



Effect of Time Constraint in Exploring Spatial Differences With Balanced Allocation of Performance Factors in a Redrawn Mental Rotation Test

Hi-Lian Jeng^{1*} and Yi-Lung Lin²

¹Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology, Taipei, Taiwan, ²Department of Educational Technology, Tamkang University, New Taipei City, Taiwan

OPEN ACCESS

Edited by:

April Lynne Zenisky,
University of Massachusetts Amherst,
United States

Reviewed by:

Nikolay Lutke,
London Metropolitan University,
United Kingdom
Martina Rahe,
University of Koblenz-Landau, Institute
of Psychology, Germany

*Correspondence:

Hi-Lian Jeng
jenghl@mail.ntust.edu.tw

Specialty section:

This article was submitted to
Assessment, Testing and Applied
Measurement,
a section of the journal
Frontiers in Education

Received: 21 May 2021

Accepted: 26 July 2021

Published: 31 August 2021

Citation:

Jeng H-L and Lin Y-L (2021) Effect of
Time Constraint in Exploring Spatial
Differences With Balanced Allocation
of Performance Factors in a Redrawn
Mental Rotation Test.
Front. Educ. 6:712691.
doi: 10.3389/feduc.2021.712691

To achieve a comprehensive and unbiased measurement, a mental rotation test (MRT) (cube form) was redrawn and administered with influential performance factors, namely, time constraint, item type, angular disparity, and rotation/flipping. Item type, angular disparity, and rotation/flipping were systematically balanced into the items of the redrawn Pentomino-MRT, and two time-constraint conditions were randomly assigned to 813 Grade 4 to 6 primary students when administering the test. Children of these ages are of investigative interest because they are at crucial stages of spatial ability development and are at an age where associated gender differences emerge. The study demonstrates that spatial gender differences can be detected in Grade 4, are more marked in Grade 5, and become stable in Grade 6. The importance of time constraint is acknowledged in how and at what grade gender differences emerge under the conditions of the performance factors investigated. In particular, the performance of girls reminds us to focus on their spatial ability development if later STEM-related field participation is of concern.

Keywords: spatial difference, gender difference, mental rotation, spatial ability, primary students of grades 4 to 6, STEM

INTRODUCTION

The under-representation of women in STEM (science, technology, engineering, and mathematics) careers has been a worldwide issue of concern (Ceci and Williams, 2007; Diekmann et al., 2010; Miyake et al., 2010). Women's representation in STEM careers is much lower than that of men, largely due to the decreasing number of women graduating with STEM majors each year, as well as the number of women who are early dropouts from STEM subjects. Different endeavors are being made from a range of perspectives to scaffold an individual's potential onto STEM study paths, especially girls.

A series of longitudinal studies indicate that spatial ability is a promising predictor of attainment in STEM-related fields (Kell et al., 2013; Shea et al., 2001; Stieff and Uttal, 2015; Wai et al., 2009). In following up a group of intellectually talented 13-year-olds, Kell et al. (2013) found that spatial ability provides notable incremental validity for predicting their attainment in creativity and technical innovation 30 years later. Wai et al. (2009) found that STEM professionals have higher spatial ability than non-STEM professionals. Within the current technical era, the importance of spatial ability is becoming more recognized since visual representation is prevailing.

Spatial ability has long been one of the classic structures of intelligence (Spearman, 1904; Thurstone, 1934; Vernon, 1950), although it does not receive as much attention as verbal reasoning and mathematics because the latter are considered the two most important literacies and subjects to learn and people tend to regard spatial ability as innate. However, Arnheim. (1969) claimed that visual thinking is indispensable and is not inferior to the other “higher” thinking and cognitive domains. The question arises as to whether spatial ability is malleable and to what extent. In a meta-analysis of 217 spatial studies, Uttal et al. (2012) found that spatially enriched education can pay substantial dividends in increasing participation in STEM-related fields. The effects of spatial training were shown to be malleable, durable, and transferable across gender, age, and a variety of research designs (such as within/ between/mixed subjects, type of training, and initial level of performance; see Table 4 of Uttal et al., 2012). Overall, the effect sizes of both male and female improvements are the same (0.54 and 0.53), children have the largest effect size compared to adolescents and adults (0.61, 0.44, and 0.44, respectively), and studies using only low-scoring initial level of performance subjects have a larger effect size than studies that do no separate subject levels (0.68 and 0.44). It was shown that spatial training and experiences are promising in promoting an individual’s spatial ability in magnitude, durability, and potential of transferability. Considering that girls in general are reported as lower scoring than boys, it is important for differential gender spatial development to be identified at an earlier stage so that spatial interventions and strategies can be implemented. In a short spatial intervention of a 14-h lesson spanning five weekends for Grades 4 to 6 primary students, both boys and girls improved; however, the increase in the girls’ average score was larger than that of boys, so in the posttest evaluation, the gender gap diminished (Jeng et al., 2016).

Spatial research on Grades 4 to 6 primary students is mainly based on Johnson and Meade (1987), who found that age ten is crucial in spatial development. In a study of 442 Grades 4 to 6 students (average ages 9.67–12.58), Jeng and Liu. (2016) found that girls did not develop mental rotation in a linear and logically progressive trend, which is the case with the other cognitive domains in general. Their results showed that 1) there was no gender difference in Grade 4, 2) a significant advantage for boys started in Grade 5 and increased with age and grade, and finally, 3) the largest gender gap occurred in Grade 6 due to the opposite directions of the growth path for both genders; that is, boys followed a progressively linear path, while girls followed a curvilinear path.

Following up from the finding that spatial gender differences increase with age and grade, the next step is to explore these differences further by considering performance factors which are reported to be influential in identifying differences. In the next sections, spatial studies are introduced with a specific focus on the exploration of mental rotation in children (*Exploring Mental Rotation in Children*) and factors related to performance in mental rotation tests (MRTs) of cubic forms (*Factors Related to Performance in Cube-Mental Rotation in Children*).

Exploring Mental Rotation in Children

Spatial ability is not a unitary concept but, rather, a collection of spatial components (Jeng and Liu, 2016). Various component

categorizations that have been proposed are spatial visualization and spatial orientation by McGee. (1979), spatial perception, mental rotation, and spatial visualization by Linn and Petersen. (1985), and visualization, spatial orientation, and speeded rotation which are the three most notable among the ten factors raised by Lohman. (1988). There seems to have no consensus regarding the definition of spatial ability and its subcomponents. As noted by Jeng and Liu (2016, p.201), “Although different number and names of spatial factors were proposed in these works, some factors with same or similar names may carry different definitions, and some factors with different names may carry similar definitions, which altogether contributed to complicate the constructs of spatial ability and its factors as well.” Until recently, characterizations of visual cognition still continued, for example, Kozhevnikov et al. (2005), Hegarty et al. (2006), and Uttal et al. (2012). Kozhevnikov et al. (2005) distinguished between processes of object properties (such as shape and color) and spatial properties (such as location and spatial relation) in the visual system. Hegarty et al. (2006) classified spatial ability into smaller- and larger-spatial scales. Uttal et al. (2012) proposed a two-dimensional function distinction: intrinsic vs extrinsic information and static vs dynamic task.

Mental rotation is the target of study interest because it is considered an important spatial factor with strong evidence relating it to STEM learning (Newcombe and Frick, 2010). Moreover, among the various spatial factors, the largest gender difference in favor of males is found with mental rotation (Halpern, 1989; Hegarty et al., 2006; Lauer et al., 2019; Linn and Petersen, 1985; Voyer et al., 1995). Mental rotation refers to the ability to mentally rotate two- or three-dimensional objects rapidly and accurately (Linn and Petersen, 1985). The classic item configuration used since the Vandenberg and Kuse (1978) mental rotation test (VMRT) is a 3D cubic form (cube-MRT). However, the cube-MRTs are usually regarded difficult and rarely used on children under age 13 because they cannot be reliably tested (Linn and Petersen, 1985). Even later, Hoyek et al. (2012) concluded the same that the cube aggregates were too difficult for school children. When children of Grades 2 and 4 (mean ages 7.94 and 10.06) were investigated on their use of chronometric cube-MRTs, no gender difference was found because the chronometric nature of the test further increased the task difficulty (Jansen et al., 2013). In order to reduce the difficulty of cube-MRTs for conducting spatial studies on children, cube combinations may be replaced by familiar animal pictures, letter cards, and body parts in the item design of MRTs; however, fewer gender differences were found than the traditional cube-MRTs (Hoyek et al., 2012; Jansen et al., 2013; Jansen-Osmann and Heil, 2007; Krüger and Krist, 2009; Linn and Petersen, 1985; Quaiser-Pohl, 2003; Quaiser-Pohl et al., 2010; Titze et al., 2010; Voyer et al., 1995). For example, in Hoyek et al.’s (2012) study, both the VMRT and 2D MRTs provided evidence of gender differences in 10- and 11-year-old children but not in 7- and 8-year-old children. As a consequence, cube-MRT research on children is limited.

On the other hand, as reported in the meta-analytic review of 128 studies by Lauer et al. (2019), the significant developmental

change in the magnitude of the gender difference can be found between 3 and 17 years of age. There is also some remarkable research exploring the mental rotation performance of much younger children, such as the first-year infants (Lauer and Lourenco, 2016); however, in these cases, extra human assistance, alternative data collection, or measurement technology (e.g., eye tracking) would be required in design and testing. The present study, however, mainly focused on children that can independently perform and be evaluated under a common testing or school environment, using the cube form of MRT. The spatial research involving much younger children is therefore largely constrained from the scope of discussion at the time.

However, some recent studies reported that Grade 4 students (ages 9–10) can be reliably tested with cube-MRTs (e.g., Jeng and Liu, 2016 and Titze et al., 2010), which showed at least 3 years of age advancement in mental rotation development compared to the research of Linn and Petersen. (1985). These studies have made available the target ages and cube-MRT instruments to continue the search for the initial emerging ages and patterns of spatial gender differences. Studies found significant gender differences in Grade 4 students (e.g., Neuburger et al., 2011), especially when the mean age exceeds ten (Titze et al., 2010). When consecutively increasing age/grade groups were included, Jeng and Liu. (2016) found that a gender difference in favor of boys emerged as early as when children can first mentally rotate, which is in Grade 4, although not in a significant manner. This advantage in favor of boys reached statistical significance in Grade 5 and increased by grade and age: the boys showed a persistently upward trajectory of development, while the girls first showed an increasing trajectory from Grades 4 to 5 and later a decreasing path from Grades 5 to 6. It was then suggested that efforts to promote spatial development for girls should not take place later than Grade 5. Overall, these studies support the notion that 10 years is a crucial age for spatial development. Spatial training may benefit students the most if it is delivered prior to or concurrent with taking introductory STEM courses (Stieff and Uttal, 2015). Research shows that students who perform poorly on spatial ability measures are more likely to struggle in entry-level STEM courses and are less likely to enjoy STEM instructions (Wai et al., 2009; Wai et al., 2010).

Factors Related to Performance in the Cube-Mental Rotation Tests in Children

In addition to age, performance factors (or procedural factors as named by Lauer et al., 2019), including task or stimulus characteristics, have been studied to account for spatial gender differences, such as test interactivity (Jeng and Liu, 2016), test time constraints (Jansen et al., 2013; Peters et al., 1995), angular disparity (Cheung et al., 2009), scoring schemes (Voyer, 1997), response formats (Glück and Fabrizi, 2010), task difficulty (Cherney, 2008), and larger versus smaller spatial scales (Hegarty et al., 2006). These studies differ in their perspectives, focusing on a single main independent variable of interest at a time.

It is reported that males and females differ in the amount of time needed to take MRTs and that when enough time is provided, gender differences disappear (Goldstein et al., 1990; Voyer, 1997).

In the meta-analysis of time limits and gender differences on mental rotation, gender differences are significantly larger when the task is administered With time constraints compared to when such constraints are absent (Voyer, 2011). In Peters' (2005) study 1, the gender difference was found to increase with the position of problems when undergraduate students solved two sets of 12 problems under a standard time provision (6 min). Females spent a longer time on the problems and performed more slowly; therefore, they tackled fewer problems in the end. As a result, gender differences were smaller at the problems positioned at the beginning of the task and became larger in the latter positioned problems. The question "What would be the gender difference if time provision is increased?" led to study 2 in which two time provisions (standard vs doubled) were manipulated. In the doubled time provision, the number of problems solved did indeed increase in both genders; however, the magnitude of gender difference was still sizable. Given these findings, Peters commented that it is not yet possible to conclude that the speed of mental rotation is the sole factor accounting for the gender differences, suggesting that the time factor "is a complex undertaking that requires manipulation of more variables than just the time allowed for the test" (p. 182).

Mental rotation involves mental transformation procedures that usually depend on angular disparity. Longer response time increased with increases in angularity (Vandenberg and Kuse, 1978). Some studies explored mental rotation in differing angular disparities. For example, in a study with 5- and 8-year-olds (Marmor, 1975), children were asked whether two stimuli were the same or different in shape by the clockwise rotation of 30°, 60°, 120°, or 150°. For both groups, reaction times increased as a linear function of angular discrepancy between stimuli. Later, Marmor. (1977) found a linear increase in reaction times already in 4- to 5-year-old children when geometric stimuli were used. In Neuburger et al. (2011) study, six rotation angles were used, 45°, 90°, 135°, 225°, 270°, and 315° by a constant increment of 45°, without reporting how the six angles were decided and how or whether they were systematically allocated into the item choices.

Performance is often impaired linearly with increasing angular disparity between two comparison stimuli (Cheung et al., 2009). "A viewpoint cost" is inferred as a linear increase in response time and/or a reduction in accuracy with an increase in angular disparity between two stimuli (Cheung et al., 2009), or similarly "performance costs" inferred by Shepard and Cooper, 1982; Shepard and Metzler, 1971. Images differing by a smaller rotation are more similar to each other compared to images differing by a larger rotation (Lawson and Humphreys, 1996). However, this may not be necessarily the case. An angular disparity of 180° between two images of an object is no more difficult than a smaller disparity (Hayward, 1998; Hayward et al., 2006). The effect of angular disparity is complicated yet has been seldom systematically examined or controlled. Until recently, Lütke and Lange-Küttner (2021) systematically factorial-analyzed and confirmed that the test difficulty was influenced by the angle of rotation for 4-cube aggregates (instead of 10-cube aggregates as in the study) in 4- to 11-year-old children. They also observed the main effects of age showing increase in accuracy and of sex as boys outperformed girls.

As suggested above, a cube-MRT design incorporating and balancing more performance factors is important for a more comprehensive overall picture of spatial differences to be obtained, including gender differences. An innovative scheme for a cube-MRT construction was therefore developed in the present study. The details on how the performance factors were implemented are described in the Methods section (*Methods*). Following recent research, students of primary Grades 4 to 6 were investigated because children of these ages are undergoing critical developmental stages in the transition from concrete operation to formal operation (Piaget, 1970).

METHODS

Operationalization of the Pentomino-Mental Rotation Test

Using Pentomino to Design Mental Rotation Test Items

The MRT used in the study was constructed and drawn using Pentomino. Pentomino has been adopted to design MRT items in some studies. The present study followed Jeng and Liu's (2016) rationale for adopting Pentomino: a combination of any two pieces of Pentomino results in a ten-cube figure, which is the same as the item figures used in the classic VMRT (1978), yet with more variations in the item configuration. The complete set of 12 shapes of Pentomino and example MRT items using Pentomino can be found in Jeng and Liu (2016). The construction of Pentomino-MRT items according to the selected performance variables (item type and angular disparity) is discussed in *Item Type and Allocation of Angular Disparities*. The Pentomino-MRT thus designed adheres to the definition of mental rotation by Linn and Petersen. (1985) and the intrinsic \times static definition by Utall et al. (2012). After the draft Pentomino-MRT was established, the test standardization procedure was initiated where pretesting (*Pretesting and Participants*) was administered to collect response data for item analysis to ensure test quality and suggest further test revision. Once the quality of the Pentomino-MRT was ascertained through pretesting, formal testing (*Formal Testing and Participants*) of the Pentomino-MRT was administered at which time the treatment variable (time constraint) was manipulated into the testing and the personal attributes (gender and grade) were studied.

Item Type

The variable image similarity was manipulated into the Pentomino-MRT by item type: Mirror vs Different types. There were two correct choices and two incorrect choices in the multiple-choice Pentomino-MRT items. The focus was on how to arrange the incorrect choices for each item. The incorrect choices were either mirror images of the target item figure (that is, the Mirror type) or totally different configurations (the Different type). The Mirror type is more difficult than the Different type because more effort and time would be required to distinguish the incorrect choices from the target figure (Boone and Hegarty, 2017). A spatial gender difference is found in Mirror vs Different types, since males and females differ qualitatively in response style (Cheung et al., 2009; Glück and Fabrizi, 2010). The Pentomino-MRT followed the scoring scheme in the VMRT, where a score of 1 was given when both correct choices were selected; otherwise, a score of 0 was given. There were 12 Mirror-

type items (IDs 1–12) and 12 Different-type items (IDs 13–24) in the Pentomino-MRT (**Table 1**), resulting in a total score of 24.

There is a personal characteristic related to the item type: the analytical vs spatial problem-solving strategy (Geiser et al., 2006; Geiser et al., 2008; Linn and Petersen, 1985). The analytic strategy refers to comparing features of stimuli to identify matching characteristics, while the spatial strategy involves active visualization of object rotation (as in the meta-analysis of Lauer et al., 2019). Structurally different targets and choices (as the Different item type in the study) can be easily discernible using the analytic strategy, and more females than males are inclined to using the analytic strategy, which in turn may disadvantage the females' mental rotation development. On the other hand, the Mirror item type requires active mental rotation of objects and more males apply the spatial strategy, resulting in a male advantage in task performance. However, Kozhevnikov et al. (2005) differentiated the holistic vs analytical strategy in a different way that object visualizers encode and process images "holistically," as a single perceptual unit, whereas spatial visualizers generate and process images "analytically," part by part; a person's score on one need not be correlated with the same person's score on the other. Nevertheless, it all implies that people are inclined to apply different strategies.

The above differentiation of strategies mainly belongs to personal characteristic domains and cannot be experimentally manipulated or controlled; the study instead enclosed the above ingredients of strategies in designing the item types of Pentomino-MRT, that is, the object and spatial properties were either balanced equally or controlled constant. According to Lauer et al. (2019), previous meta-analyses did not directly assess Different vs Mirror item types as to 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 discrimination between rotated mirrored images). By incorporating the two item types into MRT design, the present study would be able to disclose this hypothesis.

Allocation of Angular Disparities

As shown in **Table 1**, two smaller angular disparities (40° and 80°) and two larger angular disparities (120° and 160°) were equally allocated to 1) Mirror and Different items; 2) four choices: A, B, C, and D; and 3) Rotation, Flipping, and Rotation and Flipping (that is, the selected figure was either rotated, flipped, or both rotated and flipped by its allocated angular disparity). These angular disparities were selected according to Cheung et al. (2009). An example item is shown in **Table 2** representing the resultant combination and allocation (Different, Rotation, and small angular disparities 40° and 80°).

Pretesting and Participants

The pretest was intended to provide item response data for item analysis. A public middle-sized primary school located in Taipei, a northern Metropolitan of Taiwan, participated in the study via a formal consent inquiry. Following previous research results that age ten is crucial and that primary fourth graders are at an

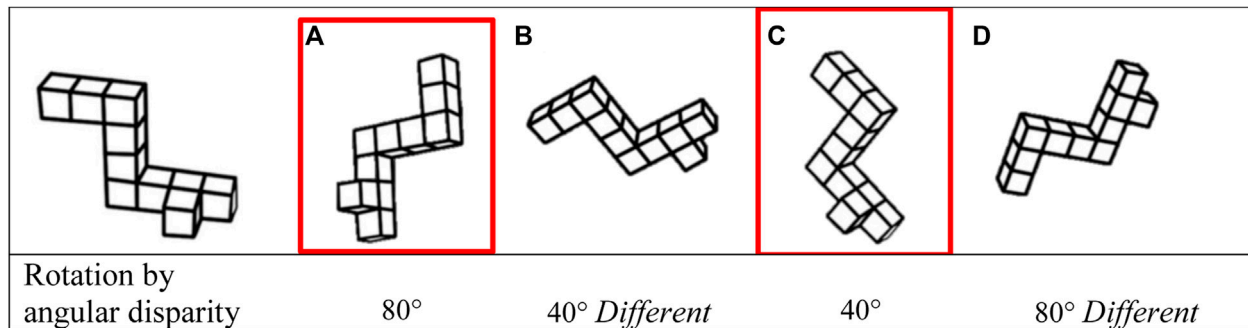
TABLE 1 | Allocation of angular disparities.

Item ID; Mirror/Different	Rotation				Flipping				Rotation and Flipping			
					Choices							
	A	B	C	D	A	B	C	D	A	B	C	D
1/13	40°	80°	40°	80°								
2/14	120°	160°	120°	160°								
3/15	40°	80°	120°	160°								
4/16	40°	80°	120°	160°								
5/17					40°	80°	40°	80°				
6/18					120°	160°	120°	160°				
7/19					40°	80°	120°	160°				
8/20					120°	160°	120°	160°				
9/21									40°	80°	40°	80°
10/22									120°	160°	120°	160°
11/23									40°	80°	120°	160°
12/24									40°	80°	40°	80°
No. of items w/small angular disparities(40°, 80°)			2				2					4
No. of items w/large angular disparities (120°, 160°)			2				4					2
No. of items w/large + small angular disparities (40°, 80°, 120°, 160°)			4				2					2
Total no. of items							24					

¹The systematic allocation plan of angular disparities into item types/ID, Rotation and/or Flipping, and the choices A, B, C, and D as shown in the table is for readers' examination. In actual testing, the order of items and the choices were randomized to prevent a fixed-order effect.

²In some figures, rotating and/or flipping the allocated angular disparity may result in occluded vision. In the case of this occurrence, there is an allowance to adjust ±5° from the allocated angle to increase visibility of the figure.

TABLE 2 | An example Pentomino-MRT item representing the combination and allocation (Different, Rotation, and small angular disparities of 40° and 80°).



Choices A and C are correct answers.

appropriate age for taking the MRT, Grades 4, 5, and 6 were selected for the study. In the school, there were ten classes in each grade, and one class out of ten in each of Grades 4, 5, and 6 was randomly selected for the pretesting, which comprised a sample of three classes (90 students; 43 girls and 47 boys). The remaining nine classes in each grade were reserved for later formal testing.

The pretest was administered Without time constraint conditions in order to collect baseline item statistics. The results of the pretest item analysis are summarized in the following:

- (1) The test internal consistency was 0.87.
- (2) For the Mirror items, the indices of item difficulty ranged between 0.30 and 0.63 and item discrimination ranged between 0.40 and 0.87.
- (3) For the Different items, the indices of item difficulty ranged between 0.30 and 0.92 and item discrimination ranged between 0.17 and 0.74.

- (4) Two Different items (D02 and D06) were in need of minor revision. They were too easy (Difficulty = 0.92) and then had low discrimination (Discrimination = 0.17). Malfunctioning choices were revised. Improvement of this revision can be seen in the summary item analysis for format testing (Table 3).

Formal Testing and Participants

After the two items were revised, the formal version of the Pentomino-MRT was established and administered to the remaining nine classes of students in each of Grades 4, 5, and 6, a total of 27 classes consisting of 814 students. They aged between 8 years 9 months and 12 years 6 months, with a mean of 11 years 2 months. The time constraint conditions were manipulated in this formal testing: With vs Without time constraint. The two conditions were randomly and equally assigned (installed) into the lab computers so that neither the students nor the teachers were aware of the result of student-condition assignment.

TABLE 3 | Summary item analysis of formal testing.

Item type	Item ID	Time constraint			
		With constraint		Without constraint	
		Difficulty	Discrimination	Difficulty	Discrimination
Mirror	M01	0.35	0.39	0.49	0.70
	M02	0.39	0.46	0.49	0.70
	M03	0.27	0.33	0.36	0.64
	M04	0.45	0.51	0.54	0.62
	M05	0.31	0.50	0.46	0.70
	M06	0.33	0.26	0.41	0.44
	M07	0.47	0.70	0.59	0.73
	M08	0.28	0.41	0.41	0.62
	M09	0.39	0.62	0.53	0.74
	M10	0.20	0.29	0.31	0.31
	M11	0.18	0.29	0.39	0.55
	M12	0.14	0.17	0.50	0.75
Average for Mirror		0.31	0.41	0.41	0.63
Different	D01	0.52	0.26	0.82	0.29
	D02	0.64	0.43	0.69	0.43
	D03	0.54	0.35	0.56	0.45
	D04	0.76	0.42	0.80	0.33
	D05	0.79	0.36	0.85	0.21
	D06	0.51	0.60	0.67	0.50
	D07	0.60	0.55	0.66	0.45
	D08	0.62	0.49	0.78	0.30
	D09	0.49	0.51	0.68	0.46
	D10	0.14	0.10	0.26	0.29
	D11	0.26	0.24	0.56	0.36
	D12	0.17	0.17	0.51	0.63
Average for Different		0.53	0.37	0.65	0.39
Average for total		0.42	0.39	0.51	0.53
Internal consistency		0.76		0.82	

1. Based on the Classical Test Theory (Crocker and Algina, 1986), item difficulty (P) is the average of two proportions, answering the item correctly in the higher-scoring and lower-scoring groups, respectively, and item discrimination (D) is the difference of the two proportions. Possible values of P are in the range of $[0, 1]$. Higher values of P represent easier items. Possible values of D are in the range of $[-1, 1]$. Higher positive values of D represent more discriminating items; items with negative D s usually suggest worse item functioning.

2. There were two correct choices per item in the Pentomino-MRT; the total number of choice combinations used for judging the correct answer was then six, AB, AC, AD, BC, BD, and CD in calculating the P s and D s.

As with the pretesting, the test instructions were first read to the students and simultaneously shown on their computer monitors. Particularly in the formal testing, students were informed that there were two time-constraint conditions randomly assigned to them. In the With time constraint condition, the time allowance was 6 minutes (based on Peters, 2005). The Pentomino-MRT would be automatically terminated once the time had elapsed. In the Without time constraint condition, there was no termination time and the students could answer at their own pace. For those students who had terminated or finished the testing, they were required to remain seated and work on their scheduled computer assignments so that they would not disturb other students who were still working on the Pentomino-MRT.

In the primary school, each class period lasts for 45 min. In the Without time constraint condition, students would have about 35 min to work on the Pentomino-MRT, plus 10 min for the initial test instruction and preparation. However, the general observation was that even in the Without time constraint

condition, students finished well before the class period ended. Therefore, the Without time constraint condition as implemented in the primary school can still be regarded as an abundant test time that does not require students to rush. The item analysis results are summarized in **Table 3**. The internal consistency was 0.76 and 0.82, respectively, in the two time-constraint conditions.

RESULTS

In this study, the students selected both Mirror items and Different items in one testing, which is a repeated-measures within-subjects design, while the manipulated factor, time constraint, constitutes a grouping variable to make a between-subject design.

After removing a student with an abnormal response pattern, the remaining number of participants was 813 (384 girls and 429 boys). The descriptive statistics of the Pentomino-MRT scores are

TABLE 4 | Descriptive statistics of test scores.

Time constraint	Grade	N	Mean	Sd	Skewness	Kurtosis
With	4	140	7.61	3.16	0.57	0.42
	5	123	9.74	3.79	0.29	-0.18
	6	142	11.35	4.40	0.43	-0.13
Without	4	132	10.80	4.15	0.57	-0.01
	5	146	14.41	4.91	-0.03	-0.94
	6	130	12.85	5.05	0.02	-0.90
With + Without	4	272	9.16	4.00	0.72	0.47
	5	269	12.28	5.00	0.29	-0.67
	6	272	12.07	4.77	0.24	-0.64

reported in **Table 4**. For the conditions of With time constraint, Without time constraint, and the combined With and Without time constraints, the distributions were normal since all the values of skewness and kurtosis were in the ranges $[-1, 1]$ (Noar, 2003). After normality was ascertained, a three-way analysis of variance was run for Grades 4, 5, and 6, respectively, to investigate the main effects of time constraint, item type, gender, and interaction effects on the test scores. A consistent finding was that for each of Grades 4, 5, and 6, all the three-way and two-way interaction effects were not significant. Only the main effects were significant, which can be summarized as follows:

- 1) Boys scored higher than girls. The direction of the gender difference was in line with most of the spatial literature.
- 2) The Different items scored higher than the Mirror items, indicating that the Different items were easier than the Mirror items.
- 3) The items taken Without time constraint scored higher than those taken With time constraint, indicating that the items taken Without time constraint were easier.

For results (2) and (3), the directions of difference regarding item type and time constraint conformed to the way in which the Pentomino-MRT was designed, providing empirical evidence that the Pentomino-MRT performed as expected. Although the above results of the three main effects emerged in a straightforward and typical manner, subsequent examinations of their simple main effects would reveal further details. The results of the simple main-effect analysis are reported and discussed in *Grade Difference by Time Constraint and Item Type* and *Gender Difference by Item Type, Grade, and Time Constraint*.

DISCUSSION

Grade Difference by Time Constraint and Item Type

As shown in **Table 5**, all the grade differences by time constraint and item type were significant. In the With time constraint condition, the mean scores increased with grade ($6 > 5 > 4$) for both Mirror and Different items, which can be regarded as in line with gradual cognitive development and maturation. In the Without time

constraint condition, however, the Grade 5 students scored either better than Grade 6 with the Mirror items or the same as Grade 6 with the Different items. This implies that the Grade 5 students had the potential to outperform the Grade 6 students on more difficult items (as the Mirror items) and even with easier items (as the Different items), the Grade 5 students scored no worse than the Grade 6 students. The Grade 4 students scored the lowest in all circumstances, aligning with previous study results indicating that Grade 4 is the crucial age of spatial ability development and that, after this age, spatial differences emerge. The most dramatic change and potential period for development is seen in Grade 5, especially with difficult items without a time constraint.

Gender Difference by Item Type, Grade, and Time Constraint

The effect of gender was then added into the analysis. This was inspired by Jeng and Liu. (2016), who found that girls showed different spatial developments compared to boys from Grades 4 to 6. As shown in **Table 6**, the results were mixed and distinctive gender patterns were observed. The effect sizes ranged from 0.13 to 0.53, showing small to marginal-medium effects according to Cohen. (1992). Power analysis with G*Power showed that these effect sizes all have a power ($1-\beta$) of 0.95. In the meta-analysis by Voyer et al. (1995), the effect sizes of the gender difference in mental rotation performance were observed small prior to adolescence ($d = 0.33$) and increased across the teenage years to reach a moderate effect size by adulthood ($d = 0.66$). Therefore, the effect sizes obtained in the study conforms to what were observed previously at these ages of children, yet with more elaboration in terms of time constraint, grade, and item type being considered. In response to the aforementioned hypothesis whether the Different items produced smaller effect sizes than the Mirror items, it can be seen in **Table 6** that there are only three pairs of effect size showing this trend of suppression or superiority: Grade 4 in the With time constraint condition as 0.50 vs 0.25 and Grades 5 and 6 in the Without time constraint condition as 0.53 vs 0.47 and 0.50 vs 0.35, respectively. Obviously, the magnitude of effect size in gender difference does not only depend on the item type; the other considerations are also involved. Furthermore, by referring to the results obtained by Voyer. (2011) that gender differences are significantly larger when the task is administered With time constraints than when such constraints are absent, such comparative results are not obviously replicated in the study. The magnitudes of effect size reported in each segment of gender comparison showed mixed patterns, and again, the reason is the same that more factors are considered in the study. Overall, the average of all the effect sizes in the With time constraint condition (0.32) is even slightly smaller than the average of the effect sizes in the Without time constraint condition (0.42).

In Grade 6, boys scored significantly higher than girls regardless of item type and time constraint (considering $p = 0.051$ as marginally significant), indicating that gender differences were reliably observed at this grade/age. The effect size of Mirror type is larger than that of Different type in Without time constraint condition (0.50 and 0.35).

TABLE 5 | Grade difference by time constraint and item type.

Item type	Time constraint	Grade	N	M	Sd	df	F	p-value	Pairwise post hoc
Mirror	With	4	140	2.35	1.82	2	23.39 ^a	<0.001	6 > 5>4
		5	123	3.37	2.42				
		6	142	4.27	2.76				
	Without	4	132	3.66	2.62	2	17.89 ^a	<0.001	5 > 6>4
		5	146	5.87	3.33				
		6	130	4.82	3.24				
Different	With	4	140	5.26	2.01	2	25.78	<0.001	6 > 5>4
		5	123	6.37	2.13				
		6	142	7.07	2.26				
	Without	4	132	7.14	2.29	2	13.21	<0.001	6 > 4 6 = 5
		5	146	8.54	2.13				
		6	130	8.03	2.42				

^aNonhomogenous F-value using Brown-Forsythe's correction.

TABLE 6 | Gender differences by item type, grade, and time constraint.

Time constraint	Grade	Item type	Gender	N	M	SD	T	df	p	Cohen's d ^b
With	4	Mirror	M	73	2.77	2.01	2.953 ^a	131.16	0.004	0.50
			F	67	1.90	1.46				
		Different	M	73	5.49	2.21	1.469 ^a	135.40	0.144	0.25
			F	67	5.00	1.76				
	5	Mirror	M	61	3.52	2.68	0.719 ^a	114.62	0.474	0.13
			F	62	3.21	2.14				
		Different	M	61	6.54	2.15	0.864	121	0.389	0.16
			F	62	6.21	2.11				
	6	Mirror	M	73	4.84	2.94	2.555 ^a	137.66	0.012	0.43
			F	69	3.68	2.43				
		Different	M	73	7.55	2.26	2.642	140	0.009	0.44
			F	69	6.57	2.17				
Without	4	Mirror	M	69	4.00	2.70	1.571	130	0.119	0.27
			F	63	3.29	2.51				
		Different	M	69	7.58	2.21	2.328	130	0.021	0.41
			F	63	6.67	2.29				
	5	Mirror	M	79	6.66	3.17	3.201	144	0.002	0.53
			F	67	4.94	3.31				
		Different	M	79	8.99	2.10	2.808	144	0.006	0.47
			F	67	8.01	2.07				
	6	Mirror	M	74	5.49	3.32	2.799 ^a	124.78	0.006	0.50
			F	56	3.95	2.94				
		Different	M	74	8.39	2.50	1.972	128	0.051	0.35
			F	56	7.55	2.26				

^aNonhomogenous t-values.

^bCalculated based on studies by Cohen (1992), Rosnow and Rosenthal. (1996), Rosnow et al. (2000), and Thalheimer and Cook (2002). Power analysis with G*Power showed that these effect sizes all have a power (1-β) of 0.95.

however, the effect sizes of the two item types are about the same in the With time constraint condition (0.43 and 0.44).

In Grade 5 and in the Without time constraint condition, there were gender differences in both item types ($p = 0.002, 0.006$),

implying that the gender effect favoring boys was becoming fixed when there was no test time pressure; however, no gender difference was found in the With time constraint condition ($p = 0.474, 0.389$). The test speed (as With time constraint) seemed to contribute to task

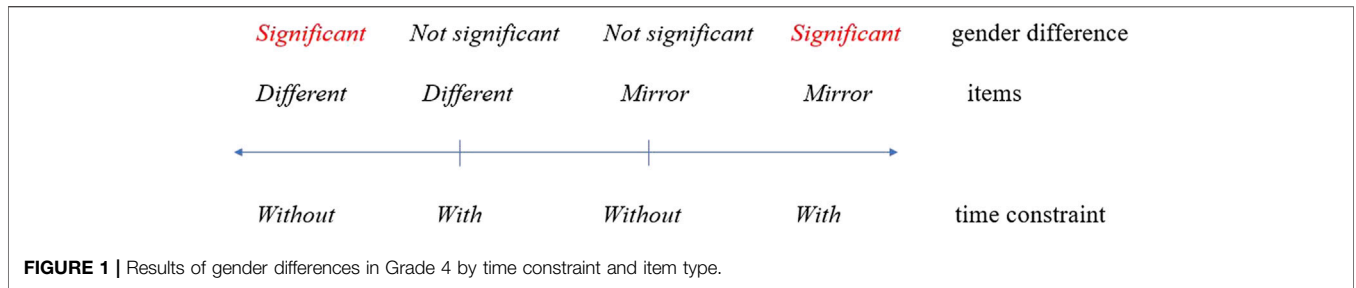


FIGURE 1 | Results of gender differences in Grade 4 by time constraint and item type.

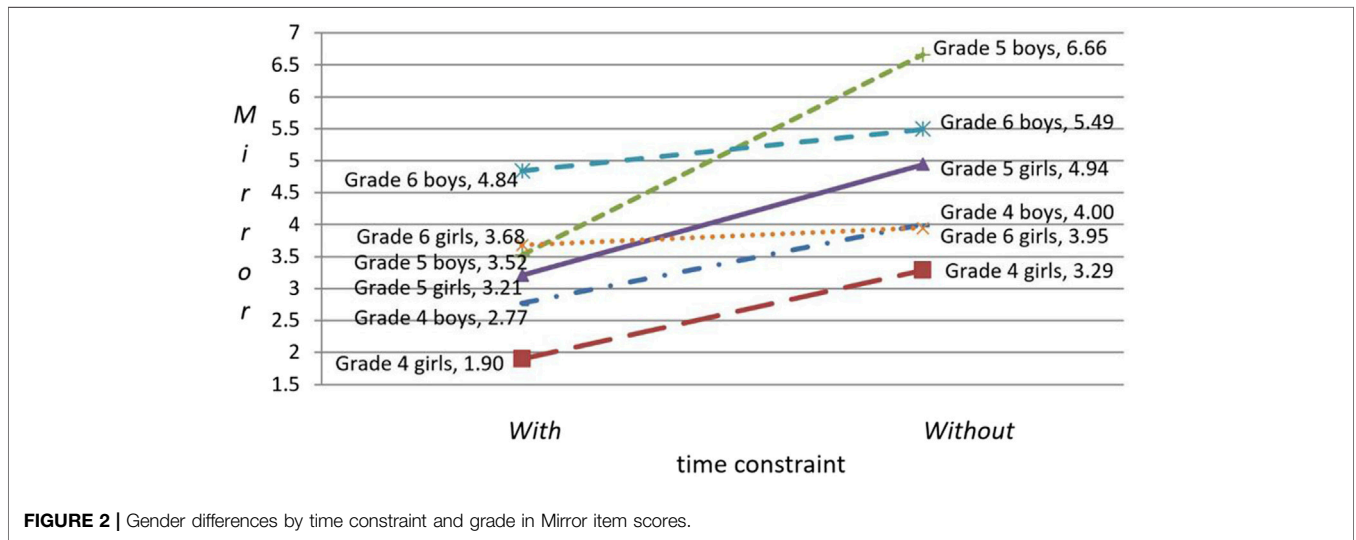


FIGURE 2 | Gender differences by time constraint and grade in Mirror item scores.

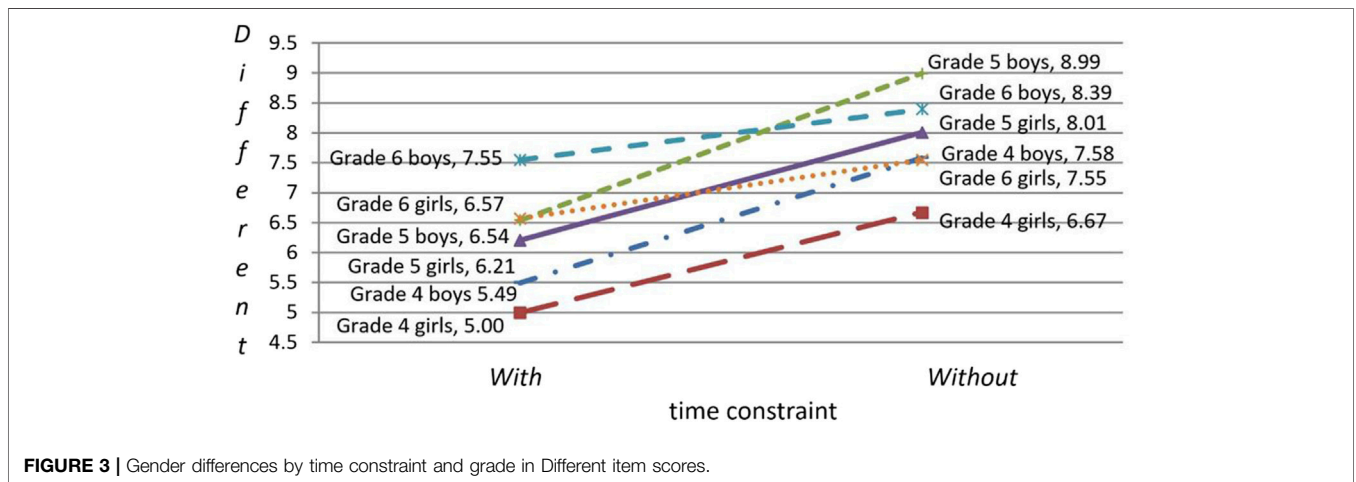


FIGURE 3 | Gender differences by time constraint and grade in Different item scores.

difficulty for both genders so that boys as well as girls experienced rushing regardless of item type and therefore the gender difference disappeared. The magnitudes of effect size reflect reverse directions and large discrepancies in the two pairs of difference, that is, 0.13 and 0.16 in the With time constraint condition and 0.53 and 0.47 in the Without time constraint condition, respectively.

In Grade 4, significant advantages for boys were found 1) With time constraint in taking the Mirror items, which represents extreme difficulty in the study, giving larger effect size (0.50) than its alternative (0.25), and 2) Without time constraint in taking the Different items, which represents the opposite extremity as easier items were taken in no rush at all, giving larger effect (0.41) than

its alternative (0.27). There were no gender differences in-between the two extremities (**Figure 1**).

Figures 2 and **3** provide a visual illustration of the gender differences by time constraint and grade in taking the two item types, respectively. In the With time constraint condition, the means in the Mirror items showed exactly the same ranking as that in the Different items: G6 boys > G6 girls > G5 boys > G5 girls > G4 boys > G4 girls (the left-hand parts of the ranking in **Figures 2, 3**). In the Without time constraint condition, the means again showed the same ranking in both the Mirror and Different items: G5 boys > G6 boys > G5 girls > G4 boys > G6 girls > G4 girls (the right-hand parts of the ranking in **Figures 2, 3**). The visual details help expand our understanding of **Table 5**. That is, in the Without time constraint condition, the Grade 5 students outperformed the Grade 6 students, as revealed by the partial ranking of G5 boys > G6 boys > G4 boys and G5 girls > G6 girls > G4 girls. The spatial performance of the Grade 6 girls is the most concerning result because they scored not only lower than Grade 5 girls but also lower than Grade 4 boys and only scored higher than Grade 4 girls.

CONCLUSION

Previous studies examining gender differences in mental rotation performance during middle childhood produced notable inconsistent results, and evidence of developmental change during middle childhood is also mixed (Lauer et al., 2019). That would usually be the case since different manipulations are involved in experiments when reviewing the aggregate literature such as the meta-analyses. However, the results of male advantage across middle childhood is still majority whilst the increase in male advantage is related to age. In general, the results obtained in the study correspond to recent studies showing that children as young as primary Grade 4 can mentally rotate and be reliably tested on mental rotation; this represents at least 3 years of age advancement in the development of mental rotation compared to an earlier conclusion in the last century that the cube-MRTs were difficult for children under age 13. The results were obtained based on the Pentomino-MRT as operationalized and used in the study, providing evidence that the Pentomino-MRT can still be reliably qualified, in terms of the internal consistency indices obtained and the test standardization procedures undertaken, after incorporating a number of performance factors. The development of the Pentomino-MRT not only represents an innovative perspective on the MRT design but also suggests that influential performance factors of interest can be explored in a similar vein in future research.

Given the refinement in the MRT design and the experimental manipulation as implemented in the present study, the results disclose that time constraint is influential in a number of ways. It is influential in whether Grade 5 students can perform the same as or better than Grade 6 students and whether gender differences can be diminished, although in opposite directions. That is, the former is obtained when time is not constrained and the latter is obtained when time is constrained. These results remind us to attend the perspectives of measuring and interpreting differences

so as to scaffold children's later participation in STEM fields. When the junior Grade 5 students are allowed more time to perform, they have the potential to compete with the elder Grade 6 students, especially on tougher tasks. On the other hand, when time is strictly constrained, providing equally tough challenges for the Grade 5 boys and girls, gender differences are diminished regardless of item type. In Grade 4, intermediate challenging tasks (mixed time constraint and item type) help to diminish gender differences. Unexpectedly, the most difficult extremities yield gender differences in the same way as the easiest extremities (**Figure 1**). It seems that, in the intermediate challenging tasks, the Grade 4 girls regain their confidence and capability to perform and compete with the Grade 4 boys. Rahe and Quaiser-Pohl. (2019) investigated middle- and high-school students and found gender, perceived ability of stereotypically masculine activities, and female gender stereotype beliefs predict mental rotation performance. Tasks such as those in the cube-MRTs would usually be perceived as masculine and therefore hinder girl's interest in involvement and self-beliefs in efficiency and competence. In their study, boys' performance increased during surveyed age from 11 to 20 years while girls' performance remained stable. The participants in the present study were primary Grade 4 to 6 students, and therefore, the gender stereotype can be expected to tune to a certain extent before it is psychologically or physically stable. In the study, spatial gender differences can be detected in Grade 4, are more marked in Grade 5, and become stable in Grade 6. Grades 4 and 5, around the age of ten, are crucial in the life span for the development of spatial potential and for gender differences to diminish. From Grade 6 onwards, spatial development and gender performance become stereotyped. At the time when spatial malleability is possible, interventions and strategies can be expected to help individuals to progress onto STEM trajectories, especially girls. For example, strengthening girls' confidence in more masculine-stereotyped abilities could help reduce the gender difference (Rahe and Quaiser-Pohl, 2019). Inspired by Deikman et al. (2010) that girl's communal orientation was hypothesized to be one of the major factors negatively relating to their pursuit of STEM careers, Jeng et al. (2016) implemented a collaborative and game-based spatial intervention for Grades 4 to 6 primary students and found that the gender difference was diminished in the posttest evaluation.

Gender differences in MRT, usually in the form of male advantage, have been commonly reported previously across a large range of age groups. Although the study is not the first one to explore mental rotation and its gender and grade differences in middle childhood, the study explores this issue in line with Jeng and Liu. (2016), with a large sample of participants and a further refined MRT design and methodological design. The large sample of 813 participants in the consecutive Grades 4 to 6 can be regarded a small succinct population in the school and therefore is able to give us a somewhat comprehensive understanding about the mental rotation development and gender differences of its emergence and continuity, as compared with previous studies that recruited limited cross-sectional age groups. More consecutive age groups and longitudinal studies can be investigated in the future. Previous studies were also limited

by the number of performance factors considered and manipulated. As contended in Hoyek et al. (2012), new ideas and creative experimental paradigms are needed to better understand the reasons and emergence of gender differences in children. It is suggested that more consideration should be given to test construction and application because the gender spatial difference is no simple phenomenon at all, as shown in the study.

In most test applications, performance factors and attributes do interplay simultaneously. There is no way to segregate them like in well-controlled experiments, and it is unrealistic to do so. However, now is a promising time for giving more consideration to test implementation. Although the main findings in the study have been widely reported previously, the study manifests consistent results with previous studies and yet reveals more complex developmental gender courses in a refined MRT design accounting for more performance factors. Ongoing test efforts and facilities are encouraged to assist in exploring differences, enhancing individual's capabilities, and preventing early stereotype and early withdrawal of individuals from potential attainment in the future. As noted in Goldstein et al. (1990), Halpern. (1989), and Jeng and Liu. (2016), whether differences can be found "depends on what, who, when, and how we test." The operationalization of the Pentomino-MRT in the study therefore can be a redrawn exemplar for MRTs and test designs, including other spatial factors.

In conclusion, the results in the study conform to some previous studies, reveal more elaboration in spatial gender differences when more performance factors are considered, and therefore have implications for education and spatial measurement instruments, especially the spatial training that can be intervened to encourage primary students to participate in the STEM fields with more competence and confidence. However, some findings still represent a puzzle and intrigue further following research. The generalization of

the study is also constrained to a certain extent by the geographical coverage of the participants and the spatial factor mental rotation and performance factors as considered in the implementation.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

ETHICS STATEMENT

The studies involving human participants were reviewed and approved by the Research Ethics Committee of Graduate Institute of Digital Learning and Education, National Taiwan University of Science and Technology. Written informed consent from the participant's legal guardian/next of kin was not required to participate in this study in accordance with the national legislation and the institutional requirements at the time of study.

AUTHOR CONTRIBUTIONS

All authors listed have made a substantial, direct, and intellectual contribution to the work and approved it for publication.

FUNDING

This research is granted and funded by the National Science Council (NSC 102-2511-S-011-003) and Ministry of Science and Technology (MOST 109-2511-H-011-001).

REFERENCES

- Arnheim, R. (1969). *Visual Thinking*. Berkeley, CA: University of California Press.
- Boone, A. P., and Hegarty, M. (2017). Sex differences in mental rotation tasks: Not just in the mental rotation process!. *J. Exp. Psychol. Learn. Mem. Cogn.* 43 (7), 1005–1019. doi:10.1037/xlm0000370
- Ceci, S. J., and Williams, W. M. (2007). "Are we moving closer and closer apart? Shared evidence leads to conflicting views," in *Why Aren't There More Women in Science? Top Researchers Debate the Evidence*. Editors S. J. Ceci and W. M. Williams (Washington, DC: American Psychological Association), 213–235.
- Cherney, I. D. (2008). Mom, let me play more computer games: they improve my mental rotation skills. *Sex Roles* 59 (11), 776–786. doi:10.1007/s11199-008-9498-z
- Cheung, O. S., Hayward, W. G., and Gauthier, I. (2009). Dissociating the effects of angular disparity and image similarity in mental rotation and object recognition. *Cognition* 113, 128–133. doi:10.1016/j.cognition.2009.07.008
- Cohen, J. (1992). A power primer. *Psychol. Bull.* 112, 155–159. doi:10.1037/0033-2909.112.1.155
- Crocker, L., and Algina, J. (1986). *Introduction to Classical and Modern Test Theory*. New York: Holt Rinehart & Winston.
- Diekmann, A. B., Brown, E. R., Johnston, A. M., and Clark, E. K. (2010). Seeking Congruity between Goals and Roles. *Psychol. Sci.* 21 (8), 1051–1057. doi:10.1177/0956797610377342
- Geiser, C., Lehmann, W., Corth, M., and Eid, M. (2008). Quantitative and qualitative change in children's mental rotation performance. *Learn. Individual Differences* 18 (4), 419–429. doi:10.1016/j.lindif.2007.09.001
- Geiser, C., Lehmann, W., and Eid, M. (2006). Separating "Rotators" from "Nonrotators" in the Mental Rotations Test: A Multigroup Latent Class Analysis. *Multivariate Behav. Res.* 41 (3), 261–293. doi:10.1207/s15327906mbr4103_2
- Glück, J., and Fabrizii, C. (2010). Gender differences in the mental rotations test are partly explained by response format. *J. Individual Differences* 31 (2), 106–109. doi:10.1027/1614-0001/a000019
- Goldstein, D., Haldane, D., and Mitchell, C. (1990). Sex differences in visual-spatial ability: the role of performance factors. *Mem. Cogn.* 18 (5), 546–550. doi:10.3758/bf03198487
- Halpern, D. F. (1989). The disappearance of cognitive gender differences: what you see depends on where you look. *Am. Psychol.* 44, 1156–1158. doi:10.1037/0003-066x.44.8.1156
- Hayward, W. G. (1998). Effects of outline shape in object recognition. *J. Exp. Psychol. Hum. Perception Perform.* 24 (2), 427–440. doi:10.1037/0096-1523.24.2.427
- Hayward, W. G., Zhou, G., Gauthier, I., and Harris, I. M. (2006). Dissociating viewpoint costs in mental rotation and object recognition. *Psychon. Bull. Rev.* 13 (5), 820–825. doi:10.3758/bf03194003
- Hegarty, M., Montello, D. R., Richardson, A. E., Ishikawa, T., and Lovelace, K. (2006). Spatial abilities at different scales: individual differences in aptitude-test performance and spatial-layout learning. *Intelligence* 34, 151–176. doi:10.1016/j.intell.2005.09.005

- Hoyek, N., Collet, C., Fargier, P., and Guillot, A. (2012). The use of the Vandenberg and Kuse Mental Rotation Test in children. *J. Individual Differences* 33 (1), 62–67. doi:10.1027/1614-0001/a000063
- Jansen, P., Schmelter, A., Quaiser-Pohl, C., Neuburger, S., and Heil, M. (2013). Mental rotation performance in primary school age children: are there gender differences in chronometric tests?. *Cogn. Dev.* 28, 51–62. doi:10.1016/j.cogdev.2012.08.005
- Jansen-Osmann, P., and Heil, M. (2007). Suitable stimuli to obtain (no) gender differences in the speed of cognitive processes involved in mental rotation. *Brain Cogn.* 64 (3), 217–227. doi:10.1016/j.bandc.2007.03.002
- Jeng, H.-L., Chuang, Y.-W., and Lai, W.-Y. (2016). “September 14–16.” “Diminishing gender difference in communal-collaborative learning,” in *The Proceedings of the International Workshop on Technology-Enhanced Collaborative Learning (TECL 2016) in conjunction with the International Conference on Collaboration Technologies* (Kanazawa, Japan: CollabTech 2016 and CRIWG 2016).
- Jeng, H.-L., and Liu, G.-F. (2016). Test interactivity is promising in promoting gender equity in females’ pursuit of STEM careers. *Learn. Individual Differences* 49, 201–208. doi:10.1016/j.lindif.2016.06.018
- Johnson, E. S., and Meade, A. C. (1987). Developmental patterns of spatial ability: an early sex difference. *Child. Dev.* 58, 725–740. doi:10.2307/1130210
- Kell, H. J., Lubinski, D., Benbow, C. P., and Steiger, J. H. (2013). Creativity and Technical Innovation. *Psychol. Sci.* 24 (9), 1831–1836. doi:10.1177/0956797613478615
- Kozhevnikov, M., Kosslyn, S., and Shephard, J. (2005). Spatial versus object visualizers: a new characterization of visual cognitive style. *Mem. Cogn.* 33 (4), 710–726. doi:10.3758/bf03195337
- Krüger, M., and Krist, H. (2009). Imagery and Motor Processes - when Are They Connected? the Mental Rotation of Body Parts in Development. *J. Cogn. Dev.* 10, 239–261. doi:10.1080/15248370903389341
- Lauer, J. E., and Lourenco, S. F. (2016). Spatial processing in infancy predicts both spatial and mathematical aptitude in childhood. *Psychol. Sci.* 27 (10), 1291–1298. doi:10.1177/0956797616665977
- Lauer, J. E., Yhang, E., and Lourenco, S. F. (2019). The development of gender differences in spatial reasoning: A meta-analytic review. *Psychol. Bull.* 145 (6), 537–565. doi:10.1037/bul0000191
- Lawson, R., and Humphreys, G. W. (1996). View specificity in object processing: evidence from picture matching. *J. Exp. Psychol. Hum. Perception Perform.* 22 (2), 395–416. doi:10.1037/0096-1523.22.2.395
- Linn, M. C., and Petersen, A. C. (1985). Emergence and characterization of sex differences in spatial ability: a meta-analysis. *Child. Dev.* 56 (6), 1479–1498. doi:10.2307/1130467
- Lohman, D. F. (1988). “Spatial abilities as traits, processes, and knowledge,” in *Advances in the psychology of human intelligence*. Editor R. J. Sternberg (Hillsdale, NJ: Lawrence Erlbaum Associates), 181–248.
- Lütke, N., and Lange-Küttner, C. (2021). The magical number four in children’s mental rotation of cube aggregates. *Dev. Psychol.* 57 (2), 211–226. doi:10.1037/dev0001139
- Marmor, G. S. (1975). Development of kinetic images: when does the child first represent movement in mental images?. *Cogn. Psychol.* 7 (4), 548–559. doi:10.1016/0010-0285(75)90022-5
- Marmor, G. S. (1977). Mental rotation and number conservation: are they related?. *Dev. Psychol.* 13 (4), 320–325. doi:10.1037//0012-1649.13.4.32010.1037/0012-1649.13.4.320
- McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal, and neurological influences. *Psychol. Bull.* 86 (5), 889–918. doi:10.1037/0033-2909.86.5.889
- Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., and Ito, T. A. (2010). Reducing the gender achievement gap in college science: a classroom study of values affirmation. *Science* 330, 1234–1237. doi:10.1126/science.1195996
- Neuburger, S., Jansen, P., Heil, M., and Quaiser-Pohl, C. (2011). Gender differences in pre-adolescents’ mental-rotation performance: Do they depend on grade and stimulus type?. *Personal. Individual Differences* 50, 1238–1242. doi:10.1016/j.paid.2011.02.017
- Newcombe, N. S., and Frick, A. (2010). Early education for spatial intelligence: why, what, and how. *Mind, Brain Edu.* 4 (3), 102–111. doi:10.1111/j.1751-228x.2010.01089.x
- Noar, S. M. (2003). The role of structural equation modeling in scale development. *Struct. Equation Model. A Multidisciplinary J.* 10 (4), 622–647. doi:10.1207/s15328007sem1004_8
- Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., and Richardson, C. (1995). A Redrawn Vandenberg and Kuse Mental Rotations Test - Different Versions and Factors that Affect Performance. *Brain Cogn.* 28, 39–58. doi:10.1006/brcg.1995.1032
- Peters, M. (2005). Sex differences and the factor of time in solving Vandenberg and Kuse mental rotation problems. *Brain Cogn.* 57 (2), 176–184. doi:10.1016/j.bandc.2004.08.052
- Piaget, J. (1970). ““Piaget’s theory,”” in *Handbook of Child Psychology*. Editor P. Mussen 3rd ed. (New York: Wiley). doi:10.7312/piag91272
- Quaiser-Pohl, C., Rohe, A. M., and Amberger, T. (2010). The Solution Strategy as an Indicator of the Developmental Stage of Preschool Children’s Mental-Rotation Ability. *J. Individual Differences* 31 (2), 95–100. doi:10.1027/1614-0001/a000017
- Quaiser-Pohl, C. (2003). The Mental Cutting Test “Schnitte” and the Picture Rotation Test—Two New Measures to Assess Spatial Ability. *Int. J. Test.* 3, 219–231. doi:10.1207/s15327574ijt0303_2
- Rahe, M., and Quaiser-Pohl, C. (2019). Mental-rotation performance in middle and high-school age: influence of stimulus material, gender stereotype beliefs, and perceived ability of gendered activities. *J. Cogn. Psychol.* 31 (5–6), 594–604. doi:10.1080/20445911.2019.1649265
- Rosnow, R. L., and Rosenthal, R. (1996). Computing contrasts, effect sizes, and counternulls on other people’s published data: General procedures for research consumers. *Psychol. Methods* 1, 331–340. doi:10.1037/1082-989x.1.4.331
- Rosnow, R. L., Rosenthal, R., and Rubin, D. B. (2000). Contrasts and correlations in effect-size estimation. *Psychol. Sci.* 11, 446–453. doi:10.1111/1467-9280.00287
- Shea, D. L., Lubinski, D., and Benbow, C. P. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: a 20-year longitudinal study. *J. Educ. Psychol.* 93, 604–614. doi:10.1037/0022-0663.93.3.604
- Shepard, R. N., and Cooper, J. A. (1982). *Mental Images and Their Transformations*. Cambridge, MA: MIT Press.
- Shepard, R. N., and Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science* 171, 3972701–3972703. doi:10.1126/science.171.3972.701
- Spearman, C. (1904). “General Intelligence,” Objectively Determined and Measured. *Am. J. Psychol.* 15, 201–292. doi:10.2307/1412107
- Stieff, M., and Uttal, D. (2015). How much can spatial training improve STEM achievement?. *Educ. Psychol. Rev.* 27, 607–615. doi:10.1007/s10648-015-9304-8
- Thalheimer, W., and Cook, S. (2002). *How to calculate effect sizes from published research articles: A simplified methodology*. Retrieved July 09, 2021 from http://work-learning.com/effect_sizes.htm.
- Thurstone, L. L. (1934). The vectors of mind. *Psychol. Rev.* 41, 1–32. Address of the president before the American Psychological Association, Chicago meeting, September 1933. doi:10.1037/h0075959
- Titze, C., Jansen, P., and Heil, M. (2010). Mental rotation performance in fourth graders: no effects of gender beliefs (yet?). *Learn. Individual Differences* 20 (5), 459–463. doi:10.1016/j.lindif.2010.04.003
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., et al. (2012). The malleability of spatial skills: a meta-analysis of training studies. *Psychol. Bull.* 139 (2), 352–402. doi:10.1037/a0028446
- Vandenberg, S. G., and Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Percept Mot. Skills* 47 (2), 599–604. doi:10.2466/pms.1978.47.2.599
- Vernon, P. E. (1950). *The Structure of Human Abilities*. London: Methuen.
- Voyer, D. (1997). Scoring procedure, performance factors, and magnitude of sex differences in spatial performance. *Am. J. Psychol.* 110 (2), 259–276. doi:10.2307/1423717
- Voyer, D. (2011). Time limits and gender differences on paper-and-pencil tests of mental rotation: a meta-analysis. *Psychon. Bull. Rev.* 18 (2), 267–277. doi:10.3758/s13423-010-0042-0
- Voyer, D., Voyer, S., and Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: a meta-analysis and consideration of critical variables. *Psychol. Bull.* 117 (2), 250–270. doi:10.1037/0033-2909.117.2.250

- Wai, J., Lubinski, D., and Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *J. Educ. Psychol.* 101, 817–835. doi:10.1037/a0016127
- Wai, J., Lubinski, D., Benbow, C. P., and Steiger, J. H. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM educational dose: a 25-year longitudinal study. *J. Educ. Psychol.* 102, 860–871. doi:10.1037/a0019454

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Jeng and Lin. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.