73 (4.23–17.60)
2000s	.47 [.33, .60]	37	97	12.06 (5.50–17.97)
2010s	.27 [.20, .35]	53	131	8.57 (3.50–17.99)
Region of testing				
Europe	.45 [.32, .59]	42	100	11.51 (5.00–17.99)
North America	.34 [.25, .44]	69	156	9.84 (3.50–17.50)
Other	.44 [.29, .59]	22	47	10.64 (4.88–17.50)

Table 1). Eighty percent of articles were obtained from peer-reviewed publications; the remaining 20% consisted of student theses, conference proceedings, or unpublished reports provided by authors.

Sample-level variables. For each effect size extracted from the literature, we coded the mean age of the participants as well as the country in which mental rotation was assessed. Mean sample age ranged from 3.50 to 17.99 years (M = 10.51, SD = 4.15) and was approximately evenly distributed across childhood and adolescence (see Table 1). Fifty-one percent of the samples were tested in North American countries, and the majority of the remaining samples were tested in Western European countries (see Table 1).

Table 2
Summary of Procedural Moderators, Including the Mean Weighted Effect Size (g), the Number of Articles (m), the Number of Effect Sizes (k), and the Mean and Range of the Mean Sample Ages (in Years) Corresponding to Each Moderator Level

Procedural moderator	*g* [95% CI of *g*]	*m*	*k*	*M* age (range)
Task characteristics				
Task				
CMTT	.19 [.03, .34]	13	26	5.57 (4.23–7.58)
PMA-SR	.37 [.23, .51]	15	44	11.16 (5.50–17.50)
VMRT	.58 [.47, .69]	42	80	13.64 (7.20–17.99)
Stimulus dimensionality				
Two-dimensional	.30 [.22, .37]	61	152	9.19 (3.96–17.50)
Three-dimensional	.50 [.41, .59]	77	148	11.94 (3.50–17.99)
Stimulus demands				
Mirror discrimination	.44 [.36, .51]	101	239	11.14 (3.50–17.99)
No mirror discrimination	.16 [.06, .26]	22	50	6.86 (4.23–14.00)
Stimulus type				
Abstract	.43 [.36, .51]	106	244	11.21 (3.96–17.99)
Animate	.19 [.10, .29]	22	41	7.35 (3.50–12.00)
Testing conditions				
Presentation method				
Digital	.28 [.10, .46]	25	49	9.16 (4.00–17.60)
Paper	.41 [.34, .49]	99	230	11.27 (4.23–17.99)
Test setting				
Group	.45 [.37, .54]	75	189	12.48 (6.50–17.99)
Individual	.26 [.17, .35]	54	113	7.22 (3.50–17.50)
Time constraints				
Time limit	.45 [.36, .53]	81	190	12.40 (5.50–17.99)
No time limit	.28 [.19, .38]	49	111	7.28 (3.50–17.00)

Note. CMTT = Children's Mental Transformation Task (Levine, Huttenlocher, Taylor, & Langrock, 1999); PMA-SR = Primary Mental Abilities-Space Relations (Thurstone & Thurstone, 1943); VMRT = Vandenberg-Kuse Mental Rotation Test (Vandenberg & Kuse, 1978).

Measure-level variables. We coded seven procedural variables that were identified as potential moderators of interest given their effect on the male advantage in adulthood and confounding relation with mean sample age. Four of these variables described features of the mental rotation measure and stimuli (i.e., task, stimulus dimensionality, stimulus discrimination demands, and stimulus type), and three of these variables described features of the testing conditions (i.e., presentation method, test setting, time constraints). To maintain adequate power for moderator and subgroup analyses, we determined the levels of each procedural variable based on whether they applied to 10 or more studies included in the meta-analysis. For example, three tasks were administered in 10 or more studies included in the meta-analysis: the Children's Mental Transformation Task (CMTT; Levine et al., 1999), the Primary Mental Abilities-Space Relations (PMA-SR) subtest (Thurstone & Thurstone, 1943), and the Vandenberg-Kuse Mental Rotation Test (VMRT; Vandenberg & Kuse, 1978). Therefore, when conducting analyses that included task as a predictor variable, we omitted effect sizes produced using tasks other than the CMTT, PMA-SR, or VMRT. See Table 2 for the levels of each procedural moderator and Appendix B for the coding procedure.

Reliability analyses. The first author assessed the inclusion eligibility of all articles retained for full-text review and completed the data coding and effect size extraction for all articles included in the meta-analysis. A second author independently reviewed one third of the records retained for full-text review, assessing their inclusion eligibility and completing data coding and effect size extraction when appropriate. Using this subset of records, intercoder reliability was examined for the following variables: country of testing, effect size, gender centrality, inclusion eligibility, mean sample age, publication status, sample size, stimulus dimensionality, stimulus discrimination demands, stimulus type, task, task presentation method, test setting, time constraints, and year of publication (see Appendix B for descriptions of each variable). There was 97% agreement regarding inclusion eligibility. Intercoder agreement for all other variables was high as measured by Krippendorff's alpha (mean α = .94, range = 0.86–1.00; see Table S1).

Data Analyses

The present study aimed to characterize the developmental trajectory of the male advantage in mental rotation from early childhood to late adolescence and to assess the influence of mean sample age and procedural factors on the magnitude of reported gender differences. To accomplish these objectives, we conducted an initial meta-analysis to quantify the gender difference in mental rotation across all effect sizes extracted from the literature. Then, we examined age-related change in the male advantage by conducting a hierarchical metaregression analysis that included age as a continuous predictor of effect size. Finally, a series of hierarchical metaregression analyses were conducted to assess the potential moderating effects of procedural variables of interest (e.g., task, stimulus dimensionality, stimulus type) and to examine the unique variability in effect sizes attributable to age and to procedural moderators.

In addition to the analyses described above, we conducted exploratory analyses to assess the moderating effects of article-level and sample-level characteristics, such as year of publication and economic development in the country of testing (see Table 1). Because we did not have a priori expectations regarding the influence of these variables on the development of the male advantage, we did not consider them as moderators in the analyses reported in the main text. However, analyses examining these variables as moderators can be found in the Supplemental Material.

Meta-analytic approach. Articles that met our inclusion criteria often reported multiple effect sizes obtained from samples of different ages or via different mental rotation tasks (see Appendix A). The inclusion of multiple effect sizes per article created dependencies within the effect size distribution, violating the assumption of independence required by traditional meta-analytic techniques (Borenstein, Hedges, Higgins, & Rothstein, 2009). To allow for dependencies in the meta-analytic dataset, metaregression analyses were performed using the robust variance estimation (RVE) method developed by Hedges and colleagues (2010a, 2010b). The RVE approach was optimal for use in this meta-analysis because it tolerates the inclusion of both hierarchical and correlated effects (Tipton & Pustejovsky, 2015) and effectively accounts for the resulting inflation in Type I error that occurs in moderator analyses (López-López, Van den Noortgate, Tanner-Smith, Wilson, & Lipsey, 2017). However, a notable caveat to the use of this approach is that, unlike traditional meta-analytic methods, RVE analyses do not support hypothesis testing based on heterogeneity statistics (Tanner-Smith, Tipton, & Polanin, 2016). Thus, heterogeneity statistics are not reported in the Results.

Metaregression analyses were conducted in R using the "robumeta" package (*v*. 2.0; Fisher, Tipton, & Zhipeng, 2017; Tanner-Smith et al., 2016) and adopting a mixed-model approach with inverse variance weights (Tipton & Pustejovsky, 2015). In line with the recommendations of Tanner-Smith and Tipton (2014), we specified hierarchical effects weights for all metaregression analyses as this type of dependency was the most prevalent in our dataset. We assessed the significance of continuous moderators via robust t tests and the significance of categorical moderators via robust F tests on $k - 1$ dummy-coded variables. F tests were conducted using the Wald Test function in the "clubSandwich" R package (*v*. 0.3.0; Pustejovsky, 2017). We required that each level of a categorical variable pertain to at least 10 articles included in the meta-analysis to allow for the accurate estimation of moderating effects (i.e., $df > 4$; Tanner-Smith & Tipton, 2014). Small sample corrections were implemented in all metaregression analyses.

Publication bias. We investigated the effects of publication bias within our sample in a number of ways. First, we considered the role of publication status as a moderator, comparing the magnitude of effect sizes reported in published and unpublished articles. We then searched for evidence of publication bias in our meta-analytic sample by examining the distribution of reported effect sizes via a funnel plot. Furthermore, we conducted Egger's Test of Asymmetry (Egger, Davey Smith, Schneider, & Minder, 1997) and a trim-and-fill analysis (Duval & Tweedie, 2000) to characterize any potential skew in the effect size distribution and to explore its effects on the mean estimated effect size. Finally, we conducted selection model analyses to estimate the mean weighted effect size when adjusting for potential publication bias via weight-function modeling (Vevea & Hedges, 1995; Vevea &

Woods, 2005), per the recommendation of McShane and colleagues (2016).

Prior to conducting publication bias analyses, we aggregated effect sizes at the article level using the "MAd" package in R (*v.* 0.8–2; Del Re & Hoyt, 2014) via the Borenstein et al. (2009) aggregation method to remove dependencies among effect size estimates. We then used mixed-effects models with the article-level aggregated effect sizes for all publication bias analyses. The "metafor" package in R was used to compute Egger's Test and for the trim-and-fill analysis (Viechtbauer, 2010). Selection model analyses were implemented using the "weightr" package in R (*v.* 1.1.2; Coburn & Vevea, 2017).

Assessment of outliers. Four effect sizes were identified as outliers (i.e., greater than 2.5 *SD*s $\pm$ *M*; see Appendix A). Removal of these outliers had no effect on the mean weighted effect size or on the significance of any analyses except when noted in the Results.

Results

Across all articles included in the meta-analysis ($m = 128$, $k = 303$), the mean weighted effect size equaled 0.39 (95% CI [0.33, 0.46], $p < 1.00 \times 10^{-15}$), indicating a small to moderate male advantage in mental rotation performance during childhood and adolescence. As expected, there was substantial heterogeneity among effect sizes that could be attributed to both between-study variability ($\tau^2 = 0.06$) and within-study variability ($\omega^2 = 0.02$). To determine whether age and procedural moderators contributed to this heterogeneity, we next conducted a series of hierarchical metaregression analyses that assessed their unique effects on reported effect sizes.

Developmental Change in the Male Advantage

Mean sample age was positively related to the magnitude of the male advantage in mental rotation performance ($B = 0.04$, $SE = 0.01$, $p = 2.64 \times 10^{-8}$; see Figure 3). Metaregression results suggested that the gender difference reaches a small effect size ($g = 0.20$) by approximately 6 years of age and a moderate effect size ($g = 0.50$) at approximately 14 years of age (Cohen, 1988).

Moderating Effects of Task and Stimulus Characteristics

Task. Reported effect sizes differed by task ($F = 11.41$, $df = 15.43$, $p = .0009$). As in earlier meta-analyses (Linn & Petersen, 1985; Voyer et al., 1995), effect sizes obtained using the VMRT were significantly larger than effect sizes obtained using the PMA-SR, with the VMRT producing a moderate male advantage in performance ($g = 0.58$, 95% CI [0.47, 0.69], $p = 1.46 \times 10^{-10}$, $m = 42$, $k = 80$) and the PMA-SR producing a small male advantage ($g = 0.37$, 95% CI [0.23, 0.51], $p = .0002$, $m = 15$, $k = 44$). In addition, the CMTT was found to produce a very small, but significant, gender difference in performance ($g = 0.19$, 95% CI [0.03, 0.34], $p = .023$, $m = 13$, $k = 26$).[1]

Because mean sample age was confounded with task (see Table 2), we next considered whether the observed age-related change in the male advantage resulted from the differential use of the CMTT, PMA-SR, and VMRT across development. When controlling for

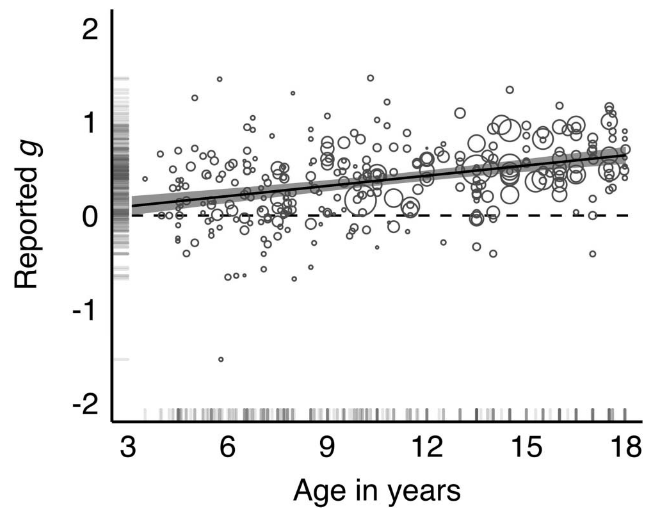

Figure 3. Bubble plot depicting the relation between mean sample age and reported gender differences in mental rotation performance with marginal rug plots that illustrate the distributions of age and effect size. Effect sizes greater than 0 indicate a male advantage in performance, whereas effect sizes less than 0 indicate a female advantage. Marker sizes are proportional to the meta-analytic weight of the corresponding effect size. The solid line denotes the metaregression line for a model with age as the only predictor; the shaded area represents its 95% confidence interval.

the effect of task via metaregression, age remained a significant predictor of reported gender differences ($F = 13.96$, $df = 19.34$, $p = .001$), with the male advantage reaching a small effect size of 0.20 by 7 years of age and a moderate effect size of 0.50 at 16 years of age. Furthermore, the effect of age did not vary statistically by task; $F = 1.46$, $df = 12.66$, $p = .270$ (see Table S2 for effect size estimates by age and task). Thus, the observed age-related change in the magnitude of the male advantage across childhood and adolescence could not be attributed solely to a confounding relation between age and task within our sample.

Given the robust effect of age on reported effect sizes, we next considered whether age-related change in the tasks used to assess mental rotation skills resulted in apparent differences across tasks (see Table 2). Task remained a significant predictor of reported effect sizes when controlling for mean sample age ($F = 4.97$, $df = 16.80$, $p = .020$), indicating that task accounted for unique variability in observed gender differences in mental rotation performance. Follow-up contrasts revealed that this unique effect of task was driven by differences between the VMRT and the other two mental rotation measures. Specifically, the VMRT produced larger effect sizes than the PMA-SR when controlling for mean sample age ($F = 9.47$, $df = 11.95$, $p = .010$), whereas effect sizes produced by the PMA-SR and CMTT did not differ ($F = 0.02$, $df = 13.86$, $p = .903$). Taken together, these findings establish unique roles for age and task as moderators of the gender difference in mental rotation performance across development.

Stimulus dimensionality. As found in adulthood (e.g., Peters et al., 1995), effect sizes derived from tasks containing 3D stimuli

[1] When outliers were removed from the dataset, the mean weighted effect size for the CMTT equaled 0.16, 95% CI [−0.01, 0.33], $p = .059$, $m = 13$, $k = 25$.

($g = 0.50$, 95% CI [0.41, 0.59], $p = 3.11 \times 10^{-14}$, $m = 77$, $k = 148$) were significantly greater than effect sizes derived from tasks containing 2D stimuli ($g = 0.30$, 95% CI [0.22, 0.37], $p = 5.28 \times 10^{-9}$, $m = 61$, $k = 152$); $F = 18.95$, $df = 52.72$, $p = 6.21 \times 10^{-5}$. When both age and stimulus dimensionality were included as predictors in the same metaregression model, age ($F = 30.71$, $df = 36.08$, $p = 2.83 \times 10^{-6}$) and stimulus dimensionality ($F = 9.06$, $df = 33.95$, $p = .005$) each accounted for unique variability in reported effect sizes. Thus, tasks requiring the mental rotation of 3D stimuli produced larger gender differences than tasks requiring the mental rotation of 2D stimuli regardless of the age at which mental rotation was assessed, and the magnitude of the male advantage increased with age regardless of the dimensionality of the stimuli used to measure performance.

To further consider the influence of the confounding relation between stimulus dimensionality and age on our results, we next examined developmental change as a function of stimulus dimensionality. Metaregression analyses indicated that the effect of age on effect sizes obtained via 3D stimuli ($B = 0.04$, $SE = 0.01$, $p = 2.42 \times 10^{-5}$) may have been greater than the effect of age on effect sizes obtained via 2D stimuli ($B = 0.02$, $SE = 0.01$, $p = .004$); $F = 3.17$, $df = 34.99$, $p = .084$. The male advantage was estimated to reach a small effect size of 0.20 by 6 years of age regardless of stimulus dimensionality; however, the gender difference reached a moderate effect size of 0.50 at approximately 13 years of age when 3D stimuli were employed to assess mental rotation abilities, but not until after 17 years of age when 2D stimuli were employed. Taken together, these findings demonstrate that both age and stimulus dimensionality account for unique variability in the gender difference in mental rotation across development and suggest that the effect of age may be more pronounced when 3D stimuli are used to assess performance.

Stimulus discrimination demands. Effect sizes obtained via tasks that required mirror discrimination ($g = 0.44$, 95% CI [0.36, 0.51], $p = 1.60 \times 10^{-14}$, $m = 101$, $k = 239$) were significantly larger than those obtained via tasks that did not require mirror discrimination ($g = 0.16$, 95% CI [0.06, 0.26], $p = .006$, $m = 22$, $k = 50$); $F = 21.41$, $df = 19.30$, $p = .0002$. When both stimulus discrimination requirements and age were included as predictors in a single metaregression model, age was significantly associated with reported effect size ($F = 23.74$, $df = 43.02$, $p = 1.53 \times 10^{-5}$); the effect of stimulus discrimination requirements was trending toward significance ($F = 3.79$, $df = 23.28$, $p = .064$) and was statistically significant when effect size outliers were removed from the dataset ($F = 4.58$, $df = 22.86$, $p = .043$). Follow-up analyses indicated that the effect of age on effect sizes derived from tasks that required mirror discrimination ($B = 0.03$, $SE = 0.01$, $p = 1.10 \times 10^{-5}$) did not differ from the effect of age on effect sizes derived from tasks that did not require mirror discrimination ($B = 0.01$, $SE = 0.03$, $p = .843$); $F = 0.63$, $df = 3.71$, $p = .475$. These results suggest that age and stimulus discrimination demands may both influence the magnitude of the male advantage across childhood and adolescence.

Stimulus type. Abstract stimuli ($g = 0.43$, 95% CI [0.36, 0.51], $p = 4.44 \times 10^{-16}$, $m = 106$, $k = 244$) were associated with larger effect sizes than animate stimuli ($g = 0.19$, 95% CI [0.10, 0.29], $p = .0005$, $m = 22$, $k = 41$); $F = 18.46$, $df = 19.27$, $p = .0004$. When both mean sample age and stimulus type were included as predictors in the same metaregression model, age remained a significant predictor of reported effect sizes ($F = 32.54$, $df = 43.69$, $p = 9.35 \times 10^{-7}$), whereas stimulus type was no longer statistically significant ($F = 3.94$, $df = 21.24$, $p = .060$). Follow-up analyses revealed that the effect of age did not differ by stimulus type ($F = 0.07$, $df = 10.57$, $p = .792$). Taken together, these results demonstrate that the reported effect of age could not be attributed to the greater use of animate stimuli among younger samples and leave open the possibility that abstract stimuli may produce larger gender differences in mental rotation performance relative to animate stimuli during childhood.

Unique effects of stimulus characteristics. The above results suggest that all of the stimulus characteristics examined as potential moderators of the gender difference in mental rotation may have accounted for unique variation in reported effect sizes when controlling for the effect of age. However, it remains possible that confounds among these stimulus characteristics influenced our results. To address this possibility, we assessed the unique roles of the three stimulus characteristics in moderating the male advantage by including age, stimulus dimensionality, stimulus discrimination demands, and stimulus type as predictors in a single metaregression model. Although task was also found to be a significant predictor of reported effect sizes when controlling for age, task could not be included in the same metaregression model as the three stimulus characteristics because of one-to-one correspondences between the levels of task and of the three stimulus characteristics (e.g., all three of the tasks examined in the meta-analysis consisted of abstract stimuli).

We found that stimulus dimensionality ($F = 4.62$, $df = 28.79$, $p = .040$), stimulus discrimination demands ($F = 6.49$, $df = 23.66$, $p = .018$), and stimulus type ($F = 9.76$, $df = 23.69$, $p = .005$) were all significantly related to reported effect sizes, demonstrating that multiple stimulus characteristics contribute unique variability to reported effect sizes across development. Moreover, age remained a significant moderator of the male advantage ($F = 6.14$, $df = 34.02$, $p = .018$), with metaregression results suggesting that the gender difference in mental rotation performance was expected to reach a small effect size of 0.20 by approximately 8 years of age and a moderate effect size of 0.50 in early adulthood when controlling for the effects of stimulus characteristics. Thus, developmental change in the magnitude of the male advantage in mental rotation performance could not be attributed solely to the influence of procedural confounds.

Given the unique effects of age, stimulus dimensionality, stimulus discrimination demands, and stimulus type documented above, we next examined whether age-related change in the male advantage varied as a function of the stimuli used in evaluating performance. When interaction terms between age and each of the stimulus characteristics were added to the metaregression model, the effect of age was greater for effect sizes derived from 3D stimuli than those derived from 2D stimuli ($F = 11.65$, $df = 29.66$, $p = .002$), but did not vary by stimulus discrimination demands or stimulus type ($p\text{s} > .31$). Taken together, these results demonstrate that both age and stimulus characteristics influence the magnitude of reported gender differences and further suggest that the rate of developmental change in the male advantage may be influenced by procedural factors.

Relations Between Testing Conditions and Reported Gender Differences

Presentation method. Effect sizes derived from measures presented on paper ($g = 0.41$, 95% CI [0.34, 0.49], $p = 5.68 \times 10^{-14}$, $m = 99$, $k = 230$) did not differ from those derived from computerized mental rotation tasks ($g = 0.28$, 95% CI [0.10, 0.46], $p = .005$, $m = 25$, $k = 49$); $F = 2.01$, $df = 19.70$, $p = .172$. When both age and presentation method were included as predictors in the same metaregression model, the effect of presentation method remained nonsignificant ($F = 0.15$, $df = 18.78$, $p = .705$), whereas age continued to be significantly associated with reported effect sizes ($F = 46.86$, $df = 42.55$, $p = 2.27 \times 10^{-8}$). Furthermore, the effect of age did not differ for effect sizes derived from paper measures ($B = 0.04$, $SE = 0.01$, $p = 3.10 \times 10^{-7}$) and those derived from digital measures ($B = 0.05$, $SE = 0.02$, $p = .039$), $F = 0.21$, $df = 7.62$, $p = .661$. These findings do not provide evidence of a moderating role of presentation method in reported gender differences during development, at least when comparing effect sizes derived from computerized measures to those derived from paper tasks.

Test setting. Metaregression revealed a significant effect of test setting ($F = 9.88$, $df = 56.84$, $p = .003$), such that reported gender differences were larger when tasks were administered in a group setting ($g = 0.45$, 95% CI [0.37, 0.54], $p = 4.64 \times 10^{-12}$, $m = 75$, $k = 189$) than when tasks were administered to participants individually ($g = 0.26$, 95% CI [0.17, 0.35], $p = 2.38 \times 10^{-6}$, $m = 54$, $k = 113$). When both age and test setting were included as predictors in a single metaregression model, the effect of test setting was no longer significant ($F = 0.02$, $df = 38.38$, $p = .891$), whereas age continued to account for unique variance in reported effect sizes ($F = 25.75$, $df = 35.06$, $p = 1.28 \times 10^{-5}$). In addition, the effect of age on effect sizes collected in group settings ($B = 0.04$, $SE = 0.01$, $p = 1.91 \times 10^{-5}$) did not differ from the effect of age on effect sizes obtained through individual testing ($B = 0.02$, $SE = 0.01$, $p = .205$), $F = 1.95$, $df = 19.64$, $p = .178$. Thus, test setting did not appear to uniquely influence the magnitude of reported gender differences.

Time constraints. As found in previous meta-analyses (Maeda & Yoon, 2013; Voyer, 2011), tasks that were administered with time limits ($g = 0.45$, 95% CI [0.36, 0.53], $p = 1.06 \times 10^{-12}$, $m = 81$, $k = 190$) produced significantly greater effect sizes than tasks that were administered without time limits ($g = 0.28$, 95% CI [0.19, 0.38], $p = 1.91 \times 10^{-6}$, $m = 49$, $k = 111$); $F = 6.81$, $df = 49.50$, $p = .012$. However, when both time constraints and age were included as predictors in the same metaregression model, time constraints no longer related to reported effect sizes ($F = 0.51$, $df = 33.55$, $p = .481$), whereas age remained significantly associated with gender differences in performance ($F = 37.58$, $df = 38.79$, $p = 3.49 \times 10^{-7}$). Moreover, the effect of age on effect sizes that were derived from tasks with time limits ($B = 0.04$, $SE = 0.01$, $p = 1.02 \times 10^{-5}$) did not differ from the effect of age on effect sizes derived from tasks without time limits ($B = 0.04$, $SE = 0.01$, $p = .031$), $F = 0.01$, $df = 14.67$, $p = .945$. Thus, the male advantage in mental rotation performance did not depend on the presence of time constraints when controlling for the effects of age.

Publication Bias

Published articles reported larger gender differences in mental rotation performance ($g = 0.42$, 95% CI [0.34, 0.50], $p = 4.29 \times 10^{-14}$, $m = 103$, $k = 245$) in comparison with unpublished articles ($g = 0.28$, 95% CI [0.17, 0.39], $p = .0002$, $m = 25$, $k = 58$; $F = 5.17$, $df = 15.54$, $p = .038$),[2] signaling the presence of publication bias within the mental rotation literature. To evaluate the effect of this bias on our results, we constructed a funnel plot displaying the distribution of article-level aggregated effect sizes for all studies included in the current meta-analysis ($m = 128$). As shown in Figure 4, effect sizes were largely centered symmetrically around the weighted mean effect size, and there was no significant asymmetry in the effect size distribution (Egger's $z = -1.09$, $p = .274$). Moreover, a trim-and-fill analysis (Duval & Tweedie, 2000) did not impute any additional effect sizes below the mean estimated effect size. These results are not consistent with the presence of publication bias in our effect size distribution. However, given the significant moderating effect of publication bias, we sought to quantify the potential effects of publication bias on our effect size estimates via selection model analyses, as recommended by McShane and colleagues (2016).

We first implemented the Vevea and Hedges (1995) weight-function model to estimate the population effect size after adjusting for selection bias. When specifying p value cutpoints of 0.01, 0.05, and 0.10, the mean weighted effect size estimated by the unadjusted model ($g = 0.40$) did not differ from the mean weighted effect size estimated by the adjusted model ($g = 0.39$); $\chi^2(3) = 1.14$, $p = .766$). We then conducted Vevea and Woods' (2005) selection model analyses that provide estimates of the mean weighted effect size assuming varying degrees of selection bias. These analyses produced adjusted g values that, although lower than unadjusted effect size of 0.40, were all small in magnitude (i.e., 0.35 when assuming moderate one-tailed selection, 0.37 when assuming moderate two-tailed selection, 0.23 when assuming severe one-tailed selection, and 0.33 when assuming severe two-tailed selection). In sum, these results suggest that our effect size estimate may have been slightly inflated because of publication bias in the literature, but do not provide direct evidence that publication bias unduly influenced the outcomes of our meta-analysis. See the Supplemental Material for additional analyses examining publication bias.

Discussion

Previous meta-analyses have established the presence of a robust male advantage in mental rotation performance during adulthood (Linn & Petersen, 1985; Maeda & Yoon, 2013; Voyer et al., 1995), but produced inconsistent evidence of age-related change in its magnitude. In the present research, we meta-analyzed findings of 128 articles that evaluated mental rotation performance in more than 30,000 participants aged 3 to 17 years. The main objectives of

[2] The moderating effect of publication status was no longer statistically significant when outliers were removed from the analysis: $F = 4.12$, $df = 15.91$, $p = .059$, $m = 128$, $k = 299$.

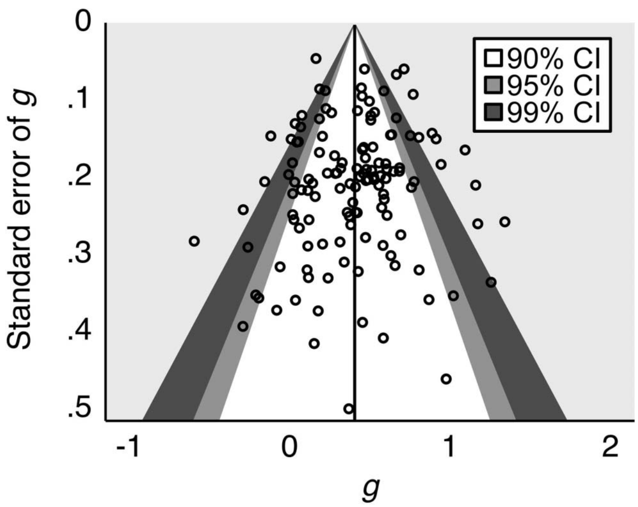

Figure 4. Funnel plot displaying article-level aggregated effect sizes by their standard error. The solid line denotes the mean weighted effect size, and the shaded regions correspond to its 90%, 95%, and 99% confidence intervals. See the Supplemental Material for funnel plots by age group.

this meta-analytic review were to characterize the developmental trajectory of the male advantage in mental rotation performance between early childhood and late adolescence and to elucidate procedural factors that influence its magnitude across development. In the following sections, we consider the findings related to each of these objectives, situate our results within theoretical accounts of the gender difference, and offer recommendations for future research.

Emergence and Age-Related Change in the Male Advantage Across Development

The results of the present meta-analysis demonstrate considerable age-related change in the magnitude of the male advantage in mental rotation performance across childhood and adolescence. More specifically, our findings suggest that a small gender difference in children's performance on mental rotation tasks emerges during the first years of formal schooling, and this small gender difference subsequently increases with age, tripling in size by late adolescence. Importantly, this age-related change was shown to be robust to procedural confounds as the effect of age remained significant even when controlling for the effects of influential procedural moderators, such as task and stimulus characteristics. What could account for the extensive developmental change observed? Prominent theoretical perspectives on the gender difference in mental rotation posit that gender-linked environmental, psychosocial, and hormonal influences give rise to the male advantage in performance (for review, see Levine, Foley, Lourenco, Ehrlich, & Ratliff, 2016). Despite empirical efforts to evaluate the validity of these varying perspectives, the mechanisms underlying the male advantage and its emergence in ontogeny have remained elusive. The results of the present meta-analysis, although not intended to adjudicate between causal models, indicate that the optimal theoretical account of the gender difference in mental rotation performance must account for its developmental emergence during the elementary-school years as well as age-related gains in its magnitude across middle childhood and adolescence. Therefore, we next consider our findings within the context of predominating perspectives on the origins of gender differences in spatial reasoning.

One popular account of gender differences in spatial cognition emphasizes the importance of early experiential factors in the development of the male advantage, proposing that men outperform women because of their relatively greater experience with activities that promote spatial learning. In support of this contention, prior research has shown that greater experience with spatial activities is associated with enhanced spatial reasoning abilities (Baenninger & Newcombe, 1989; Levine et al., 2012), and training studies suggest that this relation may be causal (for meta-analysis, see Uttal et al., 2013). Importantly, gender differences in children's experience with spatial activities, including play with spatially relevant toys and videogames, are significant by preschool age (Jirout & Newcombe, 2015; Lauer, Ilksoy, & Lourenco, 2018; Serbin et al., 1990) and increase with age (Baenninger & Newcombe, 1989; Cherney & London, 2006). From this viewpoint, one would predict that the gender difference in mental rotation would develop in synchrony with the gender difference in children's exposure to spatially relevant activities, as found in the present meta-analysis. Furthermore, one would expect that the increasing gender disparity in children's experience with spatial activities across development would result in a concomitant increase in the male advantage in mental rotation performance, a conjecture supported by our findings of developmental change between early childhood and adolescence.

Research investigating experiential influences on the gender difference in mental rotation has principally focused on spatial activities. However, recent studies have documented positive effects of parental spatial language use on children's spatial thinking (e.g., Polinsky, Perez, Grehl, & McCrink, 2017; Pruden et al., 2011; Verdine et al., 2014) as well as gender differences in exposure to parental spatial language (Pruden & Levine, 2017), implicating an additional gender-linked environmental factor in spatial development. As with the scaffolding effects of spatial activities, parental use of spatial language (i.e., references to shapes, spatial properties, and dimensional characteristics of objects) may orient children's attention to the relevant information necessary to solve spatial problems and support their understanding of spatial concepts. If boys begin to accrue relatively greater exposure to spatial language within the first years of life, as suggested by the findings of Pruden and Levine (2017), then environmental factors extending beyond direct experience with spatial activities may contribute to the increasing gender disparity in mental rotation performance over the course of development. However, spatial play has been shown to elicit greater spatial language use during parent–child interactions (Ferrara, Hirsh-Pasek, Newcombe, Golinkoff, & Lam, 2011), meaning that children who often engage in spatial activities may receive greater exposure to parental spatial language. Thus, it remains unclear to what extent these different gender-linked experiences exert dissociable effects on spatial development, calling for future research that examines the relative contributions of spatial activities, spatial language exposure, and other environmental factors to the male advantage in spatial performance during childhood.

Our findings may also be interpreted as support for perspectives implicating psychosocial factors in the development of gender differences in spatial reasoning. Most notably, it has been posited that societal stereotypes regarding the relative spatial intelligence of women and men contribute to the male advantage in spatial performance. For example, some previous studies have reported that increasing the salience of gender stereotypes concerning spatial aptitude has deleterious effects on girls' mental rotation performance as early as middle childhood (Neuburger, Ruthsatz, Jansen, & Quaiser-Pohl, 2015; Titze, Jansen, & Heil, 2010a), although meta-analytic findings raise doubts about the reliability of stereotype-threat effects in adulthood (Doyle & Voyer, 2016). In addition to potential transient stereotype-threat effects, emerging awareness of gender stereotypes regarding spatial intelligence during elementary school (Moè, 2018; Neuburger et al., 2015) may exacerbate the gender difference in mental rotation performance through others means, such as discouraging girls from engaging in male-typed spatial activities that foster spatial development and leading caregivers to provide boys with relatively more opportunities to develop and utilize their spatial skills. Given that the multifarious effects of gender stereotypes likely accumulate with age, their rising influence on children's mental rotation performance may result in the developmental trends documented in the present meta-analysis. Nevertheless, in comparison with other factors associated with gender differences in spatial abilities, we know relatively little about the effects of gender stereotypes on spatial aptitude. Therefore, further research on this topic will be necessary to evaluate the validity of this theoretical account.

In addition to gender stereotypes, there is growing evidence that other psychosocial influences, including gender-related affective factors, contribute to the male advantage in mental rotation performance (e.g., Estes & Felker, 2012; Moè & Pazzaglia, 2010; Ramirez, Gunderson, Levine, & Beilock, 2012). For example, spatial anxiety, which refers to anxiety induced by tasks that require spatial reasoning, is a gender-linked trait that has been associated with poorer mental rotation performance in childhood (Cardillo, Vincenzi, & Gallani, 2017; Lauer et al., 2018) and adulthood (Lawton, 1994; Lyons et al., 2018). Studies on school-age children have found that girls express significantly more spatial anxiety than boys within the first years of formal schooling (ds $>$ .3), even when girls and boys exhibit similar performance on mental rotation tasks (Lauer et al., 2018; Ramirez et al., 2012). These results suggest that early gender differences in spatial anxiety may play a role in the later development of gender differences in spatial performance. Furthermore, it is likely that the relation between spatial anxiety and spatial ability exhibits reciprocity over time, meaning that the detrimental effects of spatial anxiety on girls' mental rotation performance would be expected to compound across development, a prediction that aligns with our findings of age-related change in the magnitude of the male advantage. Thus, evidence of early emerging gender differences in spatial anxiety and the potential bidirectional relation between spatial anxiety and mental rotation performance could account for the developmental trajectory of the male advantage observed in the present meta-analysis. More research examining the role of gender differences in spatial anxiety and their effects on spatial performance longitudinally will be needed to directly test questions about the origins and consequences of gender differences in children's spatial anxiety over the course of development.

A final theoretical perspective on the origins of the male advantage posits that the effects of sex hormones on brain development and function give rise to the gender difference in mental rotation performance at varying points in development (e.g., Constantinescu, Moore, Johnson, & Hines, 2018; Hooven, Chabris, Ellison, & Kosslyn, 2004; Silverman & Phillips, 1993). For instance, it has been proposed that the surge in sex hormones during puberty is influential for gender-linked spatial skills (e.g., Davison & Susman, 2001; Hier & Crowley, 1982). We found that a small but significant male advantage in mental rotation is present by age 6, refuting the possibility that pubertal onset drives the emergence of the gender difference in mental rotation. However, puberty spans much of late childhood and adolescence (Dorn, 2006), a period of development that we found to coincide with substantial gains in the magnitude of the male advantage. Thus, our results leave open the possibility that the effects of sex hormones concomitant with puberty magnify the gender difference in mental rotation performance following pubertal onset. Such sex hormones could contribute to gender differences in mental rotation performance by exerting activational effects on the neural bases of visuospatial reasoning, although previous reviews on this topic cast doubt on the association between circulating sex hormones and spatial performance (for meta-analysis, see Puts et al., 2010; for review, see Miller & Halpern, 2014). Alternatively, pubertal hormones could have organizational effects on neurodevelopment that endure beyond puberty, differentially impacting the mental rotation performance of women and men throughout adolescence and adulthood. Further research on both the organizational and activational effects of pubertal hormones on spatial processing will be essential to adjudicate among these possibilities.

Another prominent hypothesis regarding the role of sex hormones in the male advantage contends that early androgens have organizational effects on the nervous system pre- and postnatally, influencing the development of the neural substrates of visuospatial processing in female and male infants. In particular, it has been proposed that boys' relatively greater exposure to androgens in utero and during the first six months of life has enduring effects on neurodevelopment that result in gender-related variation in spatial performance across the life span. The argument that organizational effects of early androgens account for later gender differences in mental rotation performance furnishes the prediction that the male advantage in mental rotation will emerge in the first years of life and continue to be present throughout childhood. This expectation is not directly congruent with the findings of our meta-analysis, as a small male advantage in mental rotation performance was not present until the elementary school years. Thus, our results do not support the proposal that early androgens produce an early emerging and developmentally continuous gender difference in mental rotation performance. Nevertheless, it remains possible that organizational effects of early androgens only become detectable following a sufficient period of maturation, and additional primary studies evaluating this hypothesis will be needed to conclude whether early androgen exposure relates to later mental rotation skills.

As described above, the results of our meta-analysis suggest that the male advantage in mental rotation is not present during the preschool years; however, several factors should be considered when interpreting our findings relating to early childhood. First, we required that articles measured response accuracy on a behavioral mental rotation task to be included in the meta-analysis. This inclusion criterion prevented us from examining gender differences in mental rotation performance in children younger than 3 years of age, as young children's mental rotation skills are typically assessed via implicit looking-time measures. A handful of studies that have employed looking-time tasks to evaluate early mental rotation processes have reported moderate to large gender differences in infants' looking behavior (e.g., Lauer, Udelson, Jeon, & Lourenco, 2015; Moore & Johnson, 2008; Quinn & Liben, 2008), leaving open the possibility that the male advantage in mental rotation may be detectable early in development when implicit measures are used to assess performance. However, most studies examining infants' looking behaviors during mental rotation tasks have not reported significant gender differences (e.g., Christodoulou, Johnson, Moore, & Moore, 2016; Erdmann, Kavšek, & Heil, 2018; Möhring & Frick, 2013; Schwarzer, Freitag, & Schum, 2013). Therefore, further investigation of infants' mental rotation abilities will be necessary to determine whether gender differences in implicit mental rotation performance are indeed present in the first year of life and, if so, whether these gender differences represent the origins of the later male advantage in explicit mental rotation performance.

Another factor relevant to our findings on the emergence of the male advantage during childhood is the considerable growth in children's mental rotation abilities during the preschool years. Children's mental rotation skills increase dramatically between the ages of 3 and 6, with young children often experiencing difficulty with mental rotation tasks prior to age 5 (for review, see Frick, Möhring, & Newcombe, 2014). It has been suggested that gender differences are not commonly found during early childhood because measures used to assess mental rotation at this age are unreliable, as young children often perform at chance levels on mental rotation tasks. However, our findings suggest that the male advantage in mental rotation remains small in magnitude throughout childhood, well beyond the age at which most children reliably exhibit above-chance performance. Thus, our findings warrant the conclusion that the male advantage is only small in magnitude during the elementary school years and increases substantially in size with age.

To conclude, our findings of developmental change are particularly compatible with theoretical perspectives that emphasize the contributions of experiential and psychosocial factors as these accounts predict the emergence of the male advantage during childhood as well as an escalation in its magnitude across development. Nonetheless, support for various environmental and psychosocial influences does not preclude hormonal underpinnings of the male advantage, and experiential, psychosocial, and biological factors likely interact to influence mental rotation performance across development. Consequently, future research exploring the relations between these factors will be critical to understanding how they act in tandem to produce, amplify, and sustain the male advantage throughout childhood and adolescence.

Procedural Moderators of the Gender Difference

In addition to age-related change, our results demonstrate that several procedural factors influence the magnitude of the gender difference in mental rotation performance across development. Specifically, moderator analyses revealed that the magnitude of reported gender differences varied across task and stimulus characteristics, even when controlling for age and other methodological factors. These procedural moderators provide insight into the cognitive mechanisms that underlie between-gender variability in mental rotation skills, furthering our understanding of potential pathways through which differential cognitive processes may yield a male advantage in performance.

First, we found evidence of a robust task effect such that the Vandenberg and Kuse (1978) MRT produced larger gender differences in mental rotation performance relative to the PMA-SR (Thurstone & Thurstone, 1943) and the CMTT (Levine et al., 1999) when controlling for mean sample age. Prior meta-analyses also reported a significant effect of task on the magnitude of reported gender differences (Linn & Petersen, 1985; Voyer et al., 1995), but did not address the source of this task-related variability. Our results suggest that differences in stimulus characteristics may underlie the task differences observed in our meta-analysis as well as in previous reviews. In particular, we found that mental rotation tasks composed of 3D stimuli, such as the VMRT, produced larger effect sizes than those composed of 2D stimuli, such as the CMTT and PMA-SR. Furthermore, tasks that required participants to discriminate between rotated mirrored images, including the VMRT, produced larger effect sizes than tasks that required participants to discriminate between rotated objects that were structurally unique, such as the CMTT. Thus, our findings indicate that task-related variation in the magnitude of the male advantage across meta-analyses may reflect differences in the dimensionality and discrimination demands of their stimuli.

One potential explanation for the observed effects of procedural factors is that gender-linked cognitive strategies for solving mental rotation tasks vary in their efficacy across these dimensions. A prominent hypothesis in the cognitive literature asserts that the gender difference in mental rotation results from a gender difference in the cognitive strategies utilized to solve mental rotation problems (Geiser et al., 2006; Geiser et al., 2008; Pezaris & Casey, 1991). Specifically, males may rely more on spatial strategies that involve visualizing the rotation of objects or object parts, whereas females may rely more on analytical strategies that involve comparing different stimulus features (e.g., size, shape, and color of component parts and relations among them) to determine if they are matched across stimuli (Just & Carpenter, 1985; Pezaris & Casey, 1991). Importantly, analytical strategies can be effective in solving some mental rotation tasks, but are more prone to error on others. Examination of the three tasks included in our analyses suggests that these tasks are differentially amenable to the use of analytical strategies. In both the CMTT (Figure 1A) and the PMA-SR (Figure 1B), the shape of incorrect answer choices often differ greatly from that of correct answer choices, meaning that task items may be effectively solved via analytical feature-based comparisons. In contrast, the VMRT (Figure 1C) requires discriminating between rotated figures that possess nearly identical features (i.e., objects and their rotated mirrored images), diminishing the efficacy of comparing these features across stimuli. Therefore, the VMRT may have produced larger gender differences than the CMTT and PMA-SR because it is less amenable to the use of analytical strategies that are more often employed by female participants. In support of this proposal, we found a broader effect of stimulus discrimination demands across all effect sizes included in the meta-analysis. That is, tasks that required participants to discriminate between rotated mirrored stimuli produced larger gender differences in performance relative to tasks that did not require mirror discrimination, even when controlling for age and other stimulus characteristics. Although the origins of the gender differences in strategy use remain unknown, these results support the contention that differential strategy use represents one source of between-gender variation in mental rotation abilities across development.

Our results also revealed that stimulus dimensionality is influential in moderating the gender difference in mental rotation, even when controlling for age and other stimulus characteristics. Tasks consisting of 3D stimuli, including the VMRT, produced larger gender differences in mental rotation performance in comparison to tasks consisting of 2D stimuli, including the CMTT and PMA-SR, indicating that task-related variability in the magnitude of the male advantage resulted in part from the differential use of 2D and 3D stimuli across tasks. Why does stimulus dimensionality influence the role of gender in mental rotation performance? One possibility is that tasks containing 3D stimuli place greater demands on visuospatial working memory (VSWM) resources than 2D tasks. Because males possess slightly better VSWM than females ($d = 0.16$, for meta-analysis, see Voyer, Voyer, & Saint-Aubin, 2017), gender differences in VSWM may account for the observed male advantage in mental rotation performance, a contention supported by findings that VSWM mediates the relation between gender and mental rotation performance in adulthood (Kaufman, 2007). On this view, larger gender differences in mental rotation performance would be expected on tasks that place greater demands on VSWM. Furthermore, given evidence that the male advantage in VSWM increases with age (Voyer et al., 2017), one would expect that the male advantage in mental rotation performance would also increase with age, particularly when measured via mental rotation tasks that heavily tax VSWM resources. Consistent with these predictions, we found that 3D tasks not only produced larger gender differences in performance but also resulted in greater age-related gains in the magnitude of the male advantage. An important caveat to this account, however, is that the gender difference in VSWM appears to be small in size in relation to the observed male advantage in mental rotation performance, particularly during childhood and adolescence (see Voyer et al., 2017). Thus, more research is warranted to test whether VSWM is a central contributing factor in the male advantage in mental rotation during development.

Lastly, moderator analyses indicated that stimulus type accounts for unique variability in reported effect sizes beyond the effects of stimulus dimensionality, stimulus discrimination demands, and age. More specifically, animate stimuli produced smaller gender differences in mental rotation performance in comparison to abstract stimuli, similar to the findings of a previous study on adults (Alexander & Evardone, 2008). A potential explanation for this stimulus type effect is that children are more familiar with animate stimuli than with abstract stimuli, and this greater familiarity allows girls to more readily form holistic mental representations of animate stimuli and therefore engage in spatial problem-solving strategies that are too challenging with unfamiliar, abstract stimuli. Alternatively, greater familiarity may allow children to maintain mental images of animate stimuli without placing increased demands on their working memory resources, mitigating the effects of boys' relatively greater visuospatial working memory capacities. Because our meta-analysis is the first to examine the effect of stimulus type on gender differences in mental rotation performance, our findings highlight the need for further exploration of the role of stimulus type in producing the male advantage in both children and adults.

Limitations

Several limitations of the current meta-analysis should be considered when interpreting its results. First, the meta-analysis synthesized effect sizes obtained via diverse methods. Although the consideration of effect sizes measured using differing procedures allowed us to characterize the unique effects of age and to document procedural factors that influence the male advantage, the diversity of the methodologies employed within our sample reduced our power to detect more subtle effects of procedural moderators, as many effect sizes were excluded from moderator analyses because of the infrequent use of the corresponding methods in the meta-analytic sample. Moreover, there was skew in the distributions of many procedural moderators, with some procedures being more commonly utilized in the broader literature (see Table 2), as well as substantial confounding between sample age and the procedures used to assess performance (see Table 2). This skew decreased our power to detect procedural moderation when controlling for age. Therefore, future empirical studies testing the effects of procedural variables within the same sample and at different time points in development will be required to fully address the role of the various procedural factors examined in this meta-analysis. In addition, although developmental change was shown to be robust to the effects of all procedural moderators assessed, it is possible that untested confounding factors contributed to the developmental trends we observed. This possibility highlights the importance of further research aimed at disentangling the effects of age and other moderators of the male advantage across childhood and adolescence.

Implications

Spatial skills during childhood are predictive of later academic, professional, and creative success in STEM fields (Kell et al., 2013; Wai et al., 2009; Wai et al., 2010), and recent studies suggest that mental rotation abilities begin to have important implications for STEM achievement within the first years of formal education (e.g., Cheng & Mix, 2014; Gunderson et al., 2012; Lauer & Lourenco, 2016). Central to the present work, gender differences in mental rotation performance account for gender differences in STEM achievement among middle-school and high-school students (Casey et al., 1995; Ganley et al., 2014), indicating that the male advantage in mental rotation contributes to gender disparities in STEM attainment by early adolescence. Although it is evident that many other factors contribute to the gender gap in STEM (Cheryan, Ziegler, Montoya, & Jiang, 2017; Su, Rounds, & Armstrong, 2009; Wang & Degol, 2013), these findings suggest that interventions aimed at reducing the gender difference in mental rotation may help mitigate one gender-related barrier to STEM success.

What types of interventions have the potential to attenuate the gender difference in mental rotation? Meta-analyses have demonstrated that individual differences in spatial abilities are malleable to spatial training (Baenninger & Newcombe, 1989; Uttal et al., 2013), introducing the possibility that increasing girls' experience with spatial activities could enhance their mental rotation performance. Nonetheless, spatial training interventions thus far have been shown to benefit the performance of girls and boys to similar degrees (Uttal et al., 2013), leaving the gender difference in children's performance intact. Alternatively, interventions that specifically target factors implicated in girls' underperformance on mental rotation tasks (e.g., spatial anxiety, strategy use) could be effective in reducing gender disparities during development. Because these factors have been linked to individual differences in mental rotation performance among girls and boys alike, such interventions could not only ameliorate gender differences in children's mental rotation skills but also foster spatial development more broadly.

Conclusion

The present meta-analytic review is the first to synthesize the contemporary developmental literature on gender differences in spatial reasoning, characterizing the developmental trajectory of the male advantage in mental rotation performance from its emergence during childhood to its apex in late adolescence. Our results demonstrate that this age-related change is robust to procedural variation but importantly also indicate that procedural factors influence the magnitude of the gender difference in performance across development. These findings have both theoretical and practical implications, providing support for contributions of environmental, psychosocial, and cognitive factors in the male advantage during childhood and adolescence, and calling for the implementation of educational interventions that promote children's spatial development within the first years of formal schooling.

References

*References preceded by an asterisk correspond to articles included in the meta-analysis.

Alexander, G. M., & Evardone, M. (2008). Blocks and bodies: Sex differences in a novel version of the Mental Rotations Test. *Hormones and Behavior, 53,* 177–184. http://dx.doi.org/10.1016/j.yhbeh.2007.09.014

*Amponsah, B. (2000). A comparison of sex differences in visual-spatial performance from preadolescence to adulthood in Ghana and Norway. *South African Journal of Psychology, 30,* 25–31. http://dx.doi.org/10.1177/008124630003000404

*Anderson, K. L., Casey, M. B., Thompson, W. L., Burrage, M. S., Pezaris, E., & Kosslyn, S. M. (2008). Performance on middle school geometry problems with geometry clues matched to three different cognitive styles. *Mind, Brain and Education, 2,* 188–197. http://dx.doi.org/10.1111/j.1751-228X.2008.00053.x

*Auyeung, B., Knickmeyer, R., Ashwin, E., Taylor, K., Hackett, G., & Baron-Cohen, S. (2012). Effects of fetal testosterone on visuospatial ability. *Archives of Sexual Behavior, 41,* 571–581. http://dx.doi.org/10.1007/s10508-011-9864-8

Baenninger, M., & Newcombe, N. S. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles: A Journal of Research, 20,* 327–344. http://dx.doi.org/10.1007/BF00287729

*Battista, M. T. (1990). Spatial visualization and gender differences in high school geometry. *Journal for Research in Mathematics Education, 21,* 47–60. http://dx.doi.org/10.2307/749456

Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental rotation: Effects of stimulus complexity and familiarity. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 12–23. http://dx.doi.org/10.1037/0096-1523.14.1.12

*Boakes, N. J. (2006). *The effects of origami lessons on students' spatial visualization skills and achievement levels in a seventh-grade mathematics classroom* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3233416)

Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Educator, 2,* 1–17. http://dx.doi.org/10.1007/s00897970138a

Borenstein, M., Hedges, L. V., Higgins, J., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. Chichester, UK: Wiley.

*Boulter, D. R. (1992). *The effects of instruction on spatial ability and geometry performance* (Master's thesis). Available from ProQuest Dissertations and Theses database. (UMI No. MM76356)

Cardillo, R., Vincenzi, I., & Gallani, A. (2017). Spatial tasks and emotional factors: A study conducted with the Italian adaptation of the Child Spatial Anxiety Questionnaire (CSAQ). *Psicologia Clinica dello Sviluppo, 21,* 483–502.

*Carr, M., Steiner, H. H., Kyser, B., & Biddlecomb, B. (2008). A comparison of predictors of early emerging gender differences in mathematics competency. *Learning and Individual Differences, 18,* 61–75. http://dx.doi.org/10.1016/j.lindif.2007.04.005

*Casey, B. M., Andrews, N., Schindler, H., Kersh, J. E., Samper, A., & Copley, J. (2008). The development of spatial skills through interventions involving block building activities. *Cognition and Instruction, 26,* 269–309. http://dx.doi.org/10.1080/07370000802177177

*Casey, M. B., Colón, D., & Goris, Y. (1992). Family handedness as a predictor of mental rotation ability among minority girls in a math-science training program. *Brain and Cognition, 18,* 88–96. http://dx.doi.org/10.1016/0278-2626(92)90113-Z

*Casey, M. B., Nuttall, R. L., & Pezaris, E. (1999). Evidence in support of a model that predicts how biological and environmental factors interact to influence spatial skills. *Developmental Psychology, 35,* 1237–1247. http://dx.doi.org/10.1037/0012-1649.35.5.1237

*Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial-mechanical reasoning skills versus mathematical 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. http://dx.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. http://dx.doi.org/10.1037/0012-1649.31.4.697

Casey, B. M., Pezaris, E., Fineman, B., Pollock, A., Demers, L., & Dearing, E. (2015). A longitudinal analysis of early spatial skills compared to arithmetic and verbal skills as predictors of fifth-grade girls' math reasoning. *Learning and Individual Differences, 40,* 90–100. http://dx.doi.org/10.1016/j.lindif.2015.03.028

*Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15,* 2–11. http://dx.doi.org/10.1080/15248372.2012.725186

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: A Journal of Research, 54,* 717–726. http://dx.doi.org/10.1007/s11199-006-9037-8

Cheryan, S., Ziegler, S. A., Montoya, A. K., & Jiang, L. (2017). Why are some STEM fields more gender balanced than others? *Psychological Bulletin, 143,* 1–35. http://dx.doi.org/10.1037/bul0000052

*Chien, S. (1986). *The effectiveness of animated and interactive microcomputer graphics on children's development of spatial visualization ability/mental rotation skills* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8618765)

*Choi, J. (2000). *Route-learning strategies and spatial processing: A developmental perspective* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. NQ59124)

Christodoulou, J., Johnson, S. P., Moore, D. M., & Moore, D. S. (2016). Seeing double: 5-month-olds' mental rotation of dynamic, 3D block stimuli presented on dual monitors. *Infant Behavior & Development, 45,* 64–70. http://dx.doi.org/10.1016/j.infbeh.2016.09.005

*Clements, D. H., Battista, M. T., Sarama, J., & Swaminathan, S. (1997). Development of students' spatial thinking in a unit on geometric motions and area. *The Elementary School Journal, 98,* 171–186. http://dx.doi.org/10.1086/461890

Coburn, K. M., & Vevea, J. L. (2017). weightr: Estimating weight-function models for publication bias. R Package Version 1.1.2. https://CRAN.R-project.org/package=weightr

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Constantinescu, M., Moore, D. S., Johnson, S. P., & Hines, M. (2018). Early contributions to infants' mental rotation abilities. *Developmental Science, 21,* e12613. http://dx.doi.org/10.1111/desc.12613

*Davison, K. K., & Susman, E. J. (2001). Are hormone levels and cognitive ability related during early adolescence? *International Journal of Behavioral Development, 25,* 416–428. http://dx.doi.org/10.1080/01650250042000438

*Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex-related differences in mathematics. *Intelligence, 32,* 25–32. http://dx.doi.org/10.1016/S0160-2896(03)00061-8

*De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *The Journal of Genetic Psychology: Research and Theory on Human Development, 163,* 272–282. http://dx.doi.org/10.1080/00221320209598683

Del Re, A. C., & Hoyt, W. T. (2014). MAd: Meta-analysis with mean differences. R Package Version 0.8–2. Available at http://cran.r-project.org/web/packages/Mad

Dorn, L. D. (2006). Measuring puberty. *Journal of Adolescent Health, 39,* 625–626. http://dx.doi.org/10.1016/j.jadohealth.2006.05.014

Doyle, R. A., & Voyer, D. (2016). Stereotype manipulation effects on math and spatial test performance: A meta-analysis. *Learning and Individual Differences, 47,* 103–116. http://dx.doi.org/10.1016/j.lindif.2015.12.018

*Dunst, B., Benedek, M., Bergner, S., Athenstaedt, U., & Neubauer, A. C. (2013). Sex differences in neural efficiency: Are they due to the stereotype threat effect? *Personality and Individual Differences, 55,* 744–749. http://dx.doi.org/10.1016/j.paid.2013.06.007

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. http://dx.doi.org/10.1111/j.0006-341X.2000.00455.x

*Dziak, N. J. (1985). *Programming computer graphics and the development of concepts in geometry*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8510566)

Egger, M., Davey Smith, G., Schneider, M., & Minder, C. (1997). Bias in meta-analysis detected by a simple, graphical test. *British Medical Journal, 315,* 629–634. http://dx.doi.org/10.1136/bmj.315.7109.629

*Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology, 42,* 1259–1268. http://dx.doi.org/10.1037/0012-1649.42.6.1259

*Eikenberry, K. D. (1988).*The effects of learning to program in the computer language Logo on the visual skills and cognitive abilities of third-grade children* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8827343)

*Eraso, M. (2007). *Connecting visual and analytic reasoning to improve students' spatial visualization abilities: A constructivist approach* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3298580)

Erdmann, K., Kavšek, M., & Heil, M. (2018). Infants' looking times in a dynamic mental rotation task: Clarifying inconsistent results. *Cognitive Development, 48,* 279–285. http://dx.doi.org/10.1016/j.cogdev.2018.10.002

*Erkoç, M. F., Gecü, Z., & Erkoç, Ç. (2013). The effects of using Google SketchUp on the mental rotation skills of eighth grade students. *Educational Sciences: Theory and Practice, 13,* 1285–1294.

Estes, Z., & Felker, S. (2012). Confidence mediates the sex difference in mental rotation performance. *Archives of Sexual Behavior, 41,* 557–570. http://dx.doi.org/10.1007/s10508-011-9875-5

Ferrara, K., Hirsh-Pasek, K., Newcombe, N. S., Golinkoff, R. M., & Lam, W. S. (2011). Block talk: Spatial language during block play. *Mind, Brain and Education, 5,* 143–151. http://dx.doi.org/10.1111/j.1751-228X.2011.01122.x

Fisher, Z., Tipton, E., & Zhipeng, H. (2017). robumeta: Robust variance meta-regression. R Package Version 2.0. Available at: http://CRAN.R-project.org/package=robumeta

*Frick, A., Daum, M. M., Walser, S., & Mast, F. W. (2009). Motor processes in children's mental rotation. *Journal of Cognition and Development, 10,* 18–40. http://dx.doi.org/10.1080/15248370902966719

*Frick, A., Ferrara, K., & Newcombe, N. S. (2013). Using a touch screen paradigm to assess the development of mental rotation between 3½ and 5½ years of age. *Cognitive Processing, 14,* 117–127. http://dx.doi.org/10.1007/s10339-012-0534-0

*Frick, A., Hansen, M. A., & Newcombe, N. S. (2013). Development of mental rotation in 3- to 5-year-old children. *Cognitive Development, 28,* 386–399. http://dx.doi.org/10.1016/j.cogdev.2013.06.002

Frick, A., Möhring, W., & Newcombe, N. S. (2014). Development of mental transformation abilities. *Trends in Cognitive Sciences, 18,* 536–542. http://dx.doi.org/10.1016/j.tics.2014.05.011

*Ganley, C. M., Vasilyeva, M., & Dulaney, A. (2014). Spatial ability mediates the gender difference in middle school students' science performance. *Child Development, 85,* 1419–1432. http://dx.doi.org/10.1111/cdev.12230

Geer, E. A., Quinn, J. M., & Ganley, C. M. (2018). Relations between spatial skills and math performance in elementary school children: A longitudinal investigation. *Developmental Psychology*. Advance online publication. http://dx.doi.org/10.1037/dev0000649

*Geiser, C., Lehmann, W., Corth, M., & Eid, M. (2008). Quantitative and qualitative change in children's mental rotation performance. *Learning and Individual Differences, 18,* 419–429. http://dx.doi.org/10.1016/j.lindif.2007.09.001

Geiser, C., Lehmann, W., & Eid, M. (2006). Separating "rotators" from "nonrotators" in the mental rotations test: A multigroup latent class analysis. *Multivariate Behavioral Research, 41,* 261–293. http://dx.doi.org/10.1207/s15327906mbr4103_2

*Geiser, C., Lehmann, W., & Eid, M. (2008). A note on sex differences in mental rotation in different age groups. *Intelligence, 36,* 556–563. http://dx.doi.org/10.1016/j.intell.2007.12.003

Gilligan, K. A., Flouri, E., & Farran, E. K. (2017). The contribution of spatial ability to mathematics achievement in middle childhood. *Journal of Experimental Child Psychology, 163,* 107–125. http://dx.doi.org/10.1016/j.jecp.2017.04.016

Gilligan, K. A., Hodgkiss, A., Thomas, M. S. C., & Farran, E. K. (2018). The developmental relations between spatial cognition and mathematics in primary school children. *Developmental Science*. Advance online publication. http://dx.doi.org/10.1111/desc.12786

*Glück, J., & Fabrizii, C. (2010). Gender differences in the mental rotations test are partly explained by response format. *Journal of Individual Differences, 31,* 106–109. http://dx.doi.org/10.1027/1614-0001/a000019

*Goldin-Meadow, S., Levine, S. C., Zinchenko, E., Yip, T. K., Hemani, N., & Factor, L. (2012). Doing gesture promotes learning a mental transformation task better than seeing gesture. *Developmental Science, 15,* 876–884. http://dx.doi.org/10.1111/j.1467-7687.2012.01185.x

*Grimshaw, G. M., Sitarenios, G., & Finegan, J. A. K. (1995). Mental rotation at 7 years: Relations with prenatal testosterone levels and spatial play experiences. *Brain and Cognition, 29,* 85–100. http://dx.doi.org/10.1006/brcg.1995.1269

Guillot, A., Champely, S., Batier, C., Thiriet, P., & Collet, C. (2007). Relationship between spatial abilities, mental rotation and functional anatomy learning. *Advances in Health Sciences Education, 12,* 491–507. http://dx.doi.org/10.1007/s10459-006-9021-7

*Gunderson, E. A., Ramirez, G., Beilock, S. L., & Levine, S. C. (2012). The relation between spatial skill and early number knowledge: The role of the linear number line. *Developmental Psychology, 48,* 1229–1241. http://dx.doi.org/10.1037/a0027433

*Gurny, H. G. (2003). *High school students' performance on Vandenberg's Mental Rotations Test: Art ability, gender, activities, academic performance, strategies, and ease of taking the test* (Master's thesis). Retrieved from ERIC database. (ED479372)

*Hagevik, R. A. (2003). *The effects of online science instruction using geographic information systems to foster inquiry learning of teachers and middle school science students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 305314112)

*Hahn, N., Jansen, P., & Heil, M. (2010). Preschoolers' mental rotation: Sex differences in hemispheric asymmetry. *Journal of Cognitive Neuroscience, 22,* 1244–1250. http://dx.doi.org/10.1162/jocn.2009.21236

Hampson, E., Rovet, J. F., & Altmann, D. (1998). Spatial reasoning in children with congenital adrenal hyperplasia due to 21-hydroxylase deficiency. *Developmental Neuropsychology, 14*(2–3), 299–320. http://dx.doi.org/10.1080/87565649809540713

*Harpole, C. A. (1995). *Effects of a mental rotations curriculum on mathematical conceptualization and math anxiety in eighth-grade male and female students* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9618394)

*Harris, J., Newcombe, N. S., & Hirsh-Pasek, K. (2013). A new twist on studying the development of dynamic spatial transformations: Mental paper folding in young children. *Mind, Brain and Education, 7,* 49–55. http://dx.doi.org/10.1111/mbe.12007

*Hawes, Z., LeFevre, J. A., Xu, C., & Bruce, C. D. (2015). Mental rotation with tangible three-dimensional objects: A new measure sensitive to developmental differences in 4- to 8-year-old children. *Mind, Brain and Education, 9,* 10–18. http://dx.doi.org/10.1111/mbe.12051

*Hawes, Z., Moss, J., Caswell, B., & Poliszczuk, D. (2015). Effects of mental rotation training on children's spatial and mathematics performance: A randomized controlled study. *Trends in Neuroscience and Education, 4,* 60–68. http://dx.doi.org/10.1016/j.tine.2015.05.001

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010a). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 39–65. http://dx.doi.org/10.1002/jrsm.5

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010b). Erratum: Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1,* 164–165. http://dx.doi.org/10.1002/jrsm.17

*Heil, M., & Jansen-Osmann, P. (2007). Children's left parietal brain activation during mental rotation is reliable as well as specific. *Cognitive Development, 22,* 280–288. http://dx.doi.org/10.1016/j.cogdev.2006.10.004

Heil, M., & Jansen-Osmann, P. (2008a). Sex differences in mental rotation with polygons of different complexity: Do men utilize holistic processes whereas women prefer piecemeal ones? *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 61,* 683–689. http://dx.doi.org/10.1080/17470210701822967

*Heil, M., & Jansen-Osmann, P. (2008b). Gender differences in math and mental rotation accuracy but not in mental rotation speed in 8-years-old children. *European Journal of Developmental Science, 2*(1–2), 190–196.

*Herzberg, F., & Lepkin, M. (1954). A study of sex differences on the Primary Mental Abilities Test. *Educational and Psychological Measurement, 14,* 687–689. http://dx.doi.org/10.1177/001316445401400412

Hier, D. B., & Crowley, W. F., Jr. (1982). Spatial ability in androgen-deficient men. *The New England Journal of Medicine, 306,* 1202–1205. http://dx.doi.org/10.1056/NEJM198205203062003

*Higgins, H. J. (2006). *The relationship of sixth-grade students' mental rotation ability to spatial experience and problem-solving strategies by socioeconomic status and gender* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3239873)

*Hobson, J. R. (1947). Sex differences in primary mental abilities. *The Journal of Educational Research, 41,* 126–132. http://dx.doi.org/10.1080/00220671.1947.10881586

Hodgkiss, A., Gilligan, K. A., Tolmie, A. K., Thomas, M. S. C., & Farran, E. K. (2018). Spatial cognition and science achievement: The contribution of intrinsic and extrinsic spatial skills from 7 to 11 years. *British Journal of Educational Psychology, 88,* 675–697. http://dx.doi.org/10.1111/bjep.12211

Hooven, C. K., Chabris, C. F., Ellison, P. T., & Kosslyn, S. M. (2004). The relationship of male testosterone to components of mental rotation. *Neuropsychologia, 42,* 782–790. http://dx.doi.org/10.1016/j.neuropsychologia.2003.11.012

*Hoyek, N., Collet, C., Fargier, P., & Guillot, A. (2012). The use of the Vandenberg and Kuse Mental Rotation Test in children. *Journal of Individual Differences, 33,* 62–67. http://dx.doi.org/10.1027/1614-0001/a000063

Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology, 65,* 373–398. http://dx.doi.org/10.1146/annurev-psych-010213-115057

Hyde, J. S., & Linn, M. C. (1988). Gender differences in verbal ability: A meta-analysis. *Psychological Bulletin, 104,* 53–69. http://dx.doi.org/10.1037/0033-2909.104.1.53

*Jahoda, G. (1980). Sex and ethnic differences on a spatial-perceptual task: Some hypotheses tested. *British Journal of Psychology, 71,* 425–431. http://dx.doi.org/10.1111/j.2044-8295.1980.tb01757.x

*Jain, P. P. (2012). *Gender, video game playing, and kindergarten children's mental rotation abilities* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3515690)

*Jansen, P., Kellner, J., & Rieder, C. (2013). The improvement of mental rotation performance in second graders after creative dance training. *Creative Education, 4,* 418–422. http://dx.doi.org/10.4236/ce.2013.46060

*Jansen, P., Quaiser-Pohl, C., Neuburger, S., & Ruthsatz, V. (2015). Factors influencing mental-rotation with action-based gender-stereotyped objects—The role of fine motor skills. *Current Psychology: A Journal for Diverse Perspectives on Diverse Psychological Issues, 34,* 466–476. http://dx.doi.org/10.1007/s12144-014-9269-7

Jirout, J. J., & Newcombe, N. S. (2015). Building blocks for developing spatial skills: Evidence from a large, representative U.S. sample. *Psychological Science, 26,* 302–310. http://dx.doi.org/10.1177/0956797614563338

Jones, S., & Burnett, G. (2008). Spatial ability and learning to program. *Human Technology: An Interdisciplinary Journal on Humans in ICT Environments, 4,* 47–61. http://dx.doi.org/10.17011/ht/urn.200804151352

*Jones, T. A. (2010). *The relationship between kindergarten students' home block play and their spatial ability test scores* (Doctoral dissertation). Retrieved from ERIC database. (ED479372)

Just, M. A., & Carpenter, P. A. (1985). Cognitive coordinate systems: Accounts of mental rotation and individual differences in spatial ability. *Psychological Review, 92,* 137–172. http://dx.doi.org/10.1037/0033-295X.92.2.137

*Karádi, K., Szabó, I., Szepesi, T., Kállai, J., & Kovács, B. (1999). Sex differences on the hand mental rotation task for 9-yr.-old children and young adults. *Perceptual and Motor Skills, 89,* 969–972. http://dx.doi.org/10.2466/pms.1999.89.3.969

*Kaufman, S. B. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence, 35,* 211–223. http://dx.doi.org/10.1016/j.intell.2006.07.009

Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation: Spatial ability's unique role. *Psychological Science, 24,* 1831–1836. http://dx.doi.org/10.1177/0956797613478615

*Kerns, K. A., & Berenbaum, S. A. (1991). Sex differences in spatial ability in children. *Behavior Genetics, 21,* 383–396. http://dx.doi.org/10.1007/BF01065974

*Kirkendall, K. L. (1987). *The effect of axis of rotation on young children's judgments of spatial relations* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8809976)

*Klein, P. S., Adi-Japha, E., & Hakak-Benizri, S. (2010). Mathematical thinking of kindergarten boys and girls: Similar achievement, different contributing processes. *Educational Studies in Mathematics, 73,* 233–246. http://dx.doi.org/10.1007/s10649-009-9216-y

*Kovac, R. J. (1989). The validation of selected spatial ability tests via correlational assessment and analysis of user-processing strategy. *Educational Research Quarterly, 13,* 26–34. Retrieved from ERIC database. (EJ415906)

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science, 31,* 549–579. http://dx.doi.org/10.1080/15326900701399897

*Krüger, M., & Krist, H. (2009). Imagery and motor processes—When are they connected? The mental rotation of body parts in development. *Journal of Cognition and Development, 10,* 239–261. http://dx.doi.org/10.1080/15248370903389341

*Kucian, K., von Aster, M., Loenneker, T., Dietrich, T., Mast, F. W., & Martin, E. (2007). Brain activation during mental rotation in school children and adults. *Journal of Neural Transmission, 114,* 675–686. http://dx.doi.org/10.1007/s00702-006-0604-5

*Lange-Küttner, C., & Ebersbach, M. (2013). Girls in detail, boys in shape: Gender differences when drawing cubes in depth. *British Journal of Psychology, 104,* 413–437. http://dx.doi.org/10.1111/bjop.12010

*Lauer, J. E., Esposito, A. G., & Bauer, P. J. (2016, October). *Affective precursors to gender differences in spatial and mathematical performance*. Poster presented at the Gender Development Research Conference, San Francisco, CA.

Lauer, J. E., Esposito, A. G., & Bauer, P. J. (2018). Domain-specific anxiety relates to children's math and spatial performance. *Developmental Psychology, 54,* 2126–2138. http://dx.doi.org/10.1037/dev0000605

Lauer, J. E., Ilksoy, S. D., & Lourenco, S. F. (2018). Developmental stability in gender-typed preferences between infancy and preschool age. *Developmental Psychology, 54,* 613–620. http://dx.doi.org/10.1037/dev0000468

*Lauer, J. E., & Lourenco, S. F. (2015, October). *Characterizing the relation between spatial and mathematical competence in preschool-aged children*. Poster presented at the Biennial Meeting of the Cognitive Development Society (CDS), Columbus, OH.

*Lauer, J. E., & Lourenco, S. F. (2016). Spatial processing in infancy predicts both spatial and mathematical aptitude in childhood. *Psychological Science, 27,* 1291–1298. http://dx.doi.org/10.1177/0956797616655977

Lauer, J. E., Udelson, H. B., Jeon, S. O., & Lourenco, S. F. (2015). An early sex difference in the relation between mental rotation and object preference. *Frontiers in Psychology, 6,* 558. http://dx.doi.org/10.3389/fpsyg.2015.00558

Lawton, C. A. (1994). Gender differences in way-finding strategies: Relationship to spatial ability and spatial anxiety. *Sex Roles: A Journal of Research, 30,* 765–779. http://dx.doi.org/10.1007/BF01544230

*Lehmann, J., Quaiser-Pohl, C., & Jansen, P. (2014). Correlation of motor skill, mental rotation, and working memory in 3- to 6-year-old children. *European Journal of Developmental Psychology, 11,* 560–573. http://dx.doi.org/10.1080/17405629.2014.888995

*LeVasseur, M. L. (1999). *An analysis of selected barriers to achievement levels in geography of female students in grade six and grade nine* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9958927)

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *WIREs Cognitive Science, 7,* 127–155. http://dx.doi.org/10.1002/wcs.1380

*Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology, 35,* 940–949. http://dx.doi.org/10.1037/0012-1649.35.4.940

*Levine, S. C., Ratliff, K. R., Huttenlocher, J., & Cannon, J. (2012). Early puzzle play: A predictor of preschoolers' spatial transformation skill. *Developmental Psychology, 48,* 530–542. http://dx.doi.org/10.1037/a0025913

*Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science, 16,* 841–845. http://dx.doi.org/10.1111/j.1467-9280.2005.01623.x

*Li, C., Zhu, W., & Nuttall, R. L. (2003). Familial handedness and spatial ability: A study with Chinese students aged 14–24. *Brain and Cognition, 51,* 375–384. http://dx.doi.org/10.1016/S0278-2626(03)00041-1

Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. (2010). New trends in gender and mathematics performance: A meta-analysis. *Psychological Bulletin, 136,* 1123–1135. http://dx.doi.org/10.1037/a0021276

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479–1498. http://dx.doi.org/10.2307/1130467

*Lizarraga, M. L. S., & Ganuza, J. G. (2003). Improvement of mental rotation in girls and boys. *Sex Roles: A Journal of Research, 49,* 277–286. http://dx.doi.org/10.1023/A:1024656408294

López-López, J. A., Van den Noortgate, W., Tanner-Smith, E. E., Wilson, S. J., & Lipsey, M. W. (2017). Assessing meta-regression methods for examining moderator relationships with dependent effect sizes: A Monte Carlo simulation. *Research Synthesis Methods, 8,* 435–450. http://dx.doi.org/10.1002/jrsm.1245

*Lütke, N., & Lange-Küttner, C. (2015). Keeping it in three dimensions: Measuring the development of mental rotation in children with the Rotated Colour Cube Test (RCCT). *International Journal of Developmental Science, 9,* 95–114. http://dx.doi.org/10.3233/DEV-14154

Lyons, I. M., Ramirez, G., Maloney, E. A., Rendina, D. N., Levine, S. C., & Beilock, S. L. (2018). Spatial Anxiety: A novel questionnaire with subscales for measuring three aspects of spatial anxiety. *Journal of Numerical Cognition, 4,* 526–553. http://dx.doi.org/10.5964/jnc.v4i3.154

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences*. Stanford, CA: Stanford University Press.

Maeda, Y., & Yoon, S. Y. (2013). A meta-analysis on gender differences in mental rotation ability measured by the Purdue spatial visualization tests: Visualization of rotations (PSVT:R). *Educational Psychology Review, 25,* 69–94. http://dx.doi.org/10.1007/s10648-012-9215-x

*Mann, V. A., Sasanuma, S., Sakuma, N., & Masaki, S. (1990). Sex differences in cognitive abilities: A cross-cultural perspective. *Neuropsychologia, 28,* 1063–1077. http://dx.doi.org/10.1016/0028-3932(90)90141-A

McShane, B. B., Böckenholt, U., & Hansen, K. T. (2016). Adjusting for publication bias in meta-analysis: An evaluation of selection methods and some cautionary notes. *Perspectives on Psychological Science, 11,* 730–749. http://dx.doi.org/10.1177/1745691616662243

*Meyer, W. J., & Bendig, A. (1961). A longitudinal study of the Primary Mental Abilities Test. *Journal of Educational Psychology, 52,* 50–60. http://dx.doi.org/10.1037/h0046088

Miller, D. I., & Halpern, D. F. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences, 18,* 37–45. http://dx.doi.org/10.1016/j.tics.2013.10.011

*Miller, R. B., Kelly, G. N., & Kelly, J. T. (1988). Effects of Logo computer programming experience on problem solving and spatial relations ability. *Contemporary Educational Psychology, 13,* 348–357. http://dx.doi.org/10.1016/0361-476X(88)90034-3

*Moè, A. (2009). Are males always better than females in mental rotation? Exploring a gender belief explanation. *Learning and Individual Differences, 19,* 21–27. http://dx.doi.org/10.1016/j.lindif.2008.02.002

*Moè, A. (2012). Gender difference does not mean genetic difference: Externalizing improves performance in mental rotation. *Learning and Individual Differences, 22,* 20–24. http://dx.doi.org/10.1016/j.lindif.2011.11.001

Moè, A. (2018). Mental rotation and mathematics: Gender-stereotyped beliefs and relationships in primary school children. *Learning and Individual Differences, 61,* 172–180. http://dx.doi.org/10.1016/j.lindif.2017.12.002

*Moè, A., & Pazzaglia, F. (2006). Following the instructions!: Effects of gender beliefs in mental rotation. *Learning and Individual Differences, 16,* 369–377. http://dx.doi.org/10.1016/j.lindif.2007.01.002

*Moè, A., & Pazzaglia, F. (2010). Beyond genetics in Mental Rotation Test performance: The power of effort attribution. *Learning and Individual Differences, 20,* 464–468. http://dx.doi.org/10.1016/j.lindif.2010.03.004

Möhring, W., & Frick, A. (2013). Touching up mental rotation: Effects of manual experience on 6-month-old infants' mental object rotation. *Child Development, 84,* 1554–1565. http://dx.doi.org/10.1111/cdev.12065

Monahan, J. S., Harke, M. A., & Shelley, J. R. (2008). Computerizing the Mental Rotations Test: Are gender differences maintained? *Behavior Research Methods, 40,* 422–427. http://dx.doi.org/10.3758/BRM.40.2.422

Moore, D. S., & Johnson, S. P. (2008). Mental rotation in human infants: A sex difference. *Psychological Science, 19,* 1063–1066. http://dx.doi.org/10.1111/j.1467-9280.2008.02200.x

*Moyer, T. O. (2003). *An investigation of The Geometer's Sketchpad and van Hiele levels* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3112299)

*Neuburger, S., Jansen, P., Heil, M., & Quaiser-Pohl, C. (2012). A threat in the classroom: Gender stereotype activation and mental-rotation performance in elementary-school children. *Zeitschrift für Psychologie, 220,* 61–69. http://dx.doi.org/10.1027/2151-2604/a000097

*Neuburger, S., Ruthsatz, V., Jansen, P., & Quaiser-Pohl, C. (2015). Can girls think spatially? Influence of implicit gender stereotype activation and rotational axis on fourth graders' mental-rotation performance. *Learning and Individual Differences, 37,* 169–175. http://dx.doi.org/10.1016/j.lindif.2014.09.003

*Noda, M. (2008). Imagery and perceptual basis of matching tasks in young children. *Perceptual and Motor Skills, 107,* 419–438. http://dx.doi.org/10.2466/pms.107.2.419-438

*Noda, M. (2014). Transformation of the incomplete figure in young children. *International Journal of Behavioral Development, 38,* 23–32. http://dx.doi.org/10.1177/0165025413503422

*Oostermeijer, M., Boonen, A. J. H., & Jolles, J. (2014). The relation between children's constructive play activities, spatial ability, and mathematical word problem-solving performance: A mediation analysis in sixth-grade students. *Frontiers in Psychology, 5,* 782. http://dx.doi.org/10.3389/fpsyg.2014.00782

*Patkin, D., & Dayan, E. (2013). The intelligence of observation: Improving high school students' spatial ability by means of intervention unit. *International Journal of Mathematical Education in Science and Technology, 44,* 179–195. http://dx.doi.org/10.1080/0020739X.2012.703335

*Patrick, C. L. (1997). *The relations among cognitive strategy, mental rotation performance and mathematical reasoning* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9808841)

*Pazzaglia, F., & Moè, A. (2013). Cognitive styles and mental rotation ability in map learning. *Cognitive Processing, 14,* 391–399. http://dx.doi.org/10.1007/s10339-013-0572-2

*Perner, J., Kloo, D., & Rohwer, M. (2010). Retro- and prospection for mental time travel: Emergence of episodic remembering and mental rotation in 5- to 8-year old children. *Consciousness and Cognition, 19,* 802–815. http://dx.doi.org/10.1016/j.concog.2010.06.022

*Perry, P. C. (2013). *Influences on visual spatial rotation: Science, technology, engineering, and mathematics (STEM) experiences, age, and gender* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3557721)

Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse mental rotations test: Different versions and factors that affect performance. *Brain and Cognition, 28,* 39–58. http://dx.doi.org/10.1006/brcg.1995.1032

*Pezaris, E., & Casey, M. B. (1991). Girls who use "masculine" problem-solving strategies on a spatial task: Proposed genetic and environmental factors. *Brain and Cognition, 17,* 1–22. http://dx.doi.org/10.1016/0278-2626(91)90062-D

*Ping, R., Ratliff, K., Hickey, E., & Levine, S. C. (2011). Using manual rotation and gesture to improve mental rotation in preschoolers. In L. Carlson, C. Hölscher, & T. Shipley (Eds.), *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 459–464). Boston, MA: Cognitive Science Society.

*Pleet, L. J. (1991). *The effects of computer graphics and Mira on acquisition of transformation geometry concepts and development of mental rotation skills in grade eight* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9130918)

Polinsky, N., Perez, J., Grehl, M., & McCrink, K. (2017). Encouraging spatial talk: Using children's museums to bolster spatial reasoning. *Mind, Brain and Education, 11,* 144–152. http://dx.doi.org/10.1111/mbe.12145

*Prieto, G., & Delgado, A. R. (1999). The role of instructions in the variability of sex-related differences in multiple-choice tests. *Personality and Individual Differences, 27,* 1067–1077. http://dx.doi.org/10.1016/S0191-8869(99)00049-5

Pruden, S. M., & Levine, S. C. (2017). Parents' spatial language mediates a sex difference in preschoolers' spatial-language use. *Psychological Science, 28,* 1583–1596. http://dx.doi.org/10.1177/0956797617711968

Pruden, S. M., Levine, S. C., & Huttenlocher, J. (2011). Children's spatial thinking: Does talk about the spatial world matter? *Developmental Science, 14,* 1417–1430. http://dx.doi.org/10.1111/j.1467-7687.2011.01088.x

Pustejovsky, J. (2017). clubSandwich: Cluster-Robust (Sandwich) variance estimators with small-sample corrections. R Package Version 0.3.0. Available at: https://CRAN.R-project.org/package=clubSandwich

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,* 282–289. http://dx.doi.org/10.1016/j.yhbeh.2010.03.005

*Quaiser-Pohl, C., Neuburger, S., Heil, M., Jansen, P., & Schmelter, A. (2014). Is the male advantage in mental-rotation performance task independent? On the usability of chronometric tests and paper-and-pencil tests in children. *International Journal of Testing, 14,* 122–142. http://dx.doi.org/10.1080/15305058.2013.860148

Quinn, P. C., & Liben, L. S. (2008). A sex difference in mental rotation in young infants. *Psychological Science, 19,* 1067–1070. http://dx.doi.org/10.1111/j.1467-9280.2008.02201.x

*Ramirez, G., Gunderson, E. A., Levine, S. C., & Beilock, S. L. (2012). Spatial anxiety relates to spatial abilities as a function of working memory in children. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology, 65,* 474–487. http://dx.doi.org/10.1080/17470218.2011.616214

Reilly, D., & Neumann, D. L. (2013). Gender-role differences in spatial ability: A meta-analytic review. *Sex Roles: A Journal of Research, 68,* 521–535. http://dx.doi.org/10.1007/s11199-013-0269-0

Resnick, S. M., Berenbaum, S. A., Gottesman, I. I., & Bouchard, T. J. (1986). Early hormonal influences on cognitive functioning in congenital adrenal hyperplasia. *Developmental Psychology, 22,* 191–198. http://dx.doi.org/10.1037/0012-1649.22.2.191

*Rosser, R. A., Ensing, S. S., Glider, P. J., & Lane, S. (1984). An information-processing analysis of children's accuracy in predicting the appearance of rotated stimuli. *Child Development, 55,* 2204–2211. http://dx.doi.org/10.2307/1129792

*Rosser, R. A., Ensing, S. S., & Mazzeo, J. (1985). The role of stimulus salience in young children's ability to discriminate two-dimensional rotations: Reflections on a paradigm. *Contemporary Educational Psychology, 10,* 95–103. http://dx.doi.org/10.1016/0361-476X(85)90010-4

*Rudoff, J. (1983). *Factors influencing the ability to rotate mental images during the preschool years* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 8323919)

*Sağlam, M., & Millar, R. (2006). Upper high school students' understanding of electromagnetism. *International Journal of Science Education, 28,* 543–566. http://dx.doi.org/10.1080/09500690500339613

Sanders, B., Soares, M. P., & D'Aquila, J. M. (1982). The sex difference on one test of spatial visualization: A nontrivial difference. *Child Development, 53,* 1106–1110. Advance online publication. http://dx.doi.org/10.2307/1129153

*Scholnick, E. K., Fein, G. G., & Campbell, P. F. (1990). Changing predictors of map use in wayfinding. *Developmental Psychology, 26,* 188–193. http://dx.doi.org/10.1037/0012-1649.26.2.188

Schwarzer, G., Freitag, C., & Schum, N. (2013). How crawling and manual object exploration are related to the mental rotation abilities of 9-month-old infants. *Frontiers in Psychology, 4,* 97. http://dx.doi.org/10.3389/fpsyg.2013.00097

*Sekiyama, K., Kinoshita, T., & Soshi, T. (2014). Strong biomechanical constraints on young children's mental imagery of hands. *Royal Society Open Science, 1,* 140118. http://dx.doi.org/10.1098/rsos.140118

Serbin, L. A., Zelkowitz, P., Doyle, A. B., Gold, D., & Wheaton, B. (1990). The socialization of sex-differentiated skills and academic performance: A mediational model. *Sex Roles: A Journal of Research, 23,* 613–628. http://dx.doi.org/10.1007/BF00289251

*Shavalier, M. E. (1999). *A study of the effects of structured activities using CAD-like software on the spatial ability of fourth, fifth, and sixth grade students*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 9939273)

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science, 171,* 701–703. http://dx.doi.org/10.1126/science.171.3972.701

*Signorella, M. L., & Jamison, W. (1978). Sex differences in the correlations among field dependence, spatial ability, sex role orientation, and performance on Piaget's water-level task. *Developmental Psychology, 14,* 689–690. http://dx.doi.org/10.1037/0012-1649.14.6.689

Silverman, I., & Phillips, K. (1993). Effects of estrogen changes during the menstrual cycle on spatial performance. *Ethology & Sociobiology, 14,* 257–269. http://dx.doi.org/10.1016/0162-3095(93)90021-9

*Silverman, I., Phillips, K., & Silverman, L. K. (1996). Homogeneity of effect sizes for sex across spatial tests and cultures: Implications for hormonal theories. *Brain and Cognition, 31,* 90–94. http://dx.doi.org/10.1006/brcg.1996.0027

*Singleton, C. (2001). Computer-based assessment in education. *Educational and Child Psychology, 18,* 58–74.

*Sorby, S. A., Drummer, T., Hungwe, K., Parolini, L., & Molzan, R. (2006). *Preparing for engineering studies: Improving the 3D spatial skills of K-12 students*. Paper presented at the 9th International Conference on Engineering Education, San Juan, PR. Retrieved from http://icee.usm.edu/icee/conferences/icee2006/papers/3195.pdf

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction, 17,* 219–234. http://dx.doi.org/10.1016/j.learninstruc.2007.01.012

Su, R., Rounds, J., & Armstrong, P. I. (2009). Men and things, women and people: A meta-analysis of sex differences in interests. *Psychological Bulletin, 135,* 859–884. http://dx.doi.org/10.1037/a0017364

Tanner-Smith, E. E., & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations including a software tutorial in Stata and SPSS. *Research Synthesis Methods, 5,* 13–30. http://dx.doi.org/10.1002/jrsm.1091

Tanner-Smith, E. E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta-analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology, 2,* 85–112. http://dx.doi.org/10.1007/s40865-016-0026-5

Tanzer, N. K., Gittler, G., & Sim, C. Q. E. (1994). A cross-cultural comparison of a Rasch calibrated spatial ability test between Austrian and Singaporean adolescents. In A. M. Bouvy, F. J. R. van de Vijver, P. Boski, & P. G. Schmitz (Eds.), *Journeys into Cross-Cultural Psychology* (pp. 96–110). Lisse, The Netherlands: Swets & Zeitlinger Publishers.

*Tentomas, P. C. (2010). *Creation and assessment of a ten-day spatial transformation intervention in a secondary school geometry class* (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3424993)

Thurstone, L. L., & Thurstone, T. G. (1943). *Chicago tests of primary mental abilities: Manual of instructions*. Chicago, IL: Science Research Association.

Tipton, E., & Pustejovsky, J. E. (2015). Small-sample adjustments for tests of moderators and model fit using robust variance estimation in meta-regression. *Journal of Educational and Behavioral Statistics, 40,* 604–634. http://dx.doi.org/10.3102/1076998615606099

*Titze, C., Jansen, P., & Heil, M. (2010a). Mental rotation performance and the effect of gender in fourth graders and adults. *European Journal of Developmental Psychology, 7,* 432–444. http://dx.doi.org/10.1080/17405620802548214

*Titze, C., Jansen, P., & Heil, M. (2010b). Mental rotation performance in fourth graders: No effects of gender beliefs (yet?). *Learning and Individual Differences, 20,* 459–463. http://dx.doi.org/10.1016/j.lindif.2010.04.003

*Titze, C., Jansen, P., & Heil, M. (2011). Single-sex school girls outperform girls attending a co-educative school in mental rotation accuracy. *Sex Roles: A Journal of Research, 65,* 704–711. http://dx.doi.org/10.1007/s11199-011-9947-y

*Tlauka, M., Williams, J., & Williamson, P. (2008). Spatial ability in secondary school students: Intra-sex differences based on self-selection for physical education. *British Journal of Psychology, 99,* 427–440. http://dx.doi.org/10.1348/000712608X282806

*Toptaş, V., Çelik, S., & Karaca, E. T. (2012). Improving 8th grades spatial thinking abilities through a 3D modeling program. *TOJET: The Turkish Online Journal of Educational Technology, 11,* 128–134. Retrieved from ERIC database. (EJ989021O)

*Tracy, D. M. (1990). Toy-playing behavior, sex-role orientation, spatial ability, and science achievement. *Journal of Research in Science Teaching, 27,* 637–649. http://dx.doi.org/10.1002/tea.3660270704

*Tzuriel, D., & Egozi, G. (2007). Dynamic assessment of spatial abilities of young children: Effects of gender and task characteristics. *Journal of*

Cognitive Education and Psychology, 6, 219–247. http://dx.doi.org/10.1891/194589507787382160

*Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development, 81,* 1417–1430. http://dx.doi.org/10.1111/j.1467-8624.2010.01482.x

United Nations Development Programme. (2015). *Human development report*. Retrieved from http://hdr.undp.org/en/composite/GDI

Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin, 139,* 352–402. http://dx.doi.org/10.1037/a0028446

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47,* 599–604. http://dx.doi.org/10.2466/pms.1978.47.2.599

*Vederhus, L., & Krekling, S. (1996). Sex differences in visual spatial ability in 9-year-old children. *Intelligence, 23,* 33–43. http://dx.doi.org/10.1016/S0160-2896(96)80004-3

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., & Newcombe, N. S. (2017). Links between spatial and mathematical skills across the preschool years. *Monographs of the Society for Research in Child Development, 82,* 1–150.

Verdine, B. N., Golinkoff, R. M., Hirsh-Pasek, K., Newcombe, N. S., Filipowicz, A. T., & Chang, A. (2014). Deconstructing building blocks: Preschoolers' spatial assembly performance relates to early mathematical skills. *Child Development, 85,* 1062–1076. http://dx.doi.org/10.1111/cdev.12165

Vevea, J. L., & Hedges, L. V. (1995). A general linear model for estimating effect size in the presence of publication bias. *Psychometrika, 60,* 419–435. http://dx.doi.org/10.1007/BF02294384

Vevea, J. L., & Woods, C. M. (2005). Publication bias in research synthesis: Sensitivity analysis using a priori weight functions. *Psychological Methods, 10,* 428–443. http://dx.doi.org/10.1037/1082-989X.10.4.428

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metafor package. *Journal of Statistical Software, 36,* 1–48. http://www.jstatsoft.org/v36/i03/. http://dx.doi.org/10.18637/jss.v036.i03

Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: A meta-analysis. *Psychonomic Bulletin & Review, 18,* 267–277. http://dx.doi.org/10.3758/s13423-010-0042-0

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,* 250–270. http://dx.doi.org/10.1037/0033-2909.117.2.250

Voyer, D., Voyer, S. D., & Saint-Aubin, J. (2017). Sex differences in visual-spatial working memory: A meta-analysis. *Psychonomic Bulletin & Review, 24,* 307–334. http://dx.doi.org/10.3758/s13423-016-1085-7

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,* 817–835. http://dx.doi.org/10.1037/a0016127

Wai, J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2010). Accomplishment in Science, Technology, Engineering, and Mathematics (STEM) and its relation to STEM educational dose: A 25-year longitudinal study. *Journal of Educational Psychology, 102,* 860–871. http://dx.doi.org/10.1037/a0019454

Wang, M. T., & Degol, J. (2013). Motivational pathways to STEM career choices: Using expectancy–value perspective to understand individual and gender differences in STEM fields. *Developmental Review, 33,* 304–340. http://dx.doi.org/10.1016/j.dr.2013.08.001

*Weckbacher, L. M., & Okamoto, Y. (2014). Mental rotation ability in relation to self-perceptions of high school geometry. *Learning and Individual Differences, 30,* 58–63. http://dx.doi.org/10.1016/j.lindif.2013.10.007

*Wei, W., Lu, H., Zhao, H., Chen, C., Dong, Q., & Zhou, X. (2012). Gender differences in children's arithmetic performance are accounted for by gender differences in language abilities. *Psychological Science, 23,* 320–330. http://dx.doi.org/10.1177/0956797611427168

Wei, W., Yuan, H., Chen, C., & Zhou, X. (2012). Cognitive correlates of performance in advanced mathematics. *British Journal of Educational Psychology, 82,* 157–181. http://dx.doi.org/10.1111/j.2044-8279.2011.02049.x

*Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and Instruction, 18,* 30–41. http://dx.doi.org/10.1016/j.learninstruc.2006.09.009

Wu, H. K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education, 88,* 465–492. http://dx.doi.org/10.1002/sce.10126

*Xistouri, X., & Pitta-Pantazi, D. (2006). *Spatial rotation and perspective taking abilities in relation to performance in reflective symmetry tasks*. Paper presented at the 30th Conference of the International Group for the Psychology of Mathematics Education, Prague, Czech Republic. Retrieved from ERIC database. (ED496939)

*Yang, J. C., & Chen, S. Y. (2010). Effects of gender differences and spatial abilities within a digital pentominoes game. *Computers & Education, 55,* 1220–1233. http://dx.doi.org/10.1016/j.compedu.2010.05.019

*Yen, W. M. (1975). Sex-linked major-gene influences on selected types of spatial performance. *Behavior Genetics, 5,* 281–298. http://dx.doi.org/10.1007/BF01066180

*Young, C., Cartmill, E., Levine, S., & Goldin-Meadow, S. (2014). Gesture and speech input are interlocking pieces: The development of children's jigsaw puzzle assembly ability. In *Proceedings of the 33rd Annual Meeting of the Cognitive Science Society* (pp. 1820–1825). Quebec City, Canada: Cognitive Science Society.

*Yurt, E., & Sünbül, A. M. (2012). Effect of modeling-based activities developed using virtual environments and concrete objects on spatial thinking and mental rotation skills. *Educational Sciences: Theory and Practice, 12,* 1987–1992.

*Yurt, E., & Sünbül, A. M. (2014). A structural equation model explaining 8th grade students' mathematics achievements. *Educational Sciences: Theory and Practice, 14,* 1642–1652. http://dx.doi.org/10.12738/estp.2014.4.2193

Appendix A

Reported *g*, Sample Size (*n*), Mean Sample Age in Years, and Procedural Characteristics for Effect Sizes (ES) in Meta-Analysis

ES No.	Article	Pub. status	*n*	*M* age	Task	Stimulus dim.	Stimulus demands	Stimulus type	*g*
1	Amponsah (2000)	P	249	9.0	PMA-SR	2D	MD	Abstract	.60
2		P	266	12.0	PMA-SR	2D	MD	Abstract	.60
3		P	303	16.0	PMA-SR	2D	MD	Abstract	.44
4		P	193	9.0	PMA-SR	2D	MD	Abstract	.78
5		P	297	12.0	PMA-SR	2D	MD	Abstract	.48
6		P	310	16.0	PMA-SR	2D	MD	Abstract	.33
7		P	249	9.0	VMRT	3D	MD	Abstract	.71
8		P	266	12.0	VMRT	3D	MD	Abstract	.39
9		P	303	16.0	VMRT	3D	MD	Abstract	.61
10		P	193	9.0	VMRT	3D	MD	Abstract	.55
11		P	297	12.0	VMRT	3D	MD	Abstract	.60
12		P	310	16.0	VMRT	3D	MD	Abstract	.95
13	Anderson et al. (2008)	P	186	12.5	Other	3D	MD	Abstract	.63
14	Auyeung et al. (2012)	P	64	9.1	Other	2D	MD	Animate	.61
15	Battista (1990)	P	128	16.5	Other	3D	NA	Abstract	.94
16	Boakes (2006)	NP	25	12.5	Other	2D	MD	Abstract	.76
17		NP	31	12.5	Other	2D	MD	Abstract	−.29
18	Boulter (1992)	NP	70	13.0	Other	2D	MD	Abstract	.57
19	Carr, Steiner, Kyser, and Biddlecomb (2008)	P	241	7.5	VMRT	3D	MD	Abstract	.51
20	Casey, Colón, and Goris (1992)	P	71	16.0	VMRT	3D	MD	Abstract	1.17
21	Casey, Nuttall, Pezaris, and Benbow (1995)	P	300	16.0	VMRT	3D	MD	Abstract	.23
22	Casey, Nuttall, and Pezaris (1999)	P	365	13.9	VMRT	3D	MD	Abstract	.50
23	Casey, Nuttall, and Pezaris (2001)	P	187	13.8	VMRT	3D	MD	Abstract	.64
24	Casey et al. (2008)	P	89	6.2	Other	3D	NA	Abstract	.56
25	Cheng and Mix (2014)	P	31	7.1	CMTT	2D	NMD	Abstract	−.57
26		P	27	7.1	CMTT	2D	NMD	Abstract	−.41
27		P	31	7.1	PMA-SR	2D	NMD	Abstract	−.25
28		P	27	7.1	PMA-SR	2D	NMD	Abstract	.19
29	Chien (1986)	NP	12	6.5	Other	2D	MD	Other	−.64
30		NP	12	6.5	Other	2D	MD	Other	.58
31		NP	12	8.5	Other	2D	MD	Other	.35
32		NP	12	8.5	Other	2D	MD	Other	.55
33		NP	12	10.5	Other	2D	MD	Other	−.34
34		NP	12	10.5	Other	2D	MD	Other	.41
35	Choi (2000)	NP	25	8.0	PMA-SR	2D	NMD	Abstract	−.68*
36		NP	28	9.0	PMA-SR	2D	NMD	Abstract	.03
37		NP	40	10.0	PMA-SR	2D	NMD	Abstract	.22
38		NP	65	11.0	PMA-SR	2D	MD	Abstract	.46
39		NP	71	12.0	PMA-SR	2D	MD	Abstract	.39
40		NP	86	13.0	PMA-SR	2D	MD	Abstract	.30
41		NP	55	14.0	PMA-SR	2D	MD	Abstract	−.04
42		NP	54	15.0	PMA-SR	2D	MD	Abstract	.85
43		NP	100	16.0	PMA-SR	2D	MD	Abstract	.43
44		NP	82	17.0	PMA-SR	2D	MD	Abstract	.71
45		NP	55	14.0	Other	2D	MD	Abstract	−.41
46		NP	54	15.0	Other	2D	MD	Abstract	.71
47		NP	100	16.0	Other	2D	MD	Abstract	.35
48		NP	82	17.0	Other	2D	MD	Abstract	.76
49	Clements, Battista, Sarama, and Swaminathan (1997)	P	23	9.5	Other	2D	MD	Abstract	.58
50	Davison and Susman (2001)	P	108	12.4	PMA-SR	2D	MD	Abstract	.68
51	De Lisi and Wolford (2002)	P	24	8.5	Other	2D	MD	Abstract	.92
52		P	23	8.5	Other	2D	MD	Abstract	.82
53	Delgado and Prieto (2004)	P	455	13.5	VMRT	3D	MD	Abstract	.77
54	Dunst, Benedek, Bergner, Athenstaedt, and Neubauer (2013)	P	29	16.5	Other	3D	MD	Abstract	.04

(*Appendices continue*)

Appendix A (*continued*)

ES No.	Article	Pub. status	*n*	*M* age	Task	Stimulus dim.	Stimulus demands	Stimulus type	*g*
55	Dziak (1985)	NP	95	13.0	Other	2D	MD	Abstract	.49
56	Ehrlich, Levine, and Goldin-Meadow (2006)	P	80	5.6	CMTT	2D	NMD	Abstract	.58
57	Eikenberry (1988)	NP	108	8.5	Other	2D	MD	Abstract	.42
58	Eraso (2007)	NP	63	15.5	Other	3D	NA	Abstract	.37
59	Erkoç, Gecü, and Erkoç (2013)	P	31	13.5	VMRT	3D	MD	Abstract	.16
60		P	31	13.5	VMRT	3D	MD	Abstract	.52
61	Frick, Daum, Walser, and Mast (2009)	P	20	5.6	Other	2D	MD	Abstract	.14
62		P	20	8.5	Other	2D	MD	Abstract	−.55
63		P	20	11.4	Other	2D	MD	Abstract	−.17
64	Frick, Ferrara, and Newcombe (2013)	P	48	4.0	Other	2D	MD	Abstract	.00
65		P	48	5.0	Other	2D	MD	Abstract	.72
66	Frick, Hansen, and Newcombe (2013)	P	20	3.5	Other	3D	MD	Animate	.39
67		P	20	4.5	Other	3D	MD	Animate	.29
68		P	20	5.5	Other	3D	MD	Animate	−.04
69		P	20	4.6	Other	2D	MD	Animate	.10
70		P	20	5.5	Other	2D	MD	Animate	−.16
71	Ganley, Vasilyeva, and Dulaney (2014)	P	113	14.0	VMRT	3D	MD	Abstract	.61
72	Geiser, Lehmann, Corth, and Eid (2008)	P	519	10.5	VMRT	3D	MD	Abstract	.45
73	Geiser, Lehmann, and Eid (2008)	P	40	9.0	VMRT	3D	MD	Abstract	1.06
74		P	128	10.0	VMRT	3D	MD	Abstract	.82
75		P	90	11.0	VMRT	3D	MD	Abstract	.73
76		P	154	12.0	VMRT	3D	MD	Abstract	.61
77		P	124	13.0	VMRT	3D	MD	Abstract	.63
78		P	178	14.0	VMRT	3D	MD	Abstract	.93
79		P	102	15.0	VMRT	3D	MD	Abstract	.96
80		P	107	16.0	VMRT	3D	MD	Abstract	.68
81		P	137	17.0	VMRT	3D	MD	Abstract	.83
82	Glück and Fabrizii (2010)	P	288	16.5	VMRT	3D	MD	Abstract	.53
83	Goldin-Meadow et al. (2012)	P	158	6.1	CMTT	2D	NMD	Abstract	.05
84	Grimshaw, Sitarenios, and Finegan (1995)	P	25	7.0	Other	2D	MD	Animate	.29
85		P	16	7.0	Other	2D	MD	Animate	.06
86	Gunderson, Ramirez, Beilock, and Levine (2012)	P	42	5.4	CMTT	2D	NMD	Abstract	.43
87		P	87	6.5	PMA-SR	2D	NMD	Abstract	.22
88		P	65	7.5	PMA-SR	2D	NMD	Abstract	−.22
89	Gurny (2003)	NP	186	16.3	VMRT	3D	MD	Abstract	.91
90	Hagevik (2003)	NP	164	13.0	Other	3D	NA	Abstract	1.09
91	Hahn, Jansen, and Heil (2010)	P	97	5.4	Other	2D	MD	Animate	.44
92	Harpole (1995)	NP	56	13.5	VMRT	3D	MD	Abstract	.38
93	Harris, Newcombe, and Hirsh-Pasek (2013)	P	165	5.7	CMTT	2D	NMD	Abstract	.01
94	Hawes, LeFevre, et al. (2015)	P	25	4.5	CMTT	2D	NMD	Abstract	−.10
95		P	25	5.3	CMTT	2D	NMD	Abstract	.03
96		P	26	5.8	CMTT	2D	NMD	Abstract	1.46*
97		P	28	6.3	CMTT	2D	NMD	Abstract	−.64
98		P	25	6.8	CMTT	2D	NMD	Abstract	.49
99		P	28	7.6	CMTT	2D	NMD	Abstract	.06
100		P	25	4.5	Other	3D	MD	Abstract	.29
101		P	26	5.3	Other	3D	MD	Abstract	.13
102		P	26	5.8	Other	3D	MD	Abstract	−.13
103		P	32	6.3	Other	3D	MD	Abstract	−.04
104		P	26	6.8	Other	3D	MD	Abstract	1.04
105		P	28	7.6	Other	3D	MD	Abstract	−.12
106	Hawes, Moss, et al. (2015)	P	61	7.2	CMTT	2D	NMD	Abstract	−.02
107		P	61	7.2	Other	2D	NMD	Abstract	−.31
108		P	61	7.2	Other	2D	MD	Animate	−.02
109		P	61	7.2	Other	2D	MD	Other	−.04
110		P	61	7.2	VMRT	3D	MD	Abstract	.35
111	Heil and Jansen-Osmann (2007)	P	24	8.6	Other	2D	MD	Other	−.29
112	Heil and Jansen-Osmann (2008b)	P	109	7.7	Other	3D	MD	Abstract	.51
113		P	109	7.7	Other	2D	MD	Animate	.40
114	Herzberg and Lepkin (1954)	P	189	16.5	PMA-SR	2D	MD	Abstract	.42

(*Appendices continue*)

Appendix A (*continued*)

ES No.	Article	Pub. status	n	M age	Task	Stimulus dim.	Stimulus demands	Stimulus type	g
115		P	705	17.5	PMA-SR	2D	MD	Abstract	.48
116	Higgins (2006)	NP	86	11.7	VMRT	3D	MD	Abstract	.55
117		NP	92	11.7	VMRT	3D	MD	Abstract	.56
118	Hobson (1947)	P	472	14.5	PMA-SR	2D	MD	Abstract	.47
119		P	525	14.5	PMA-SR	2D	MD	Abstract	.44
120		P	1,436	13.5	PMA-SR	2D	MD	Abstract	.49
121	Hoyek, Collet, Fargier, and Guillot (2012)	P	28	7.8	Other	2D	MD	Abstract	−.15
122		P	66	11.4	Other	2D	MD	Abstract	.88
123		P	28	7.8	VMRT	3D	MD	Abstract	.24
124		P	66	11.4	VMRT	3D	MD	Abstract	.71
125	Jahoda (1980)	P	80	14.0	Other	3D	NMD	Abstract	.74
126		P	80	15.0	Other	3D	NMD	Abstract	.56
127	Jain (2012)	NP	92	6.1	Other	3D	NA	Abstract	.12
128	Jansen, Quaiser-Pohl, Neuburger, and Ruthsatz (2015)	P	50	9.5	Other	3D	MD	Other	.47
129	Jansen, Kellner, and Rieder (2013)	P	44	7.7	Other	3D	MD	Animate	.02
130		P	39	7.7	Other	3D	MD	Animate	.41
131		P	20	7.7	Other	3D	MD	Animate	.52
132	Jones (2010)	NP	95	6.1	Other	3D	NA	Abstract	.53
133	Karádi, Szabó, Szepesi, Kállai, and Kovács (1999)	P	89	9.0	Other	3D	MD	Animate	.03
134	Kaufman (2007)	P	100	17.5	VMRT	3D	MD	Abstract	1.16
135	Kerns and Berenbaum (1991)	P	81	10.2	Other	2D	MD	Abstract	.45
136		P	81	10.2	Other	3D	MD	Abstract	.42
137		P	42	10.8	Other	2D	MD	Abstract	.96
138		P	42	10.8	Other	3D	MD	Abstract	1.21
139	Kirkendall (1987)	NP	30	5.5	Other	3D	MD	Animate	−.21
140	Klein, Adi-Japha, and Hakak-Benizri (2010)	P	80	5.5	Other	3D	MD	Abstract	.11
141	Kovac (1989)	P	58	13.5	Other	3D	NA	Abstract	.03
142	Krüger and Krist (2009)	P	24	5.8	Other	3D	MD	Other	−1.54*
143		P	43	6.0	Other	3D	MD	Animate	−.66
144		P	48	7.1	Other	3D	MD	Animate	.40
145	Kucian et al. (2007)	P	10	9.2	Other	2D	MD	Animate	.02
146		P	10	12.0	Other	2D	MD	Animate	.72
147	Lange-Küttner and Ebersbach (2013)	P	18	6.6	Other	3D	MD	Abstract	.21
148		P	23	7.3	Other	3D	MD	Abstract	.85
149		P	37	8.6	Other	3D	MD	Abstract	.58
150		P	19	9.5	Other	3D	MD	Abstract	.06
151	Lauer and Lourenco (2015)	NP	60	4.5	CMTT	2D	NMD	Abstract	.14
152		NP	60	4.5	Other	2D	MD	Animate	.00
153	Lauer and Lourenco (2016)	P	53	4.3	CMTT	2D	NMD	Abstract	.29
154		P	53	4.3	Other	2D	MD	Animate	.49
155	Lauer, Esposito, and Bauer (2016)	NP	134	6.5	PMA-SR	2D	NMD	Abstract	−.03
156		NP	154	7.5	PMA-SR	2D	NMD	Abstract	.08
157		NP	150	8.5	PMA-SR	2D	NMD	Abstract	−.09
158		NP	146	9.5	PMA-SR	2D	NMD	Abstract	.36
159		NP	71	7.7	PMA-SR	2D	NMD	Abstract	.02
160	Lehmann, Quaiser-Pohl, and Jansen (2014)	P	65	5.0	Other	2D	MD	Animate	−.29
161	LeVasseur (1999)	NP	325	11.5	Other	2D	MD	Abstract	.06
162		NP	45	14.5	Other	2D	MD	Abstract	.47
163		NP	500	11.5	Other	2D	MD	Abstract	.10
164		NP	478	14.5	Other	2D	MD	Abstract	.42
165	Levine et al. (1999)	P	48	4.2	CMTT	2D	NMD	Abstract	.00
166		P	48	4.7	CMTT	2D	NMD	Abstract	.31
167		P	48	5.3	CMTT	2D	NMD	Abstract	.63
168		P	48	5.7	CMTT	2D	NMD	Abstract	.39
169		P	48	6.2	CMTT	2D	NMD	Abstract	.35
170		P	48	6.7	CMTT	2D	NMD	Abstract	.19
171	Levine, Vasilyeva, Lourenco, Newcombe, and Huttenlocher (2005)	P	154	7.5	PMA-SR	2D	NMD	Abstract	−.06
172		P	274	7.5	PMA-SR	2D	NMD	Abstract	.18
173		P	112	7.5	PMA-SR	2D	NMD	Abstract	.44

(*Appendices continue*)

Appendix A (*continued*)

ES No.	Article	Pub. status	*n*	*M* age	Task	Stimulus dim.	Stimulus demands	Stimulus type	*g*
174	Levine, Ratliff, Huttenlocher, and Cannon (2012)	P	53	4.5	CMTT	2D	NMD	Abstract	.69
175	Li, Zhu, and Nuttall (2003)	P	266	16.0	VMRT	3D	MD	Abstract	.51
176	Lizarraga and Ganuza (2003)	P	218	14.0	Other	3D	NMD	Abstract	.03
177	Lütke and Lange-Küttner (2015)	P	100	8.6	Other	3D	MD	Abstract	.28
178	Mann, Sasanuma, Sakuma, and Masaki (1990)	P	121	16.5	VMRT	3D	MD	Abstract	.64
179		P	175	16.5	VMRT	3D	MD	Abstract	.86
180	Meyer and Bendig (1961)	P	100	13.5	PMA-SR	2D	MD	Abstract	.23
181	Miller, Kelly, and Kelly (1988)	P	175	11.0	Other	Both	U	Other	−.12
182	Moè (2009)	P	23	18.0	VMRT	3D	MD	Abstract	.62
183		P	21	18.0	VMRT	3D	MD	Abstract	.52
184		P	26	18.0	VMRT	3D	MD	Abstract	.82
185		P	27	18.0	VMRT	3D	MD	Abstract	.52
186		P	25	18.0	VMRT	3D	MD	Abstract	.90
187		P	30	18.0	VMRT	3D	MD	Abstract	.40
188	Moè (2012)	P	201	15.5	VMRT	3D	MD	Abstract	.89
189	Moè and Pazzaglia (2006)	P	63	17.5	VMRT	3D	MD	Abstract	.27
190		P	68	17.5	VMRT	3D	MD	Abstract	1.13
191		P	66	17.5	VMRT	3D	MD	Abstract	.40
192	Moè and Pazzaglia (2010)	P	120	17.0	VMRT	3D	MD	Abstract	.60
193	Moyer (2003)	NP	108	15.7	Other	3D	NA	Abstract	.46
194	Neuburger, Jansen, Heil, and Quaiser-Pohl (2012)	P	72	10.2	VMRT	3D	MD	Abstract	−.15
195		P	72	10.2	VMRT	3D	MD	Abstract	.59
196		P	72	10.2	VMRT	3D	MD	Abstract	.51
197	Neuburger, Ruthsatz, Jansen, and Quaiser-Pohl (2015)	P	68	9.8	VMRT	3D	MD	Abstract	.88
198		P	68	9.8	VMRT	3D	MD	Abstract	−.13
199	Noda (2008)	P	30	6.8	Other	2D	MD	Abstract	.91
200		P	31	7.8	Other	2D	MD	Abstract	.19
201		P	31	8.8	Other	2D	MD	Abstract	.07
202		P	42	9.7	Other	2D	MD	Abstract	−.22
203		P	29	7.0	Other	2D	MD	Abstract	−.07
204		P	32	7.7	Other	2D	MD	Abstract	.00
205		P	31	8.8	Other	2D	MD	Abstract	.25
206		P	38	9.8	Other	2D	MD	Abstract	−.17
207	Noda (2014)	P	99	4.9	Other	3D	NMD	Other	.52
208	Oostermeijer, Boonen, and Jolles (2014)	P	128	11.7	Other	3D	MD	Animate	.28
209	Patkin and Dayan (2013)	P	22	17.5	Other	3D	MD	Abstract	.98
210	Patrick (1997)	NP	109	15.5	VMRT	3D	MD	Abstract	.48
211	Pazzaglia and Moè (2013)	P	107	17.5	VMRT	3D	MD	Abstract	.51
212	Perner, Kloo, and Rohwer (2010)	P	28	6.7	Other	2D	MD	Animate	−.08
213	Perry (2013)	NP	73	13.5	VMRT	3D	MD	Abstract	−.33
214		NP	121	13.5	VMRT	3D	MD	Abstract	.22
215		NP	77	17.0	VMRT	3D	MD	Abstract	.00
216		NP	72	17.0	VMRT	3D	MD	Abstract	.18
217	Pezaris and Casey (1991)	P	176	13.8	VMRT	3D	MD	Abstract	.20
218	Ping, Ratliff, Hickey, and Levine (2011)	P	20	4.5	CMTT	2D	NMD	Abstract	−.27
219		P	22	4.6	CMTT	2D	NMD	Abstract	.33
220		P	21	4.5	CMTT	2D	NMD	Abstract	.07
221		P	20	4.5	Other	3D	MD	Animate	−.24
222		P	22	4.6	Other	3D	MD	Animate	.38
223		P	21	4.5	Other	3D	MD	Animate	.38
224	Pleet (1991)	NP	150	13.5	Other	2D	MD	Abstract	.30
225		NP	134	13.5	Other	2D	MD	Abstract	−.04
226		NP	276	13.5	Other	2D	MD	Abstract	−.04
227	Prieto and Delgado (1999)	P	60	17.6	Other	3D	MD	Abstract	.90
228		P	60	17.6	Other	3D	MD	Abstract	.82
229		P	60	17.6	Other	3D	MD	Abstract	1.09
230		P	60	17.6	Other	3D	MD	Abstract	.30
231	Quaiser-Pohl, Neuburger, Heil, Jansen, and Schmelter (2014)	P	72	7.9	Other	3D	MD	Animate	.24
232		P	70	7.9	Other	2D	MD	Other	.06

(*Appendices continue*)

Appendix A (*continued*)

ES No.	Article	Pub. status	*n*	*M* age	Task	Stimulus dim.	Stimulus demands	Stimulus type	*g*
233		P	74	10.1	Other	3D	MD	Animate	.39
234		P	72	10.1	Other	2D	MD	Other	.22
235		P	72	7.8	Other	3D	MD	Animate	.13
236		P	72	7.8	Other	2D	MD	Other	.05
237		P	72	9.9	Other	3D	MD	Animate	.17
238		P	72	9.9	Other	2D	MD	Other	.31
239		P	70	7.9	Other	3D	MD	Abstract	.13
240		P	72	10.1	Other	3D	MD	Abstract	.03
241		P	72	7.8	VMRT	3D	MD	Abstract	.06
242		P	72	9.9	VMRT	3D	MD	Abstract	.42
243	Ramirez, Gunderson, Levine, and Beilock (2012)	P	158	7.1	PMA-SR	2D	NMD	Abstract	.04
244	Rosser, Ensing, Glider, and Lane (1984)	P	40	5.0	Other	2D	MD	Abstract	1.26
245	Rosser, Ensing, and Mazzeo (1985)	P	40	4.0	Other	2D	MD	Abstract	.66
246	Rudoff (1983)	NP	40	5.4	Other	2D	MD	Other	−.06
247	Sağlam and Millar (2006)	P	52	17.5	Other	3D	NA	Abstract	.44
248		P	119	18.0	Other	3D	NA	Abstract	.48
249	Scholnick, Fein, and Campbell (1990)	P	48	4.8	Other	Both	MD	Abstract	−.41
250		P	46	5.8	Other	Both	MD	Abstract	.44
251	Sekiyama, Kinoshita, and Soshi (2014)	P	20	6.8	Other	3D	MD	Animate	.61
252		P	18	8.0	Other	3D	MD	Animate	1.31
253		P	16	8.8	Other	3D	MD	Animate	−.09
254		P	15	9.8	Other	3D	MD	Animate	.52
255		P	15	10.9	Other	3D	MD	Animate	−.08
256	Shavalier (1999)	NP	116	10.5	VMRT	3D	MD	Abstract	.33
257	Signorella and Jamison (1978)	P	93	13.5	Other	2D	MD	Abstract	.48
258	Silverman, Phillips, and Silverman (1996)	P	70	14.5	VMRT	3D	MD	Abstract	1.34
259	Singleton (2001)	P	90	6.5	Other	2D	MD	Animate	−.16
260	Sorby, Drummer, Hungwe, Parolini, and Molzan (2006)	P	40	15.5	Other	3D	NA	Abstract	.81
261	Tanzer, Gittler, and Sim (1994)	P	652	14.3	Other	3D	MD	Abstract	.97
262		P	675	15.3	Other	3D	MD	Abstract	.36
263	Tentomas (2010)	NP	50	15.0	Other	3D	NA	Abstract	.58
264	Titze, Jansen, and Heil (2010a)	P	48	9.3	Other	2D	MD	Animate	.19
265		P	48	10.3	Other	2D	MD	Animate	.24
266		P	48	9.3	VMRT	3D	MD	Abstract	.46
267		P	48	10.3	VMRT	3D	MD	Abstract	1.47*
268	Titze, Jansen, and Heil (2010b)	P	84	10.5	VMRT	3D	MD	Abstract	.72
269		P	84	10.5	VMRT	3D	MD	Abstract	.44
270		P	84	10.5	VMRT	3D	MD	Abstract	.25
271	Titze, Jansen, and Heil (2011)	P	126	13.6	VMRT	3D	MD	Abstract	.33
272		P	126	18.0	VMRT	3D	MD	Abstract	.71
273	Tlauka, Williams, and Williamson (2008)	P	40	17.0	Other	2D	MD	Abstract	−.41
274		P	38	17.0	Other	2D	MD	Abstract	.63
275	Toptaş, Çelik, and Karaca (2012)	P	82	13.5	VMRT	3D	MD	Abstract	.02
276	Tracy (1990)	P	282	10.5	Other	2D	U	U	.42
277	Tzuriel and Egozi (2007)	P	64	5.5	Other	2D	MD	Abstract	.66
278		P	64	5.5	PMA-SR	2D	NMD	Abstract	.16
279	Tzuriel and Egozi (2010)	P	60	6.6	Other	2D	MD	Abstract	.54
280		P	60	6.6	Other	2D	MD	Abstract	.69
281		P	56	6.6	Other	2D	MD	Abstract	.94
282		P	56	6.6	Other	2D	MD	Abstract	.48
283		P	60	6.6	PMA-SR	2D	NMD	Abstract	.47
284		P	56	6.6	PMA-SR	2D	NMD	Abstract	.26
285	Vederhus and Krekling (1996)	P	193	9.5	PMA-SR	2D	MD	Abstract	.78
286		P	193	9.5	VMRT	3D	MD	Abstract	.55
287	Weckbacher and Okamoto (2014)	P	113	17.0	VMRT	3D	MD	Abstract	.68
288	Wei et al. (2012)	P	1,556	10.0	Other	3D	MD	Abstract	.16
289	Wiedenbauer and Jansen-Osmann (2008)	P	64	10.7	Other	2D	MD	Animate	.41
290	Xistouri and Pitta-Pantazi (2006)	NP	492	11.0	VMRT	3D	MD	Abstract	.19
291	Yang and Chen (2010)	P	34	10.5	Other	2D	MD	Abstract	1.02
292	Yen (1975)	P	824	14.5	PMA-SR	2D	MD	Abstract	.53

(*Appendices continue*)

Appendix A (*continued*)

ES No.	Article	Pub. status	n	M age	Task	Stimulus dim.	Stimulus demands	Stimulus type	g
293		P	711	15.5	PMA-SR	2D	MD	Abstract	.39
294		P	518	16.5	PMA-SR	2D	MD	Abstract	.44
295		P	425	17.5	PMA-SR	2D	MD	Abstract	.64
296		P	816	14.5	VMRT	3D	MD	Abstract	.91
297		P	691	15.5	VMRT	3D	MD	Abstract	.82
298		P	510	16.5	VMRT	3D	MD	Abstract	.97
299		P	412	17.5	VMRT	3D	MD	Abstract	1.00
300	Young, Cartmill, Levine, and Goldin-Meadow (2014)	P	75	4.8	CMTT	2D	NMD	Abstract	.16
301	Yurt and Sünbül (2012)	P	58	11.5	Other	2D	MD	Abstract	.11
302		P	29	11.5	Other	2D	MD	Abstract	.00
303	Yurt and Sünbül (2014)	P	470	14.5	VMRT	3D	MD	Abstract	.22

Note. P = published; NP = not published; CMTT = Children's Mental Transformation Task (Levine et al., 1999); PMA-SR = Primary Mental Abilities-Space Relations (Thurstone and Thurstone, 1943); VMRT = Vandenberg and Kuse (1978) Mental Rotation Test; MD = mirror discrimination required; NMD = no mirror discrimination required; U = unknown.
* Effect size is an outlier (i.e., $g > M \pm 2.5\ SD$).

Appendix B

Coding Procedure

Age

Mean sample age was coded as a continuous variable. When the mean age was not available but the age range of the sample was provided, mean age was approximated as the midpoint of the age range. When chronological age was not reported but educational grade level was provided, mean age was estimated using the grade level (e.g., the mean age of first graders was approximated as 6.5 years, the mean age of second graders was approximated as 7.5 years).

Country of Testing

We coded the country in which participants completed the mental rotation task as a categorical variable. When this information was not provided in the text, the country of the corresponding authors' address was coded as the country of testing.

Following data extraction, we obtained the Education Index (EI), Gender Development Index (GDI), and Human Development Index (HDI) for each country according to the 2015 Human Development report (United Nations Development Programme, 2015). The EI is a measure of educational achievement that takes into consideration the number of years of schooling completed by adults in the population and the number of years of schooling that young children are expected to complete by adulthood. The GDI is a measure of gender equality that represents the ratio of the HDI index for males relative to females for a given country. GDI values range from 0 to 1, with higher GDI values corresponding to greater gender equality. The HDI represents a composite measure of life expectancy, average educational attainment, and gross national income. HDI values range from 0 to 1, with greater HDI values indicating higher levels of these variables. See the Supplemental Material for analyses related to country of testing.

Gender Centrality

We coded the centrality of gender differences to each article's findings as a dichotomous variable (i.e., gender central, gender not central). Gender differences were coded as central to the findings of an article when references to gender or sex were contained within the article title or journal of publication (e.g., *Sex Roles*), including references to gender differences in performance (e.g., "the male advantage") or gender/sex roles (e.g., "masculine problem-solving strategies"). Gender differences were coded as not central to an article's findings when the article title and journal of publication contained no reference to gender or sex. See the Supplemental Material for analyses on gender centrality.

Publication Status

Publication status was coded as a dichotomous variable (i.e., published or unpublished). Articles were classified as published if they were obtained from (a) a peer-reviewed journal, (b) published conference proceedings, or (c) original research articles in books. The remaining articles, composed of student dissertations and theses, studies presented at conferences, and online research reports provided by authors, were considered to be unpublished work. In cases in which data were presented in an unpublished work (e.g., a dissertation) as well as in a peer-reviewed publication, data were extracted from the article that contained the greater sample size and/or number of age groups.

(*Appendices continue*)

Sample Size

The number of boys and girls in each sample was coded for effect size computation. In cases in which gender composition was not reported for a sample but all other necessary statistics were provided, the gender ratio of the sample was estimated as 1:1.

Stimulus Dimensionality

We coded whether the stimuli were two-dimensional (2D) or three-dimensional (3D). Stimuli were coded as 2D if they consisted of 2D shape or objects that lacked depth cues, and stimuli were coded as 3D if they consisted of 3D objects or 2D renderings of 3D objects that contained visual cues suggesting depth. Stimuli were considered to be 3D if they included depth information. Stimulus dimensionality was not coded in cases in which both 2D and 3D stimuli were presented during the same task ($k = 3$), and the corresponding effect sizes were excluded from analyses on stimulus dimensionality.

Stimulus Discrimination Demands

We coded whether tasks required mirror discrimination or did not require mirror discrimination as a dichotomous variable. Tasks were considered to require mirror discrimination when participants were required to determine whether a target stimulus differed from its rotated mirrored image on at least some of the task's items. Tasks were considered to not require mirror discrimination when participants were required to select from response options that were structurally different from the target stimulus and/or from other response options. This variable was not coded in cases in which all response options consisted of rotated versions of the target stimulus (e.g., Casey et al., 2008; Eraso, 2007). Effect sizes derived from such tasks ($k = 12$) were excluded from analyses examining the effects of stimulus discrimination requirements. Information regarding task requirements was not available for two effect sizes; these effect sizes were also excluded from analyses examining this variable.

Presentation Method

We coded the presentation method of the task used to obtain each effect size as a dichotomous variable, namely whether stimuli were presented on paper or through a digital medium (i.e., on a computer or projector screen). Presentation method was not coded for 20 effect sizes that were produced from tasks that employed physical objects as stimuli ($k = 18$) or presented stimuli through multiple formats ($k = 2$). These 20 effect sizes were not included in analyses in which presentation method was a predictor.

Stimulus Type

We coded stimulus type as a dichotomous variable consisting of two categories: abstract shapes, such as those presented in the PMA-SR and VMRT, and animate stimuli, including animals, human-like characters (e.g., ghosts), or human bodies and/or body parts (e.g., hands). Stimulus type was not coded for 18 effect sizes obtained from tasks with stimuli consisting of alphanumeric symbols ($k = 7$), familiar objects such as cars ($k = 1$), multiple stimulus types ($k = 8$), or unknown stimulus types ($k = 2$). Effect sizes that were not derived from tasks with abstract or animate stimuli were excluded from all analyses in which stimulus type was a predictor.

Task

We coded the name of the mental rotation task as a categorical variable with three levels: CMTT, PMA-SR, and VMRT. Many articles employed modified versions of these three tasks (e.g., fewer trials, altered time constraints), and in these cases, tasks were coded according to the labels provided in the articles. No other mental rotation task was administered in 10 or more articles included in the meta-analysis, so effect sizes obtained using mental rotation tasks other than the CMTT, PMA-SR, and VMRT were excluded from all analyses examining task differences.

Test Setting

We coded test setting as a dichotomous variable (i.e., individual or group). Specifically, we coded whether participants were administered the mental rotation task in individual setting (i.e., one-on-one with an experimenter) or in a group setting in which two or more participants completed the mental rotation task simultaneously in the same room. Information regarding test setting was not available for one effect size, which was excluded from analyses examining the effect of test setting.

Time Constraints

We coded time constraints as a dichotomous variable (i.e., timed or untimed). Tasks were coded as timed when participants were allotted a specific amount of time to complete a single trial, a portion of the task, or the entire task. All other tasks were coded as untimed, including those for which reaction time (RT) data were collected but participants were allowed an unlimited time to respond to each trial. Information regarding time constraints was not available for two effect sizes, which were excluded from analyses of time constraints.

Year of Publication

The year of publication of each article was coded as a continuous variable. For student theses, the year of publication was coded as the date of degree conferral. Moderator analyses that examine the effects of publication year can be found in the Supplemental Material.

Received January 8, 2018
Revision received January 4, 2019
Accepted January 30, 2019 ■