Chapter 8
The Improvement of Spatial Ability and its Relation to Spatial Training

Yi-Ling Cheng

8.1 Introduction

8.1.1 What Is Spatial Ability?

Spatial ability is the ability to process spatial thinking. The definition of spatial thinking can be found in a recent article published by Newcombe and Shipley (2015). They concluded that findings from neural, cognitive science, and linguistic studies suggest that spatial thinking is using spatial information for "manipulating, constructing, and navigating the physical world" (page 2). Such a thinking process can be characterized as a cognitive ability and thus can be ordered along a continuum scale from low to high levels of spatial thinking. Therefore, a person who possesses a high degree of spatial thinking is a person who has high spatial ability.

The benefit of possessing high spatial ability has been identified in previous studies. Many studies have established that people who possess high spatial ability also have a higher likelihood to be successful in professional careers in science, technology, engineering or math (STEM) fields (see reviews in Levine et al. 2016 or Wai et al. 2009). High spatial ability also plays a critical role for surgeons (Wanzel et al. 2002), dental education (Hegarty et al. 2009) or even being creative (Kell et al. 2013).

These possible benefits have led to continuous attempts to precisely measure, identify high spatial ability and further improve spatial ability to see if training on spatial ability can improve STEM domains (see discussion in Newcombe and Frick 2010; Uttal and Cohen 2012; Mix and Cheng 2012). Through these attempts, a number of spatial abilities were identified and measures of these spatial abilities were developed accordingly. However, several debates exist among the ample

---

Y.-L. Cheng (✉)
Michigan State University, East Lensing, MI, USA
e-mail: chengyil@msu.edu

M.S. Khine (ed.), *Visual-spatial Ability in STEM Education*,
DOI 10.1007/978-3-319-44385-0_8

amount of research on the studies of spatial cognition and spatial abilities. One line of debates is whether there is a single spatial factor or multiple spatial factors existing in spatial abilities. Specifically, this questions whether all spatial tasks measure the same construct or if different spatial tasks measure related but distinct constructs. Another debate in the research is whether spatial training can be generalized to other cognitive abilities. Specifically, to produce near transfer, which is the transfer effect to other untrained spatial abilities task, or to produce far transfer effect to STEM field, which is the transfer effect to another construct.

Although these two debates seem to be directed as two different questions, they are in fact deeply connected. Studies on cognitive training studies have presented how training a shared cognitive skill/component produced the training- transfer effect to structural dissimilar tasks (Karbach and Kray 2009; Schubert et al. 2014). If all spatial tasks measure the same construct, training on one spatial task should be easily transferred to other spatial tasks. On the other hand, if each spatial task measures a different construct, training on one spatial task might not transfer to the spatial tasks measuring different constructs. This mechanism will also further influence the possibility of improving performance in STEM domains. Therefore, the distinctions of spatial abilities are important from both theoretical and practical standpoints.

8.1.2 The Categorization of Spatial Abilities Tasks

Attempts to categorize these spatial tasks have made by many previous studies with different approaches. Linn and Petersen (1985) tried to categorize spatial abilities as spatial perception, mental rotation, and spatial visualization; methodologically they identify these from both psychometric approaches and underlying cognitive processes. Along the same lines, recent work by Uttal et al. (2013) and Newcombe and Shipley (2015) further proposed a two-by-two typology using recent findings from cognitive science research. The two-by-two categorization maps out static/dynamic processes with intrinsic/extrinsic processes to generate four different spatial dimensions (i.e., intrinsic static, extrinsic static, intrinsic dynamic, and extrinsic dynamic). Although this theoretical approach reflects the research from cognitive science nicely, there have not been any studies to test whether this typology verifies the cognitive structure of spatial ability empirically. Several difficulties for categorizing spatial abilities in previous research have been identified. For example, Lohman (1979) has reviewed and discussed extensively the nature of factor analysis on spatial abilities studies and how it might have led to inconclusive results, as different studies have used different spatial tests to refer to the same spatial constructs. A second issue is that the study of the structure of spatial ability has mainly been done with factor analysis. The results of factor analysis has to do with whether a sufficient number of measurement variables were involved in the analysis. For example, Fabrigar et al. (1999) suggested three to five tasks for each factor is adequate to get a decent result. For example, if a researcher is using the two-by-two typology mentioned above with at least three tasks in each category, a total of 12 spatial tasks would need to be included in order to correctly identify spatial categories. Furthermore, the accuracy of categorization of spatial measures is critically related to whether the selected measures have adequate validity and reliability in terms of their representation of the constructs that they intend to measure.

Fundamentally, the validity and reliability of spatial measures are important for identifying the cognitive structure of spatial ability for effective training, especially because increased scores of spatial measures is used as an indication of effective spatial training. The purpose of this chapter is therefore to review spatial tests from a modern psychometric perspective and provide new insights for future training studies. Although many different spatial tasks have been developed over time, the current chapter focuses on the constructs of spatial ability that have been extensively applied for further training studies in previous studies. Also, to help readers to create a mental map of the structure of spatial ability with these spatial ability measures, these tasks will also be discussed within the framework of 2×2 typology and Linn and Petersen's work as well.

In the following paragraphs, the representative spatial measures of the constructs and the validity and reliability of these measures will be discussed first. The validity and reliability of the measure then will be discussed alone with the existing individual differences. This is because any individual differences are important for the purpose of identifying whether the same construct is measured across different populations (e.g., gender or cultural differences). After the analysis of the characteristics of these measurements, the generalizability of its training effect will be discussed and interpreted.

8.2 Overview of the Spatial Tasks in Spatial Training Studies

Among all spatial training studies, several types of spatial measures have been applied more comprehensively in previous training studies. These are: water level task, mental rotation, visual spatial working memory task, and map reading task. Each of these tasks is described in detail in a separate section below. To help the reader understand each task, in the beginning of each section, the test characteristic of the specific spatial test will be explained first. The test development, the item description, the test instruction, and the range of item difficulty will then be described and explained. Next, the validity and reliability in previous studies will be reviewed. The reliability will be discussed first as reliability is prerequisite condition for validity (Thorndike 1997). The primary reliability to be reported here is test-retest reliability, which is the correlation between two testing time points with the same test. If the test-retest reliability was not established by a previous study, Crobach's alpha reliability (internal consistency) or split-half reliability would be provided instead. When the performance on spatial tests is assessed using the judgments of raters, an inter-rater reliability will be identified. For validity, the dimensionality of the test will be identified first. The concurrent and discriminate validity will be discussed after that. Finally, the ways in which improvement of these spatial tasks were used as indications of effective training would be discussed.

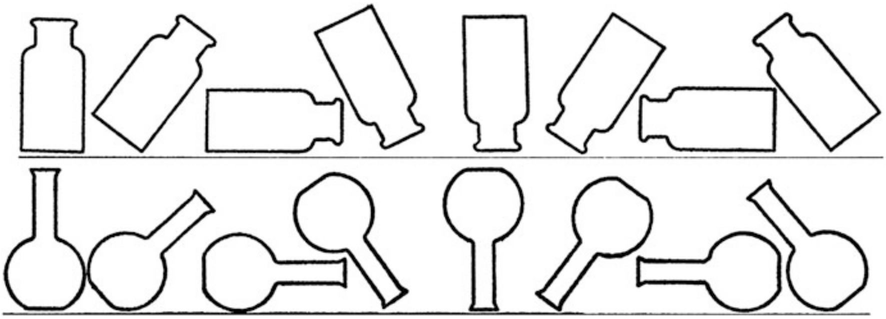

Fig. 8.1 A sample water level task item (Beilin et al. 1966, p. 326)

8.2.1 Water Level Task

The water level task was first developed by Piaget and Inhelder (1956) for understanding children's concept development of the Euclidean reference system. To administer the test, subjects were first shown pictures of titled empty bottles and then were asked to indicate (by drawing) the water level on these titled pictures. To be able to respond to the questions of the water level task correctly, the participants needed to understand and apply the invariant principles (see Fig. 8.1 for the example of the task). The invariance principle is the concept that objects remain unchanged when some transformation is applied to the objects. Specifically, in the water level task, the invariance principle means that the subjects need to understand that the water level should be horizontal despite of the changing angles of the bottles. The item difficulties changed with the different angles of titled bottles such that larger angles were indicated to be harder items (Vasta et al. 1994). Previous studies also have discussed whether the round shaped bottles or square shaped bottles are more difficult (e.g., Liben 1978; Vasta and Liben 1996; Wittig and Allen 1984). While one might hypothesize that round shape of bottles should be easier as it might create less confusions with the horizontal frame of the task, the results from previous studies did not confirm that.

In previous studies, participants were asked to produce two types of responses: reproduction and recognition. For reproduction, the subjects needed to generate the drawing of water level. For recognition, subjects were asked to pick an answer from multiple-choice items (Wittig and Allen 1984). Because recognition might produce a 20–25 % guessing rate (depending on the number of choices given), reproduction of the lines might be more precise for estimating the underlying ability.

8.2.1.1 Validity and Reliability

Wittig and Allen (1984) have computed Spearman-Brown reliability using different versions of water level tasks and identified different reliabilities between males and females. These tasks were categorized with the combination of three response methods (Draw vs. Multiple choice vs. Apparatus (using real bottles)) by two types of bottles (round vs. rectangle). The lowest reliability is found to be with the items that ask for drawing of rectangle bottles. For this condition, reliabilities are 0.78 for males and 0.83 for females, whereas the other conditions have reliabilities ranging from 0.86 to 0.96 across males and females. Other studies using the water level task have also found high reliability using Cronbach's reliability. For example, Cronbach's reliability ranges from 0.80 to 0.86 (Li 2000). These reliabilities also did not differ much between male and female, suggesting an overall good reliability.

In terms of dimensionality of this task, the Rasch item response model was found to fit this task well when using a mixture of 431 children and adults (Formann 2003). Formann (2003) suggested that this implies the unidimensionality of this task. However, Kalichman's (1988) analysis found that there are four sub-abilities: "visual perceptual skills, mental imaging and rotation skills, utilization of spatial coordinate system, and recall of relevant information" (p. 273). These sub-abilities seem to imply the existence of multidimensionality within the water level task. Although these results seem contradictory, it was demonstrated that multidimensional data could also fit well with unidimensional models when items measuring the same composite of abilities are used (Reckase et al. 1988). Therefore, it is possible that the water level task can be a multidimensional task that still fits a unidimensional model well.

In terms of its spatial category, the water level task belongs to the extrinsic static category in Uttal et al.'s typology (2013). Extrinsic static is defined as "understanding abstract spatial principles, such as horizontal invariance or verticality" (Uttal et al. 2013). Specifically, to be able to compute correct responses in the water level task, subjects need to refer to an extrinsic frame. The visual image was not transformed during the task performance, and it only requires static imagination.

However, the discriminant validity of this task has shown some different directions. Some studies showed that the water level task has low correlations with other spatial tasks, but also was indicated to have significant positive correlations with the mental rotation task and embedded figure test in previous studies (Signorella and Jamison 1978). One possibility could be that for generating a correct response, the subjects need to recruit several congitive processes such as transforming the image of the volume of water to align with the reference system. Therefore, evidence also shows that a larger angle of the bottle produces more errors compared to a small angle of the bottle (Vasta et al. 1994) as it would require more efforts. Neural evidence also indicated that people who perform well on this task and the mental rotation task both show similar advances in brain lateralization in the right hemisphere (Rilea et al. 2004), suggesting both tasks are similar in brain lateralization. Perhaps this is also the reason that although the water clock and plumb line tasks both require people to understand the invariance principle, there is little evidence of transferability between the water level and plumb line tasks (e.g., Vasta et al. 1996).

Furthermore, these different sub-abilities might develop at different rates that might be a reason that performances differ in different subgroups (such as males and females or different age groups; Thomas and Turner 1991). For example, while

Pieget indicated that children should develop their understanding of invariance principle around age 9 and therefore perform this task at ceiling, some studies have found that even college students were not able to perform this task well (e.g., Liben and Golbeck 1980; Vasta et al. 1996). There are also gender differences favoring males (e.g., Vasta et al. 1996) and cultural differences were also found favoring Chinese (Li 2000). The gender differences might be likely due to the difference of right hemisphere advances (Rilea 2008). The cultural differences might be due to the possibility that learning Chinese is itself a process of spatial training (e.g., Li et al. 1999). Further studies are needed in studying the measurement invariance of this task across cultures and gender. Overall, these different results might indicate measurement variance in different populations and therefore the interpretation of results in different cultures and genders should be done with caution.

There are several factors might influence the reliability and validity of water level task. For example, the length of water level task is varied across different studies is one of the factors influencing the reliability. Some studies used 12 items (each of them 30° apart) and some used 8 items. Though the item difficulty was varied with different angles, one might wonder whether this is really necessary to assess individual performance of understanding a single principle. For example, in Tran and Formann (2008), the participants' responses were mostly either 0 or 8 across life span (see Fig. 8.2, adapted from table 1 score distribution from their paper). Such a response pattern suggests a possible bimodal distribution of the score range. This distribution might be an indication of discrete attributes. Based on this bimodal distribution, the reliability would therefore be high. Different versions of the water level task might also be confounded with the item difficulty. For example, in the version of square shaped bottles, the titled angles of 90 and 180° should probably be removed as the reference frame is paralleled with the bottom of the bottle, which causes the confusion of interpreting whether subjects answered these two items correctly based on true understanding of invariance principle or they were using the bottom of the bottle as the reference system.

The method of scoring the water level task might also create a threat of its validity. Specifically, researchers often set a certain degree of tolerance level (such as 4 or 5° of deviation) to judge whether the items were correct and add these up to a sum score without further evidence to support why a certain degree was chosen as the cut off criteria. A possible cut off value could be done by having follow-up questions assessing whether participants truly understand the invariance principle and then calculating the deviation degrees only among participants who do understand the principle with computing the standard errors from such deviations. This might provide support for a particular cut off degree for the scoring process. Other than using a sum score, researchers also calculate the exact degree deviation to estimate the performance of subjects (e.g., Vasta et al. 1996). Specifically, by using the deviation of the angle from the horizontal level, the score is estimated with the degree of variation between the participants' response and reference horizontal line. However, considering the possibility that a person with 10° deviation might not have understood more about the invariance principle compared to people had 100°

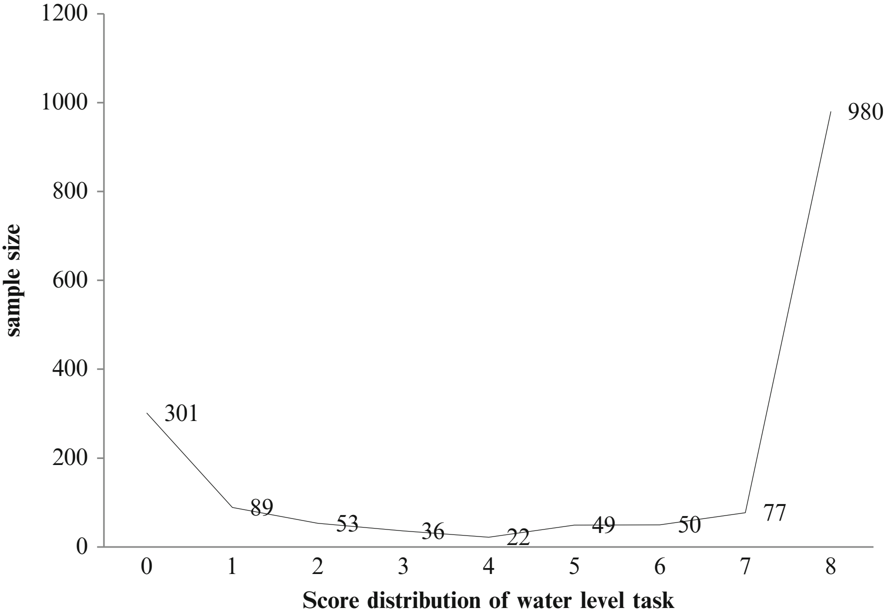

Fig. 8.2 The overall score distribution of water level task (Adapted from Table 1 in Tran and Formann 2008)

of deviation, a summed score might not less differentiate individual level of performances compared to using deviation scores.

8.2.1.2 Training

Improvement

Several approaches have been developed to improve the performance of the water level task and they were all shown effective results. One of the major approaches is using observation (or experience) training (Krekling and Noedvik 1992; Smedslund 1993; Vasta et al. 1996). The results show significant improvements, though Smedslund (1963) noted that observation training has almost no effect on people who did not answer any items correctly in the pretest phase, suggesting the training might not work on people who have no understanding of the invariance concept.

Another approach that might overcome this shortcoming is using instructional approach. By explicitly explaining to the participants what is the invariance principle, the research hopes to see participants understand and apply the concept directly in the task. For example, in Li's (2000) study, the instruction was to remember that no matter how the water bottle was rotated, the water level is always horizontal. The improvement was also significant, though it was difficult to identify whether the participants really understood the principle or simply memorized the principle.

Some individual differences were also found from training effects. For example, there were some age differences revealed in the training results. For example, in Li's (2000) study, the training effect was only significant for 6th and 8th graders but not for 4th, 5th, or 11th graders (the 11th graders were at ceiling with a 97 % correct rate). Perhaps this result also supports that the training has no effect on children (e.g., 4th or 5th grades) who do not have any understanding of the invariance concept. Aside from the age difference, training of the water level task was also found to eliminate pre-existing gender differences (e.g., self-discovery training in Vasta et al. 1996).

8.2.1.3 Transfer

The transfer effect of the water level task is elusive in previous research. First, even though the correct responses of the plumb-line test and the water level task both require the understanding of invariance principles, Vasta et al. (1996) found that the significant improvement from water level training on the water level task did not transfer to Piaget's plumb-line test. This might suggest that although both tasks involve the understanding of invariance principles, they are somehow different. One possibility is that training in water level does not accumulate enough effect to be transferred, or perhaps it is simply the case that water level and plumb line tasks require different cognitive processes. For example, water level might be related to a mental rotation process (e.g., Signorella and Jamison 1978) whereas plumb-line task does not. Secondly, in terms of transfer to academic achievement, there has not been many studies identifying the specific relationship between water level task and STEM achievement. The application of the principle of invariance has been discussed in the process of mathematical problem solving (e.g., Perels et al. 2005) as well as the understanding of multiplication and division (Greer 1994), which might suggest the possibility that training in water level task might improve understanding of these math tasks. In addition, Li et al. (1999) found that SAT scores are highly correlated with the performance in the water level task.

8.2.2 Mental Rotation Test

The very first mental rotation task was developed by Shepard and Metzler (1971), though the most common version people used today was completed by Vandenberg and Kuse (1978), which was adapted from the stimuli from Shepard and Metzler (1971), In the Vandenberg and Kuse's version of the mental rotation task (MRT), 20 three-dimensional mental rotation items were generated with different angles and different lengths of blocks (see Fig. 8.3 for an example item). A redrawn version was also produced by Peters et al. (1995) because the originals of MRT were distorted after many reproductions. The new version consists of 24 items. There are several test characteristics of Vandenberg and Kuse that were similar yet distinct

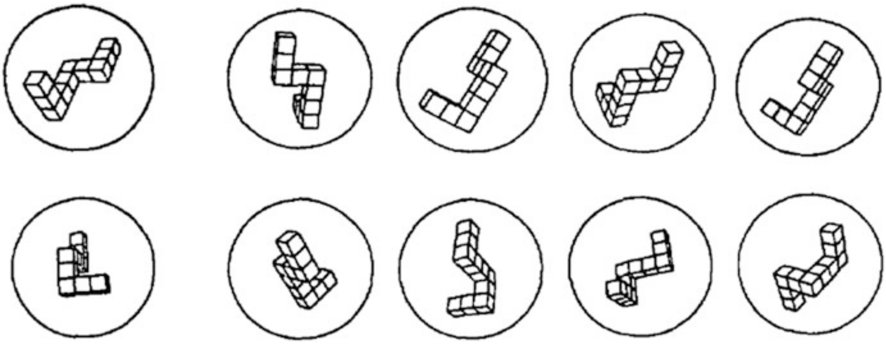

Fig. 8.3 A sample mental rotation task item (Vandenberg and Kuse 1978)

from other spatial tests. First, like the water level task, the item difficulty also is varied by changing the item rotated angles, such that bigger rotated angles result in longer response times (Shepard and Metzler 1971). The most distinct part of MRT is that it has two answers and the test-takers have to answer both targets correctly. According to Vandenberg and Kuse (1978), this is to avoid the possibility of guessing so subjects who only have one correct answer received no credit, although there was no further explanation of how this decreases guessing. By computing a probability of correct rate here, the guessing rate seems to be decreased. For example, using a four foils of multiple choice questions as an instance, the possibility of guessing is decreased from 25 % (1/4) to 17 % (1/2 × 1/3) by asking children to pick two correct answers.

Furthermore, other than the angles of the items changed its difficulty, the types of the distractors (the choices that were not answers) in Vandenberg and Kuse's task are arranged systematically differently. Half of the items have the distractors using the mirror-imaged of rotated targets and half of the items have the distractors using different block configurations from the answer. Because the foils type is also critical for the probability of correctly answering an item, it was hypothesized that that the items have different block configurations as distractors are easier (e.g., Voyer et al. 2004). Voyer and Hou (2006) found the distractors using different block configurations has a high rate of correct responses (Mirrored: 66 %, Structure: 69 %) but the difference between these two different distractors was not significant.

8.2.2.1 Validity and Reliability

Compared to other spatial tests, mental rotation test has been researched broadly in cognitive psychology and neural studies. Previous studies have found that the mental rotation test has proper reliability. For example, the test-retest reliability is 0.83 (Vandenberg and Kuse 1978). For Peters et al. (1995) version, it was found that Cronbach's reliability is 0.87 and split-half reliability is 0.80 (Geiser et al. 2006). Overall, the reliability of MRT is generally good across different versions.

Many studies have suggested that the dimensionality of MRT has a critical influence on participants' performances (Shepard and Metzler 1988). Neural studies also found that 2D and 3D activated different brain areas (Kawamichi et al. 2007; Tagaris et al. 1997). Furthermore, in some studies, mental rotation tasks (such as two dimensional mental rotation task) does not result in male adults advantage in performance (e.g., Rilea et al. 2004). However, fewer studies examined the dimensionality within a single 2D or 3D task. For example, there are some evidences showed that the type of stimulus (hands or blocks) might be influential to the cognitive process recruited of mental rotation tests (Kosslyn et al. 1998), and some showed different result with slight changes on stimulus such as different number of the cubes being used in the blocks (e.g., Bryden et al. 1990). However, when considering the cognitive processes involved in the mental rotation task, multidimensional processing has appeared when the processing of specific items were examined. Studies have found that different cognitive processes were occurred within different items of the 3D task (e.g., Voyer and Hou 2006). Neural researchers have suggested mental rotation involves both analog spatial representation and motor processes (Zacks 2008). All these suggests the potential existence of multidimensionality in mental rotation tasks further research is need to validate its precise structure.

Mental rotation task is considered as an intrinsic dynamic type of spatial task in Uttal et al.'s paper (2013). It is defined as "Piecing together objects into more complex configuration, visualizing and transforming objects" (page 4). Other tasks in the intrinsic dynamic category, such as paper folding and block design, along with mental rotation, seem to be categorized differently by Linn and Petersen (1985). For example, mental rotation is in mental rotation section while paper folding is in spatial visualization in Linn and Petersen (1985). Some evidence suggested that paper folding and mental rotation might be in the same category. For example, Harris et al. (2013) compared and contrasted findings from psychometric and neural studies between mental rotation and paper folding tasks and suggested these two tasks are very similar in many aspects. In Kozhevnikov and Hegarty's (2001) study, they also found that within a confirmatory factor analysis model both tasks loaded significantly on the same latent factor, suggesting at least a significant part of both tasks are loaded on a single dimension.

8.2.2.2 Training

Training on mental rotation task has shown successful results across different populations. Specifically, mental rotation training is effective both on children (De Lisi and Wolford 2002; Ehrlich et al. 2006) and adults (McGee 1978; Terlecki et al. 2008), and also improved the gender difference in favoring males before the training (Neubauer et al. 2010). Furthermore, studies also have shown that improvement of mental rotation scores can be linked to neural efficiency (decreased brain activation; Neubauer et al. 2010). Both virtual and paper versions of mental rotation trainings are shown to be effective as well. For example, a study found that a video training can be effective (De Lisi and Wolford 2002), while another study showed that both video and practice training were effective on improving mental rotation scores (Terlecki et al. 2008).

8.2.2.3 Transfer

Interestingly, not only mental rotation performance can be improved through practicing it, several studies have shown that its transfer effect is quite effective and can be generalized to other similar tasks (Wright et al. 2008; Stransky et al. 2010) or spatial visualization task (Sanz de Acedo Lizarraga and Garcia Ganuza 2003). Furthermore, among these studies, it was found that hands-on activities can be helpful for increasing mental rotation ability (e.g., Jansen et al. 2009; Wiedenbauer and Jansen-Osmann 2008). For example, Wiedenbauer et al. (2008) found that their manual training can lead to improvement on untrained stimulus that usually did not improve through the practice type of training. Perhaps this is linked to the possibility that the strategies used to mental rotation might be connected to motor processes (Wexler et al. 1998). Mental rotation has showed significant relations to STEM achievement (Bruce and Hawes 2015; Stransky et al. 2010; Reuhkala 2001), but the transfer effect to STEM fields is mixed., It was demonstrated that the effect can transfer to improvement on math skills such as missing terms problems (Cheng and Mix 2014) or college introductory physics grades (Miller and Halpern 2013), though some other studies have found no effect (Hawes et al. 2015).

8.2.3 Visual Spatial Working Memory (VSWM)

As one of the subsystems of working memory structure, VSWM has been studied more extensively compared to other spatial abilities. Although an unidimensional view of VSWM defines it as a psychological construct that temporarily holds visual and spatial information (Baddeley 1986, 2000; Quinn 2008; Reuhkala 2001), different subsets of VSWM (e.g., static VSWM and dynamic VSWM) have been used across different studies. (e.g., Corsi 1972; Della Sala et al. 1999). Among these different VSWM tasks, there are several ways to categorize them. For example, one way is to separate these as static VSWM and dynamic VSWM constructs (or they can also be considered as simultaneous and successive VSWM tasks). Between these two categories, one of the representative VSWM tasks, which is developed from the static VSWM construct, is the visual pattern test (Della Sala et al. 1999, see Fig. 8.4 for an example item). The visual pattern test (VPT) is administered by briefly showing a matrix with patterns of blank cells and filled cells (for 3–5 s) and then asking the subjects to indicate what they remember. The score is estimated by counting the number of cells in the given most complex matrix that children can correctly mark (the score range can be from 0 to 15). A different way of score calcualtion is using sum score. See an example in Kaufman ABC test: Spatial Working Memory test; Kaufman and Kaufman 1983) Another representative VSWM task

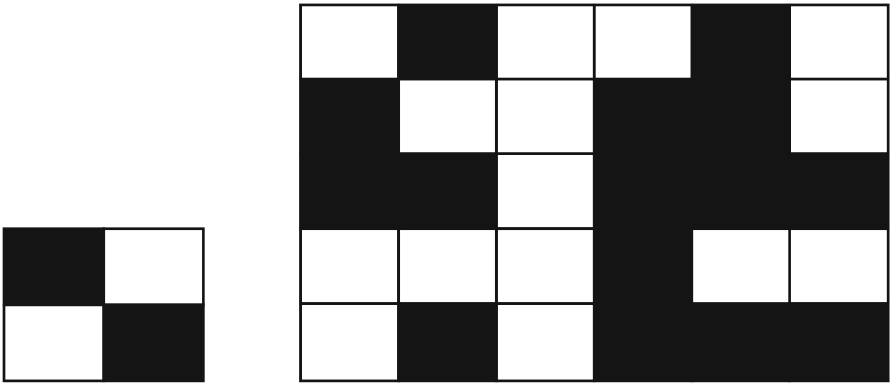

Fig. 8.4 A sample item from visual pattern test (Della Sala et al. 1999)

that is developed from the dynamic VSWM construct is the Corsi block task (Corsi 1972). The Corsi block test measures whether children can successfully reproduce the sequential order that the experimenter tapped on the three dimension blocks.

8.2.3.1 Validity and Reliability

The reliability estimate of VSWM is not always included in studies, though a few studies have provided some of them. For example, for the VPT, the reliability of test-retest is 0.75 on both forms A and B among 50 subjects aged 20–81 years (Della Sala et al. 1999). For the Corsi block test, odd-even reliability (range from 0.70 to ~0.79) is with age 11–16 year old adolescents (Orsini 1994). Friedman et al. (2006) used a spatial two-back task which indicated internal consistency of reliability with a Cronbach's alpha of 0.91. Overall, these reliabilities seem to be adequate among VSWM tests with certain ages, though it is unknown whether similar reliabilities would be found in other populations.

The validity of VSWM construct has been established by showing that a single, independent factor of VSWM can be isolated through confirmatory factor analysis when other working memory tasks were included in the same model (Kane et al. 2004; Miyake et al. 2001). For example, Kane et al. (2004) were able to separate VSWM from general working memory, and the model they proposed also confirmed that one VSWM latent factor can be extracted with three VSWM tasks (rotation span, symmetry span, and navigation span). However, although VSWM is an independent psychological construct, the question remains as to whether multidimensions exist within VSWM.

As mentioned earlier, many studies have developed different VSWM tasks, but they might measure slightly different construct. For example, both VPT and Corsi block tasks require people to remember a brief picture and recall the location of objects presented in the examples. However, one difference is in the procedure: the stimuli of VPT are presented in a simultaneous way and the Corsi block task was presented in a sequential way. Recent neural evidence has shown that these two types of tasks appear to have different patterns of activation in the brain (Darling et al. 2006). There is also behavioral evidence to show that when having people performed dual tasks, visual interference will decrease the performance of VPT task but not on the Corsi task, and spatial interference will decrease the performance of the Corsi task but not on VPT task, which also implies two separate cognitive processes (Della Sala et al. 1999). Therefore, the cognitive processes of static VSWM and dynamic VSWM are suggested to be somehow independent through factor analysis (Vecchi et al. 1995), meaning these processes can be considered two separate dimensions within the concept of VSWM.

There are also studies that examine VSWM more closely across many different VSWM tasks and find that there are more than two sub-dimensions. For example, Mammarella et al. (2008) examined the components of VSWM using confirmatory factor analysis on third and fourth graders. They found that four dimensions coexist within VSWM: sequential-spatial, simultaneous-spatial, visual, and visuospatial active factors. In their study, the division of dimensionality of VSWM seemed to be task specific, such that each factor could be extracted from two or three similar VSWM tasks, suggesting that only several similar tasks shared a common dimension.

Furthermore, in each VSWM task, perhaps the cognitive processes can be hypothesized to reveal more than one dimension. The hypothesis, theorized from Kosslyn's (1983) classical work on mental images, speaks of two types of spatial processing within image generations. One is with a categorical spatial relationship, and the other is with a coordinating spatial relationship. A categorical spatial relationship is the categorizing processing by which people remember the general structure of stimuli, and a coordinating spatial relationship is the location recoding process by which people remember where stimuli are located. This suggests that people understand spatial relationships through two different systems, according to their structure and the exact location of the objects under scrutiny. Many studies in neural imagining have also shown hemispheric specialization for categorical and coordinate relationships: categorical memory has more activation in the left hemisphere and coordinate memory has more activation in the right hemisphere, also suggesting that these are two separate memory systems (e.g., Trojano et al. 2002; van der Ham et al. 2009).

Similarly, Vecchi et al. (1995) found that the amount of information and the structure (the location of the objects) of the VSWM stimuli are two important factors that influence the storage capacity of VSWM. Concluding from these studies, how people remember these changes in structure and how people remember the number of stimuli may be allocated to different dimensions. This multidimensional view of VSWM is not only supported by researchers in behavioral studies (e.g., Logie and Van Der Meulen 2009), but also from neural imagining studies. For example, a recent neural study has demonstrated that there might be two separate cognitive subsystems operating within VSWM (Darling et al. 2006). Overall, these findings imply the possibility that multiple dimensions may exist within in one type of VSWM task.

Few studies, however, have investigated how or whether these dimensions in one type of VSWM (e.g., static VSWM, the basic form of VSWM) can be separated. Based on the separate activations found braining imagining studies, it can be speculated that there may be at least two dimensions in one VSWM task, depending on the location and appearance of the stimuli in the task. However, other researchers have also suggested that the separation between categorical spatial relation and coordinate spatial relation might not be as discrete, as they found that these two tasks show similar activations in their experiments (Martin et al. 2008).

8.2.3.2 Training

As described above, training in VSWM is often included in composite working memory-training programs. Such a composite training program shows improvements on VSWM across different programs, both immediately and in follow ups, but several studies showed a small effect size for immediate improvement (see a meta-analysis in Melby-Lervag and Hulme 2013). Neural imagining studies using VSWM tasks also have demonstrated that the improvement of VSWM performance can be associated with the brain activity in the middle frontal gyrus and superior and inferior parietal cortices. (e.g., Olesen et al. 2004), which suggests the possible connection between VSWM training and brain plasticity.

8.2.3.3 Transfer

Using VSWM training to improve other spatial abilities or academic achievement (such as in mathematics) is prevalent in previous studies. However, such effects have been mixed (Chase and Ericsson 1981; Holmes et al. 2009; Jaeggi et al. 2011; Kyttälä and Kanerva 2014; St Clair-Thompson et al. 2010; Van der Molen et al. 2010). VSWM is found to be trainable even when other cognitive abilities were less effective (e.g., Owen et al. 2010). Even so, the generalizability of VSWM training is inconsistent across measures. For example, in an experiment reported by St Clair-Thompson et al. (2010), children ages 5–8 were trained in composite working memory techniques, including visual image rehearsal strategies, for 6–8 weeks. The training did not improve their performance in the block recall task (a dynamic VSWM that is similar to the Corsi block task) or the digit recall task, but it did improve mental calculation performance. This perhaps suggests that training VSWM to improve math performance is possible, even when the training does not improve other types of VSWM tasks. However, the study itself had a composite training and it was difficult to determine where this effect came from. Furthermore, this specific relationship between VSWM and math might be bidirectional, such that math can improve VSWM as well. For example, Lee et al. (2007) discovered that a year's math training (given once a week) in the use of the mental math abacus (a Chinese math calculation tool) improved the performance of 12-year-old children in simple spatial span tasks, but not in complex spatial span tasks (a task that involves both an equation and a presentation of dots in squares). Overall, training one type of VSWM task does not necessarily improve other types of VSWM tasks, but it might improve some types of math tasks; additionally, training specific math tasks can improve specific types of VSWM. These studies show that VSWM and math have some cross-generalizability, but it is task-specific.

8.2.4 Map Reading

Compared to other spatial tasks, only a handful of tasks have been developed to assess children's potential on map understanding, and they were developed specifically for certain grades. For example, Presson's (1982) map task was designed to test kindergarten and second grade students. Kastens and Liben (2007) adapted Liben and Downs' (1989) flag-sticker field-based map skills test to create a map-reading task for fourth graders. However, across the tasks, the instructions were similar in terms of how subjects were asked to find the location of the targets (see Fig. 8.5). For example, in Kastens and Liben's (2007) map reading task, children were asked to place flags on a map when they were shown an actual flag in a real field area.

The difficulty of the items usually varied by the rotation of the targets (e.g., Presson (1982) or by varying the locations of unique or repeated map symbols, which was recommended by previous research because locations near unique locations might be easier (Kastens and Liben 2007). The scoring scheme is sometimes twofold. For example, in Kastens and Liben's (2007) study, one score was generated from measuring the linear distance from the target to the child's placement (so a longer distance means a higher error rate), and another estimate was generated from categorizing the children's errors to identify children's level of understanding of map concepts.

8.2.4.1 Validity and Reliability

Few studies have examined the map task through the traditional test construction procedure to identify the psychometric properties of the task. While in the previous studies, the development of map understanding has three stages: symbol recognition, metric understanding, and projective view (Liben and Downs 1989), some map reading tasks have identified the development of concept through children's error types (e.g., Kastens and Liben 2007).

Several studies have further differentiated map reading into several different cognitive processes and studied their neural similarity and differences (Shelton and Gabrieli 2002; Shelton and Pippitt 2007; Yamamoto and DeGirolamo 2012). For example, Shelton and Gabrieli (2002) found that the route encoding task and the map overhead reading task recruited similar neural activations on the bilateral fusiform, inferior temporal gyri, and posterior superior parietal cortex. Further, although

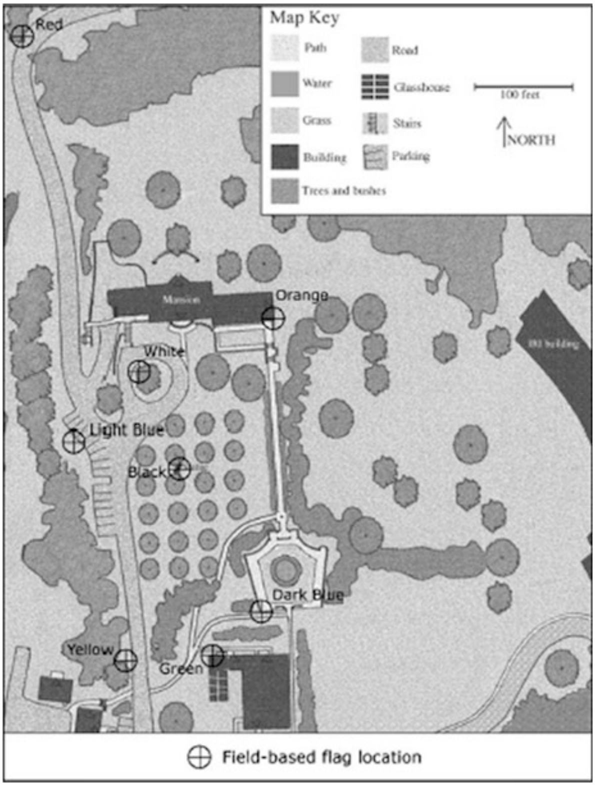

Fig. 8.5 The *black* and *white* version of map reading test (Adapted from Kastens and Liben 2007)

the map overhead reading task recruited greater activation, the route encoding task had a larger area of activation. Specifically, route coding recruited the bilateral medial temporal lobe, postcentral gyrus, right posterior cingulate, and left medial frontal gyrus. Shelton and Gabrieli (2002) considered this as supportive evidence of a hierarchical relationship between route learning and map overhead reading in child development (e.g., Herman and Siegel 1978), which might also suggest multidimensionality within map reading. In terms of spatial categories, while Uttal et al. (2013) categorized map reading task as dynamic and extrinsic, Linn and Petersen (1985) did not specify where map reading tasks should belong.

8.2.4.2 Training

Several studies also further explored the role of instructions in improving children's map reading ability (e.g., Kastens and Liben 2007). For example, Kastens and Liben (2007) compared and contrasted two experimental conditions (ask children to give self explanation on answers vs. no explanation) on a map reading task in fourth graders. Using the distance score, the researchers found that the self-explanation group had significantly less error scores compared to the no explanation group. The authors suggested that perhaps the self-explanation provided a strategy for children to identify and notice the clues in the environment and therefore improved their performances. However, the self explanation group also spent longer time to complete the task (time interval for the activity: 45 mins vs 15–20 mins); perhaps spending enough time on each question can help students to rethink about their answers better and therefore perform better. Overall, the self-explanation group committed fewer errors. The researchers used a distance score procedure, which might be more sensitive, compared to dichotomous responses (0 or 1), and therefore might produce higher sensitivity on identifying children's responses.

Another somewhat different approach used in map reading improvement is decreasing the complexity of map reading tasks by improving subjects' cognitive processing in some components of the task. Map reading contains a lot of different aspects of information processing; therefore, eliminating one aspect of cognitive load might potentially improve the performance is the rationale behind this approach. For example, Fisk and Eboch (1989) adopted automatic processing theory to develop a training protocol. The result suggested that automatic training did improve map reading task performance. However, because the task procedure was only asking the subjects to identify the legend on the maps, which requires symbols reading, there was no item measuring the concepts of spatial orientation or scaling that is often shown in other map reading tasks. Nevertheless, this study demonstrated that training on one component could facilitate learning a multi-components map reading task.

8.2.4.3 Transfer

Few tasks have examined the transfer effect of map reading task (Griffin 1995). Among these, Griffin (1995) examined two approaches (situated learning and traditional learning) in map skills of fourth graders. In the situated group, the students were brought to the environment pictured in their maps while the traditional group received instructions from overhead projectors and writing exercises of map reading skills. The results showed that the situation group performed better than traditional group on the target map performance task but not on the transfer task (which was a novel map task that required the same map skills). The author argued that cognitive skills are context-bound and hard to transfer. A few studies have looked at if other spatial related training might have effects on map reading skills. Klahr and Carver (1988) found that improvement of debugging computer programming skills correlated with a map route following task with a age mixed group including third grade through sixth grade students.

8.3 The Improvment of Spatial Ability.

8.3.1 What is Improved in Spatial Training?

Form the discussion above, one can find that the attempts to improve the performance of spatial abilities through spatial training have gained many successes (also see a meta analysis in Uttal et al. 2013; Wright et al. 2008). It was evidenced that spatial ability is quite sensitive for training effects, and perhaps part of this effect is related to the fact that spatial ability is related to spatial experience (Baenninger and Newcombe 1989). Therefore spatial trainings that simply repeated the procedures to increase spatial experiences have been shown improvement of performances of spatial tasks (e.g., Ehrlich et al. 2006; Cheng and Mix 2014).

Spatial training effects can also transfer across different spatial abilities. Uttal et al. (2013) meta-analyzed previous spatial studies along with the two by two typology categorization that they have developed to calculate the effect size of transfer effects within each category (whether tasks belong to the same category can transfer) and across category (whether tasks belong to different spatial categories can transfer). The results showed that the effect sizes (using Hedges's g) within category ($g=0.51$) and between categories ($g=0.55$) are quite similar. This might suggest that training in any spatial task could possibly improve any other spatial tasks even though they are in different categories here. Although this result seems to imply that spatial ability is a single construct, the possibility that spatial ability might still be multidimensional remains. It is possible that different spatial cognitive processes were merged in different spatial tasks, but all spatial tasks might share similar ingredients of cognitive components with different combinations of processing. Training certain components that were largely shared by several spatial tasks will improve these tasks but not otherwise. For example, all spatial tasks require visual spatial working memory (VSWM) to remember the characteristics of stimuli. Perhaps training VSWM can enhance most of the spatial abilities tasks. Another example is the rotation process. Recall that rotation processes might be recruited during several spatial tasks such as the water level task (e.g., Kalichman 1988), the mental rotation task (Vandenberg and Kuse 1978), and the map-reading task (Presson 1982). It is possible that training rotation processes will improve these spatial tasks more than other spatial tasks. Though most spatial studies showed improvement on spatial abilities, some studies showed less improvement compared to others. Although this outcome has been discussed broadly based on the design of the training such as whether it was within or between subject, or the training methods used such as video, course, or spatial task (for a review, see Uttal et al. 2013), less has been discussed on how these training effects were identified through spatial outcome measures. However, a successful training effect is not only dependent on the effective training design, it is also dependent on the sensitivity of the outcome measures.

The performances of previous spatial tasks were estimated from two main approaches: (1) Using the sum scores from dichotomous items (e.g., VSWM or mental rotation task); (2) Using the degree of error or distance scores (map reading or water level task). Both approaches were able to identify the improved ability from previous studies as they all successfully detected the improvement of spatial performances. However, estimation of degree of error or distance scores might have more advantage compared to a dichotomous item sum score because scores using dichotomous responses have constrained the judgment of participants' performances to a certain degree, which might lose some sensitivity compared to error scores in detecting how far off a participant is from the true target, which might be closer to participants' underlying true abilities. Although all spatial training seems to be effective, this difference might be important for detecting further transfer effects. Furthermore, the approach of using the sum scores to serve as an indication of training gain might have some potential issues as it weights each item equally. What could happen, specifically, is that students answered difficult items correctly might not be weighted more despite of their higher abilities. One way to avoid this problem might be using an extensive range of item difficulty when developing test items.

Secondly, the common issue with using sum scores or sum error scores from single spatial task is that such a score comes with the assumption of measuring the construct perfectly. Therefore the measurement error is to be neglected. Several studies have adapted the latent variable approach to identify the growth of improvement (Embretson 1991; Noack et al. 2014). Such an approach utilizes the latent variable modeling on several variables (that belong to the same constructs) to identify the change of the latent mean and also account for measurement errors. Another measurement issue from previous spatial studies was that certain constructs have to be measured with the format of multiple choices because the attribute might not be able to be measured as an absolute error (or distance) score from the underlying ability (e.g., choose the correct block configuration of the mental rotation task). However, the quality of multiple choice items is also related to the selection of their distractors. The choices of distractors might certainly change the discriminating degree of the items (Haladyna et al. 2002).

Other than the traditional psychometric methods, recent development of neural imaging research has introduced a new perspective to measure the effectiveness of spatial related training (some part of these cognitive training programs). People have used the structures of neural mechanisms to identify the brain areas that are linked to spatial related cognitive trainings and positive effects were found in previous studies. For example, McNab et al. (2009) found that a cognitive training involving spatial memory components is associated with changes in both prefrontal and parietal D1 binding potential, which has been identified to be associated with improved working memory capacity. Raz et al. (2013) found that less brain shrinkage (brain shrinkage happen when we get older) of the cerebellum is associated with cognitive training (spatial working memory involved). Although the results seem to be promising, the above studies involved multiple components (verbal and spatial) in the training process. It is therefore difficult to identify the part of the training associated with this change in brain.

A few studies have further explored the effect from trainings that only have spatial training processes. These studies have also identified positive effects in the brain after the training. For example, in terms of orientation (one type of spatial processing), Wenger et al. (2012) discovered that after spatial navigation training, cortical thickening in left precuneus and paracentral lobule was found with their 20–30 year old participants. Other than identifying whether the positive effect is associated with training, neural approaches also help to identify which different strategies might be involved in spatial tasks. For example, Kosslyn et al. (2001) found that when subjects were asked to imagine rotation with their own hands, this type of imagination activates the motor areas (such as primary motor cortex), whereas the subjects who were asked to imagine the rotation from external sources did not show the same activation. Some studies further examined the relations between item difficulties and brain activity. For example, Ando et al. (2009) found that compared to 10 objects in one stimulus, 30 objects in one stimulus has extended the brain activation in the frontal lobe. These studies might suggest different paths of strategies might be recruited with different numbers of objects and each of these is associated with different degree of brain connectivity.

8.3.2 The g *Factor*

As mentioned in the first section of the chapter, spatial ability has been found to have significant relations with STEM achievement. The directionality of these relations is unclear. However, the relation between STEM and spatial abilities might be discussed more broadly under the relation between academic achievement and cognitive abilities. Many studies have found significant relationships between academic achievement and cognitive abilities (Best et al. 2011; Deary et al. 2007; Gustafsson and Balke 1993; Passolunghi and Lanfranchi 2012; Rohde and Thompson 2007; Spinath et al. 2006). The range of correlations is wide, from 0.40 to 0.90, and at the extreme end, a correlation of 0.90 suggests 80 % of the variance in each is shared. This relation might be because they are both influenced by the same general intelligence factor (*g*). Lynn and Meisenberg (2010) have tested this hypothesis using factor analysis. What they found is that after correcting for unreliability (using the attenuation procedure from Ferguson 1971), that the correlation between the general factor extracted from cognitive tests and the general factor extracted from scores on achievement tests over the course of one school year is 1.0, suggesting one common general ability influences both academic achievements and cognitive abilities. Therefore, scores for an academic achievement (e.g., reading, math, or science) can be viewed as a sub-dimensions under the broader construct of general ability. Follow this logic, it is reasonable to hypothesize the significant relations between STEM achievement and spatial abilities could be because they share a general factor.

8.3.3 Can Spatial Training Improve STEM Achievement?

Because these two constructs might share a general factor, the possibility that training cognitive abilities, specifically spatial abilities, might further improve STEM has drawn attention. For example, previous researchers have argued, perhaps spatial ability is critical for early learning stage of new STEM concept (Uttal and Cohen 2012).

The question of whether spatial training improves STEM achievement actually goes beyond the question itself and can be extended to several levels of inquiries. The first inquiry is whether this predictive relation could also be considered in a reverse direction. Specifically, whether STEM training improves spatial ability or whether such an effect can be reciprocal during the developmental span (Farmer et al. 2013). For example, perhaps learning mathematics itself is a spatial training process and when a person is better at mathematics they are also improved their spatial abilities. The second inquiry, which is often brought by STEM educators, is what the benefits are of doing spatial training when training with STEM contents can directly improve STEM achievements. Indeed, STEM training improves STEM achievement, but these trainings are also context-bound. For example, in math education, Campbell et al. (2006) found training people in addition problems did not improve their performances on subtraction problems. Rickard et al. (1994) also found no transfer of skills between similar division problems (the participants practiced 56/7 = 8 and were tested on 56/8 = 7), suggesting the training effect is specific. Although it is not guaranteed that spatial training will help people transfer their knowledge to mathematics problems, spatial training has showed efficient on transfer effect to other spatial tasks, implying the flexibility of spatial thinking. For example, if we consider the spatial processing from the above mathematics problems, one possible outcome is that training underlying spatial processes might help people to be able to connect and apply the mathematics concepts more flexibly.

Whereas spatial training seems to be promising as a powerful tool for STEM achievement theoretically, spatial training studies did not have many successful cases in improving STEM achievement practically. As many studies have demonstrated the significant relations between spatial ability and STEM (Best et al. 2011; Deary et al. 2007; Gustafsson and Balke 1993; Passolunghi and Lanfranchi 2012; Rohde and Thompson 2007; Spinath et al. 2006), these mixed results from studies cast doubts on the transfer effect of spatial training to STEM. First, there are simply not enough spatial studies to attempt to improve achievement on STEM field to be judged whether this approach is effective. However, here is what is known from current studies—previous spatially related working training programs have showed some improvements (e.g., St Clair-Thompson et al. 2010), when meta-analysis tried to identify the effect of verbal and visual-spatial training, spatial memory training has limited but a convincing effect size (Melby-Lervåg and Hulme 2013). Nevertheless, training and learning a composite cognitive training program might be such a complex process and it would probably lead to learning less about each specific aspect of the task. As a result, individuals may demonstrate less mastery on the trained tasks, let alone improved overall ability. This process is similar to mastering a sport such as basketball. Athletes learn how to move the ball, how to pass a ball, and how to score a basket. The mastery of each specific process is needed to ultimately master basketball as a whole. This step-by-step learning process is likely similar in the brain. Learning one specific cognitive process at a time may help to encourage the thought process needed for a task or may even transfer to other tasks that have applied these specific processes. This type of training might then provide a better outcome than the learning of many different processes at one time.

Secondly, in order for transfer to happen, the underlying processes of transfer mechanism need to be identified. Previously, studies had posited some potential mechanisms behind a successful transfer. Perhaps some processes shared between the trained tasks and improved measures underlie the above generalizable examples. An example is Wallace and Hofelich's (1992) study. In the study they had some participants complete training in mental rotation who were then tested in a geometric analogies task. They had the other group of participants complete training in a geometric analogies task and then tested them in a mental rotation task. Specifically, in the geometric analogies task, the participants were asked to view several shapes, all of which could be combined as an object that still involved one missing piece; they were then asked to also identify which of five options was the missing piece. In the mental rotation task, the participants were asked to identify whether the orientation of objects was standard or mirrored. They found that training in either task improved the performance in both, even though these two tasks did not share similar contexts. The possible mechanism, as explained by Wallace and Hofelich (1992), is Process Theory (Kosslyn 1983), which stresses the importance of similar cognitive processes on task transfer. The argument for this effect is that when two tasks emphasize the same processes, training in either of them should be able to produce a generalized effect on the other. Additionally, this study demonstrated that the improvement between cognitive tasks could be bidirectional.

This phenomenon is not limited to mature adults. Kloo and Perner (2003), for example, found that training children ages 3–4 in an executive control task and a theory of mind task led to improvement on both. Executive control was trained and assessed with a Dimensional Change Card Sorting task (DCCS; Frye et al. 1995), and the theory of mind test was trained for and assessed using false-belief tasks. The improvement of both suggests that the generalization of training effects is possible with children. Kloo and Perner (2003) made a similar argument as Wallace and Hofelic's (1992), suggesting that there may be a specific underlying common factor in these two improved tasks. While these transfer examples showed that transfer effect was possible when the shared processing is defined, there is currently not enough evidence to define the shared processes between STEM and spatial ability, As a result, the question remains as to what is the underlying structure between spatial ability and STEM achievement can be identified and trained to make that transfer happen?

8.3.4 The Cognitive Processes Underlying Spatial Ability and STEM

Identify the underlying cognitive processes among spatial ability and STEM achievement has be developed from several approaches. From the psychometric approach, researchers have been using measurement models to identify whether the same construct is measured across different populations. For example, measurement invariance models were applied and identify the measurement and structural invariance of the construct (e.g., Byrne et al. 1989; van de Schoot et al. 2012). Furthermore, in order to identify the items have discriminating powers on participants' abilities, to estimate the participant's underlying abilities or attribute more accurately, item responses models (Embretson and Reise 2000; Reckase 1997) or diagnostic classification models (e.g., Birenbaum and Tatsuoka 1993) also has been proven to be useful.

Researchers also have worked from the perspective of cognitive science to understand the underlying cognitive process. For example, researchers discovered such a process by doing dual tasks experiments. Specifically, the participants were asked to do one task when another interference task was on the background. By doing so, researchers gained clues about whether these two tasks share the same functional processes in the human brain (e.g., Pashler 1994; McKenzie et al. 2003). Furthermore, recent developments in brain imaging research have extended this approach by determining whether two tasks activate similar brain areas and are therefore using the same processes (e.g., Hubbard et al. 2005) or connect the specific function of the memory tasks with the specific brain areas (see studies in meta analysis of Wager and Smith 2003).

These two approaches have provided many insights on the underlying cognitive processes. However, each of them has its disadvantage. While modeling human responses from a statistical perspective has providing many possible inferences, the results might be rested on relatively better models compared to other models. On the other hands, although neural studies have helped researchers to identify the cognitive processes to the level of item level, building a reliable item analysis from neural studies requires a significant sample size, which might be a reason that different results often come out with different studies. A possible way to solve this problem might be using a large sample with the hypothesized models that were generated from a meta-analysis of previous neural studies. Connecting the item response with neural network might be a new approach although revious studies have demonstrated item diagnostic methods within artificial neural networks (Lamb et al. 2014) While item response models are applied across cognitive psychology studies, few applicated have been made in brain imaging studies. For example, recent studies have used item response models to understand learning (Embretson 1991) and growth change (McArdle et al. 2009), as well as to explore the structure of a cognitive ability such as phonological awareness (Schatschneider et al. 1999). An extension of IRT models, multidimensional item response models (Reckase 2009), was also used to identify the dimensionality of a ninth grade math test (Ackerman et al.

2003) and a test for English language learners (Reckase and Xu 2015). Item response models have been applied extensively in the construction of achievement tests and might be promising to connect with the ample research from neural studies to explore the structure of spatial ability and STEM achievement.

References

Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22*(3), 37–51.

Ando, T., Momose, K., Tanaka, K., & Saito, K. (2009, January). Effects of task difficulty and training of visuospatial working memory task on brain activity. In *13th international conference on Biomedical Engineering* (pp. 657–660). Berlin: Springer.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press.

Baddeley, A. (2000). The episodic buffer: A new component of working memory? *Trends in Cognitive Sciences, 4*(11), 417–423.

Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles, 20*(5–6), 327–344.

Beilin, H., Kagan, J., & Rabinowitz, R. (1966). Effects of verbal and perceptual training on water level representation. *Child Development, 37*, 317–329.

Best, J. R., Miller, P. H., & Naglieri, J. A. (2011). Relations between executive function and academic achievement from ages 5 to 17 in a large, representative national sample. *Learning and Individual Differences, 21*(4), 327–336.

Birenbaum, M., & Tatsuoka, K. K. (1993). Applying an IRT-based cognitive diagnostic model to diagnose students' knowledge states in multiplication and division with exponents. *Applied Measurement in Education, 6*(4), 255–268.

Bruce, C. D., & Hawes, Z. (2015). The role of 2D and 3D mental rotation in mathematics for young children: What is it? Why does it matter? And what can we do about it? *ZDM, 47*(3), 331–343.

Bryden, M. P., George, J., & Inch, R. (1990). Sex differences and the role of figural complexity in determining the rate of mental rotation. *Perceptual and Motor Skills, 70*(2), 467–477.

Byrne, B. M., Shavelson, R. J., & MutheÅLn, B. O. (1989). Testing for equivalence of factor covariance and mean structures: The issue of partial measurement invariance. *Psychological Bulletin, 105*, 456–466.

Campbell, J. I. D., Fuchs-Lacelle, S., & Phenix, T. L. (2006). Identical elements model of arithmetic memory: Extension to addition and subtraction. *Memory & Cognition, 34*, 633–647.

Chase, W. G., & Ericsson, K. A. (1981). Skilled memory. In J. R. Anderson (Ed.), *Cognitive skills and their acquisition* (pp. 141–189). Hillsdale: Lawrence Erlbaum Associates.

Cheng, Y. L., & Mix, K. S. (2014). Spatial training improves children's mathematics ability. *Journal of Cognition and Development, 15*(1), 2–11.

Corsi, P. M. (1972). Human memory and the medial temporal region of the brain (Doctoral dissertation, McGill University, Montreal, Canada). Dissertation Abstracts International, 34 (02), 819B. (University Microfilms No. AA105–77717).

Darling, S., Della Sala, S., Logie, R. H., & Cantagallo, A. (2006). Neuropsychological evidence for separating components of visuo–spatial working memory. *Journal of Neurology, 253*(2), 176–180.

De Lisi, R., & Wolford, J. L. (2002). Improving children's mental rotation accuracy with computer game playing. *The Journal of Genetic Psychology, 163*(3), 272–282.

Deary, I. J., Strand, S., Smith, P., & Fernandes, C. (2007). Intelligence and educational achievement. *Intelligence, 35*(1), 13–21.

Della Sala, S., Gray, C., Baddeley, A., Allamano, N., & Wilson, L. (1999). Pattern span: A tool for unwelding visuo–spatial memory. *Neuropsychologia, 37*(10), 1189–1199.

Ehrlich, S. B., Levine, S. C., & Goldin-Meadow, S. (2006). The importance of gesture in children's spatial reasoning. *Developmental Psychology, 42*(6), 1259.

Embretson, S. E. (1991). A multidimensional latent trait model for measuring learning and change. *Psychometrika, 56*(3), 495–515.

Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah: Erlbaum.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4*(3), 272.

Farmer, G., Verdine, B. N., Lucca, K., Davies, T., Dempsey, R., Newcombe, N., Hirsh-Pasek, K., & Golinkoff, R. M. (2013, April). Putting the pieces together: Spatial skills at age 3 predict to spatial and math performance at age 5. Poster to be presented at the 2013 meeting of The Society for Research in Child Development Conference, Seattle.

Ferguson, G. A. (1971). *Statistical analysis in psychology and education*. New York: McGraw-Hill.

Fisk, A. D., & Eboch, M. (1989). An automatic/controlled processing theory application to training component map reading skills. *Applied Ergonomics, 20*(1), 2–8.

Formann, A. K. (2003). Modeling data from water-level tasks: A test theoretical analysis. *Perceptual and Motor Skills, 96*(3c), 1153–1172.

Friedman, N. P., Miyake, A., Corley, R. P., Young, S. E., DeFries, J. C., & Hewitt, J. K. (2006). Not all executive functions are related to intelligence. *Psychological Science, 17*(2), 172–179.

Frye, D., Zelazo, P., & Palfai, T. (1995). Theory of mind and rule-based reasoning. *Cognitive Development, 10*, 483–527.

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*(3), 261–293.

Greer, B. (1994). Extending the meaning of multiplication and division. In G. Harel & J. Confrey (Eds.), *The development of multiplicative reasoning in the learning of mathematics* (pp. 61–85). Albany: SUNY Press.

Griffin, M. M. (1995). You can' t get there from here: Situated learning transfer, and map skills. *Contemporary Educational Psychology, 20*(1), 65–87.

Gustafsson, J. E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*(4), 407–434.

Haladyna, T. M., Downing, S. M., & Rodriguez, M. C. (2002). A review of multiple-choice item-writing guidelines for classroom assessment. *Applied Measurement in Education, 15*(3), 309–333.

Harris, J., Hirsh-Pasek, K., & Newcombe, N. S. (2013). Understanding spatial transformations: Similarities and differences between mental rotation and mental folding. *Cognitive Processing, 14*(2), 105–115.

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*(3), 60–68.

Hegarty, M., Keehner, M., Khooshabeh, P., & Montello, D. R. (2009). How spatial abilities enhance, and are enhanced by, dental education. *Learning and Individual Differences, 19*(1), 61–70.

Herman, J., & Siegel, A. (1978). The development of spatial representations of large-scale environments. *Journal of Experimental Child Psychology, 26*, 389–406.

Holmes, J., Gathercole, S. E., & Dunning, D. L. (2009). Adaptive training leads to sustained enhancement of poor working memory in children. *Developmental Science, 12*(4), F9–F15.

Hubbard, E. M., Piazza, M., Pinel, P., & Dehaene, S. (2005). Interactions between number and space in parietal cortex. *Nature Reviews Neuroscience, 6*, 435–448.

Jaeggi, S. M., Buschkuehl, M., Jonides, J., & Shah, P. (2011). Short- and long-term benefits of cognitive training. *Proceedings of the National Academy of Sciences of the United States of America, 108*, 10081–10086. doi:10.1073/pnas.1103228108.

Jansen, P., Titze, C., & Heil, M. (2009). The influence of juggling on mental rotation performance. *International Journal of Sport Psychology, 40*(2), 351–359.

Kalichman, S. C. (1988). Individual differences in water-level task performance: A component-skills analysis. *Developmental Review, 8*(3), 273–295.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuo-spatial memory span and reasoning. *Journal of Experimental Psychology: General, 133*(2), 189.

Karbach, J., & Kray, J. (2009). How useful is executive control training? Age differences in near and far transfer of task-switching training. *Developmental Science, 12*(6), 978–990.

Kastens, K. A., & Liben, L. S. (2007). Eliciting self-explanations improves children's performance on a field-based map skills task. *Cognition and Instruction, 25*(1), 45–74.

Kawamichi, H., Kikuchi, Y., Noriuchi, M., Senoo, A., & Ueno, S. (2007). Distinct neural correlates underlying two- and three-dimensional mental rations using three-dimensional objects. *Brain Research, 1144*, 117–126.

Kell, H. J., Lubinski, D., Benbow, C. P., & Steiger, J. H. (2013). Creativity and technical innovation spatial ability's unique role. *Psychological Science, 24*(9), 1831–1836.

Klahr, D., & Carver, S. M. (1988). Cognitive objectives in a LOGO debugging curriculum: Instruction, learning, and transfer. *Cognitive Psychology, 20*(3), 362–404.

Kloo, D., & Perner, J. (2003). Training transfer between card sorting and false belief understanding: Helping children apply conflicting descriptions. *Child Development, 74*, 1823–1839.

Kosslyn, S. M. (1983). *Ghosts in the mind's machine: Creating and using images in the brain*. New York: Norton.

Kosslyn, S. M., Digirolamo, G. J., Thompson, W. L., & Alpert, N. M. (1998). Mental rotation of objects versus hands: Neural mechanisms revealed by positron emission tomography. *Psychophysiology, 35*(02), 151–161.

Kosslyn, S. M., Thompson, W. L., Wraga, M., & Alpert, N. M. (2001). Imagining rotation by endogenous versus exogenous forces: Distinct neural mechanisms. *NeuroReport, 12*(11), 2519–2525.

Kozhevnikov, M., & Hegarty, M. (2001). A dissociation between object manipulation spatial ability and spatial orientation ability. *Memory & Cognition, 29*(5), 745–756.

Krekling, S., & Noedvik, H. (1992). Observational training improves adult women' performance on Piaget's water-level task. *Scandinavian Journal of Psychology, 33*(2), 117–124.

Kyttälä, M., & Kanerva, K. (2014). Training mathematical skills in pre-school children: Specific training of phonological and visuo-spatial working memory.

Lamb, R. L., Annetta, L., Vallett, D. B., & Sadler, T. D. (2014). Cognitive diagnostic like approaches using neural-network analysis of serious educational videogames. *Computers & Education, 70*, 92–104.

Lee, Y., Lu, M., & Ko, H. (2007). Effects of skill training on working memory capacity. *Learning and Instruction, 17*, 336–344. doi:10.1016/j.learninstruc.2007.02.010.

Levine, S. C., Foley, A., Lourenco, S., Ehrlich, S., & Ratliff, K. (2016). Sex differences in spatial cognition: Advancing the conversation. *Wiley Interdisciplinary Reviews: Cognitive Science, 7*, 127–155.

Li, C. (2000). Instruction effect and developmental levels: A study on water-level task with Chinese children ages 9–17. *Contemporary Educational Psychology, 25*(4), 488–498.

Li, C., Nuttall, R. L., & Zhao, S. (1999). The effect of writing Chinese characters on success on the water-level task. *Journal of Cross-Cultural Psychology, 30*(1), 91–105.

Liben, L. S. (1978). Performance on Piagetian spatial tasks as a function of sex, field dependence, and training. *Merrill-Palmer Quarterly of Behavior and Development, 24*(2), 97–110.

Liben, L. S., & Downs, R. M. (1989). Understanding maps as symbols: The development of map concepts in children. *Advances in Child Development and Behavior, 22*, 145–201.

Liben, L. S., & Golbeck, S. L. (1980). Sex differences in performance on Piagetian spatial tasks: Differences in competence or performance? *Child Development, 51*, 594–597.

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

Logie, R. H., & Van Der Meulen, M. (2009). Fragmenting and integrating visuospatial working memory. In *The visual world in memory* (pp. 1–32). Hove: Psychology Press.

Lohman, D. F. (1979). *Spatial ability: A review and reanalysis of the correlational literature* (Stanford University, School of Education, No. TR-8). Stansford: Stanford University.

Lynn, R., & Meisenberg, G. (2010). National IQs calculated and validated for 108 nations. *Intelligence, 38*(4), 353–360.

Mammarella, I. C., Pazzaglia, F., & Cornoldi, C. (2008). Evidence for different components in children's visuospatial working memory. *British Journal of Developmental Psychology, 26*(3), 337–355.

Martin, R., Houssemand, C., Schiltz, C., Burnod, Y., & Alexandre, F. (2008). Is there continuity between categorical and coordinate spatial relations coding? Evidence from a grid/no-grid working memory paradigm. *Neuropsychologia, 46*(2), 576–594.

McArdle, J. J., Grimm, K. J., Hamagami, F., Bowles, R. P., & Meredith, W. (2009). Modeling life-span growth curves of cognition using longitudinal data with multiple samples and changing scales of measurement. *Psychological Methods, 14*(2), 126.

McGee, M. G. (1978). Effects of training and practice on sex differences in mental rotation test scores. *The Journal of Psychology, 100*(1), 87–90.

McKenzie, B., Bull, R., & Gray, C. (2003). The effects of phonological and visual–spatial interference on children's arithmetical performance. *Educational and Child Psychology, 20*, 93–108.

McNab, F., Varrone, A., Farde, L., Jucaite, A., Bystritsky, P., Forssberg, H., & Klingberg, T. (2009). Changes in cortical dopamine D1 receptor binding associated with cognitive training. *Science, 323*(5915), 800–802.

Melby-Lervåg, M., & Hulme, C. (2013). Is working memory training effective? A meta-analytic review. *Developmental Psychology, 49*(2), 270.

Miller, D. I., & Halpern, D. F. (2013). Can spatial training improve long-term outcomes for gifted STEM undergraduates? *Learning and Individual Differences, 26*, 141–152.

Mix, K. S., & Cheng, Y.-L. (2012). The relation between space and math: Developmental and educational implications. *Advances in Child Development and Behavior, 42*, 197–243.

Miyake, A., Friedman, N. P., Rettinger, D. A., Shah, P., & Hegarty, M. (2001). How are visuospatial working memory, executive functioning, and spatial abilities related? A latent-variable analysis. *Journal of Experimental Psychology: General, 130*(4), 621.

Neubauer, A. C., Bergner, S., & Schatz, M. (2010). Two-vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation. *Intelligence, 38*(5), 529–539.

Newcombe, N. S., & Frick, A. (2010). Early education for spatial intelligence: Why, what, and how. *Mind, Brain, and Education, 4*(3), 102–111.

Newcombe, N. S., & Shipley, T. F. (2015). Thinking about spatial thinking: New typology, new assessments. In *Studying visual and spatial reasoning for design creativity* (pp. 179–192). Dordrecht: Springer.

Noack, H., Lövdén, M., & Schmiedek, F. (2014). On the validity and generality of transfer effects in cognitive training research. *Psychological Research, 78*(6), 773–789.

Olesen, P. J., Westerberg, H., & Klingberg, T. (2004). Increased prefrontal and parietal activity after training of working memory. *Nature Neuroscience, 7*(1), 75–79.

Orsini, A. (1994). Corsi's block-tapping test: Standardization and concurrent validity with WISC-R for children aged 11 to 16. *Perceptual and Motor Skills, 79*(3f), 1547–1554.

Owen, A. M., Hampshire, A., Grahn, J. A., Stenton, R., Dajani, S., Burns, A. S., Howard, R. J., & Ballard, C. G. (2010). Putting brain training to the test. *Nature, 465*, 775–778. doi:10.1038/nature09042.

Pashler, H. (1994). Dual-task interference in simple tasks: Data and theory. *Psychological Bulletin, 116*(2), 220.

Passolunghi, M. C., & Lanfranchi, S. (2012). Domain-specific and domain-general precursors of mathematical achievement: A longitudinal study from kindergarten to first grade. *British Journal of Educational Psychology, 82*(1), 42–63.

Perels, F., Gürtler, T., & Schmitz, B. (2005). Training of self-regulatory and problem-solving competence. *Learning and Instruction, 15*(2), 123–139.

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*(1), 39–58.

Piaget, J., & Inhelder, B. (1956). The child's conception of space (F. J. Langdon & J. L. Lunzer, Trans.). New York: Norton.

Presson, C. C. (1982). The development of map-reading skills. *Child Development, 53*, 196–199.

Quinn, J. G. (2008). Movement and visual coding: The structure of visuo-spatial working memory. *Cognitive Process, 9*(1), 35–43.

Raz, N., Schmiedek, F., Rodrigue, K. M., Kennedy, K. M., Lindenberger, U., & Lövdén, M. (2013). Differential brain shrinkage over 6 months shows limited association with cognitive practice. *Brain and Cognition, 82*(2), 171–180.

Reckase, M. D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement, 21*, 25–36.

Reckase, M. D. (2009). *Multidimensional item response theory*. New York: Springer.

Reckase, M. D., Ackerman, T. A., & Carlson, J. E. (1988). Building a unidimensional test using multidimensional items. *Journal of Educational Measurement, 25*(3), 193–203.

Reckase, M. D., & Xu, J. R. (2015). The evidence for a subscore structure in a test of English language competency for English language learners. *Educational and Psychological Measurement, 75*, 805–825.

Reuhkala, M. (2001). Mathematical skills in ninth-graders: Relationship with visuo-spatial abilities and working memory. *Educational Psychology, 21*, 387–399.

Rickard, T. C., Healy, A. F., & Bourne, L. E., Jr. (1994). On the cognitive structure of basic arithmetic skills. Operation, order, and symbol transfer effects. *Journal of Experimental Psychology. Learning, Memory, and Cognition, 20*, 1139–1153.

Rilea, S. L. (2008). A lateralization of function approach to sex differences in spatial ability: A reexamination. *Brain and Cognition, 67*(2), 168–182.

Rilea, S. L., Roskos-Ewoldsen, B., & Boles, D. (2004). Sex differences in spatial ability: A lateralization of function approach. *Brain and Cognition, 56*(3), 332–343.

Rohde, T. E., & Thompson, L. A. (2007). Predicting academic achievement with cognitive ability. *Intelligence, 35*, 83–92.

Sanz de Acedo Lizarraga, M. L., & Garcia Ganuza, J. M. (2003). Improvement of mental rotation in girls and boys. *Sex Roles, 49*, 277–286.

Schatschneider, C., Francis, D. J., Foorman, B. R., Fletcher, J. M., & Mehta, P. (1999). The dimensionality of phonological awareness: An application of item response theory. *Journal of Educational Psychology, 91*(3), 439.

Schubert, T., Strobach, T., & Karbach, J. (2014). New directions in cognitive training: On methods, transfer, and application. *Psychological Research, 78*(6), 749.

Shelton, A. L., & Gabrieli, J. D. (2002). Neural correlates of encoding space from route and survey perspectives. *The Journal of Neuroscience, 22*(7), 2711–2717.

Shelton, A. L., & Pippitt, H. A. (2007). Fixed versus dynamic orientations in environmental learning from ground-level and aerial perspectives. *Psychological Research, 71*(3), 333–346.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects.

Shepard, S., & Metzler, D. (1988). Mental rotation: Effects of dimensionality of objects and type of task. *Journal of Experimental Psychology: Human Perception and Performance, 14*(1), 3.

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*(6), 689.

Smedslund, J. (1963). The effect of observation on children's representation of the spatial orientation of a water surface. *The Journal of Genetic Psychology, 102*(2), 195–201.

Spinath, B., Spinath, F. M., Harlaar, N., & Plomin, R. (2006). Predicting school achievement from general cognitive ability, self-perceived ability, and intrinsic value. *Intelligence, 34*(4), 363–374.

St Clair-Thompson, H. L., Stevens, R., Hunt, A., & Bolder, E. (2010). Improving children's working memory and classroom performance. *Educational Psychology, 30*(2), 203–219.

Stransky, D., Wilcox, L. M., & Dubrowski, A. (2010). Mental rotation: Cross-task training and generalization. *Journal of Experimental Psychology: Applied, 16*(4), 349.

Tagaris, G. A., Kim, S., Strupp, J. P., Andersen, P., Uğurbil, K., & Georgopoulos, A. P. (1997). Mental rotation studied by functional magnetic resonance imaging at high field (4 Tesla): Performance and cortical activation. *Journal of Cognitive Neuroscience, 9*(4), 419–432.

Terlecki, M. S., Newcombe, N. S., & Little, M. (2008). Durable and generalized effects of spatial experience on mental rotation: Gender differences in growth patterns. *Applied Cognitive Psychology, 22*(7), 996–1013.

Thomas, H., & Turner, G. F. (1991). Individual differences and development in water-level task performance. *Journal of Experimental Child Psychology, 51*(2), 171–194.

Thorndike, R. M. (1997). *Measurement and evaluation in psychology and education* (7th ed.). Columbus: Merrill Prentice-Hall.

Tran, U. S., & Formann, A. K. (2008). Piaget's water-level tasks: Performance across the lifespan with emphasis on the elderly. *Personality and Individual Differences, 45*(3), 232–237.

Trojano, L., Grossi, D., Linden, D. E., Formisano, E., Goebel, R., Cirillo, S., … & Di Salle, F. (2002). Coordinate and categorical judgements in spatial imagery. An fMRI study. *Neuropsychologia*, *40*(10), 1666–1674.

Uttal, D. H., & Cohen, C. A. (2012). Spatial thinking and STEM education: When, why and how. *Psychology of Learning and Motivation, 57*(2), 147–181.

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*(2), 352.

van de Schoot, R., Lugtig, P., & Hox, J. (2012). A checklist for testing measurement invariance. *European Journal of Developmental Psychology, 9*(4), 486–492.

van der Ham, I. J., Raemaekers, M., van Wezel, R. J., Oleksiak, A., & Postma, A. (2009). Categorical and coordinate spatial relations in working memory: An fMRI study. *Brain Research, 1297*, 70–79.

Van der Molen, M. J., Van Luit, J. E. H., Van der Molen, M. W., Klugkist, I., & Jongmans, M. J. (2010). Effectiveness of a computerised working memory training in adolescents with mild to borderline intellectual disabilities. *Journal of Intellectual Disability Research, 54*, 433–447.

Vandenberg, S. G., & Kuse, A. R. (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills, 47*(2), 599–604.

Vasta, R., & Liben, L. S. (1996). The water-level task: An intriguing puzzle. *Current Directions in Psychological Science, 5*(6), 171–177.

Vasta, R., Belongia, C., & Ribble, C. (1994). Investigating the orientation effect on the water-level task: Who? When? and why? *Developmental Psychology, 30*(6), 893.

Vasta, R., Knott, J. A., & Gaze, C. E. (1996). Can spatial training erase the gender differences on the water-level task? *Psychology of Women Quarterly, 20*(4), 549–567.

Vecchi, T., Monticellai, M. L., & Cornoldi, C. (1995). Visuo-spatial working memory: Structures and variables affecting a capacity measure. *Neuropsychologia, 33*, 1549–1564.

Voyer, D., & Hou, J. (2006). Type of items and the magnitude of gender differences on the Mental Rotations Test. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 60*(2), 91.

Voyer, D., Rodgers, M. A., & McCormick, P. A. (2004). Timing conditions and the magnitude of gender differences on the Mental Rotations Test. *Memory & Cognition, 32*(1), 72–82.

Wager, T. D., & Smith, E. E. (2003). Neuroimaging studies of working memory. *Cognitive, Affective, & Behavioral Neuroscience, 3*(4), 255–274.

Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology, 101*(4), 817.

Wallace, B., & Hofelich, B. (1992). Process generalisation and the prediction of performance on mental imagery tasks. *Memory Cognition, 20*, 695–704.

Wanzel, K. R., Hamstra, S. J., Anastakis, D. J., Matsumoto, E. D., & Cusimano, M. D. (2002). Effect of visual-spatial ability on learning of spatially-complex surgical skills. *The Lancet, 359*(9302), 230–231.

Wenger, E., Schaefer, S., Noack, H., Kühn, S., Mårtensson, J., Heinze, H. J., … & Lövdén, M. (2012). Cortical thickness changes following spatial navigation training in adulthood and aging. *Neuroimage*, *59*(4), 3389–3397.

Wexler, M., Kosslyn, S. M., & Berthoz, A. (1998). Motor processes in mental rotation. *Cognition, 68*(1), 77–94.

Wiedenbauer, G., & Jansen-Osmann, P. (2008). Manual training of mental rotation in children. *Learning and Instruction, 18*(1), 30–41.

Wittig, M. A., & Allen, M. J. (1984). Measurement of adult performance on Piaget's water horizontality task. *Intelligence, 8*(4), 305–313.

Wright, R., Thompson, W. L., Ganis, G., Newcombe, N. S., & Kosslyn, S. M. (2008). Training generalized spatial skills. *Psychonomic Bulletin & Review, 15*(4), 763–771.

Yamamoto, N., & DeGirolamo, G. J. (2012). Differential effects of aging on spatial learning through exploratory navigation and map reading. *Frontiers in Aging Neuroscience, 4*, 14.

Zacks, J. M. (2008). Neuroimaging studies of mental rotation: A meta-analysis and review. *Journal of Cognitive Neuroscience, 20*(1), 1–19.