CHAPTER FOUR

Spatial Thinking and STEM Education: When, Why, and How?

David H. Uttal *and* Cheryl A. Cohen

Contents

Abstract

We explore the relation between spatial thinking and performance and attainment in science, technology, engineering and mathematics (STEM) domains. Spatial skills strongly predict who will go into STEM fields. But why is this true? We argue that spatial skills serve as a gateway or barrier for entry into STEM fields. We review literature that indicates that psychometrically-assessed spatial abilities predict performance early in STEM learning, but become less predicative as students advance toward expertise. Experts often have mental representations that allow them to solve problems without having to use spatial thinking. For example, an expert chemist who knows a great deal about the structure and behavior of a particular molecule may not need to mentally rotate a representation of this molecule in order to make a decision about it. Novices who have low levels of spatial skills may not be able to advance to the point at which spatial skills become less important. Thus, a program of spatial training might help to increase the number of people who go into STEM fields. We review and give examples of work on spatial training, which show that spatial abilities are quite malleable. Our chapter helps to constrain and specify when and how spatial abilities do (or do not) matter in STEM thinking and learning.

1. INTRODUCTION

There is little doubt that the United States faces a serious, and growing, challenge to develop and educate enough citizens who can perform jobs that demand skill in science, technology, engineering, and mathematics (STEM) domains. We do not have enough workers to fill the demand in the short run, and the problem is only likely to get worse in the long run (Kuenzi, Matthews, & Mangan, 2007; Mayo, 2009; Sanders, 2009). Addressing the "STEM challenge" is thus a concern of great national priority. For example, President Obama noted that "Strengthening STEM education is vital to preparing our students to compete in the 21st century economy and we need to recruit and train math and science teachers to support our nation's students." (White House Press Release, September 27, 2010).

In this paper we focus on one factor that may influence people's capacity to learn and to practice in STEM-related fields: spatial thinking. The contribution of spatial thinking skill to performance in STEM-related fields holds even when controlling for other relevant abilities, such as verbal and mathematical reasoning (Wai, Lubinski, & Benbow, 2010). Moreover, substantial research has established that spatial skills are malleable–that they respond positively to training, life experiences, and educational interventions (e.g., Baenninger & Newcombe, 1989; Uttal, Meadow, Hand, Lewis, Warren, & Newcombe, Manuscript in publication. Terlecki, Newcombe, & Little, 2008; Wright, Thompson, Ganis, Newcombe, & Kosslyn, 2008).

Many STEM fields seem to depend greatly on spatial reasoning. For example, much of geology involves thinking about the transformation of physical structures across time and space. Structural geologists need to infer the processes that led to the formation of current geological features, and these processes often, if not always, are spatial in nature. For example, consider the geological *folds* shown in Figure 1. Even to the novice, it seems obvious that this structure must have stemmed from some sort of transformation of rock layers. Opposing tectonic plates created extreme forces which then pushed the rocks into the current configuration. The structural geologist's job is in essence to "undo" these processes and determine why and how the mountains take the shape and form that they do. This is but one of an almost infinite number of spatial and temporal problems that form the field of geology.

Although the importance of spatial thinking may be most obvious in geology, it is equally important in other STEM fields. For example, a great deal of attention is devoted in chemistry to the study and behavior of *isomers,* which are compounds with identical molecular compositions, but different spatial configurations. A particularly important spatial property of isomers is *chirality*, or handedness. As illustrated in Figure 2, a molecule is chiral if its mirror image cannot be superimposed on itself through rotation, translation, or scaling. Molecules that are chiral opposites are called *enantiomers*. Chemistry teachers often use a classic analogy to explain chirality, namely, the spatial relation between a person's right and left hand. Although they share the same set of objects (fingers and thumbs),
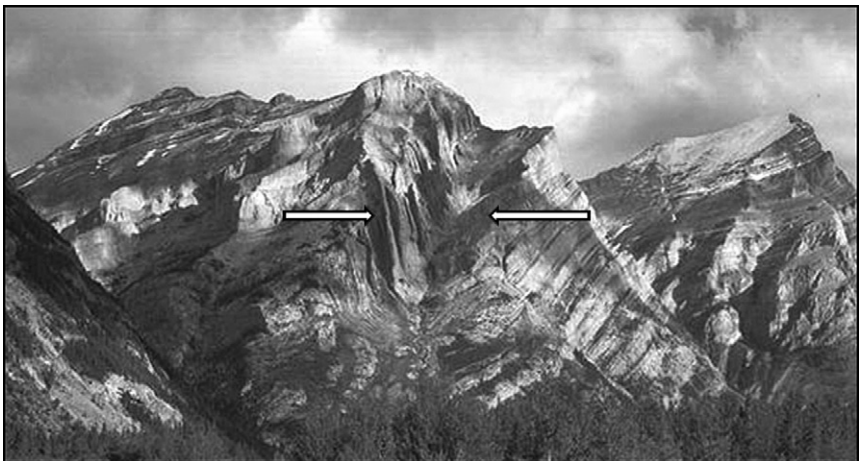

Figure 1   Geological folds in the Canadian Rockies. The arrows point to one aspect of the structure that was created through folding. (B. Tikoff, personal communication, December 28, 2011). (Photograph courtesy of Steve Wojtal, used with permission.) (For color version of this figure, the reader is referred to the web version of this book.)
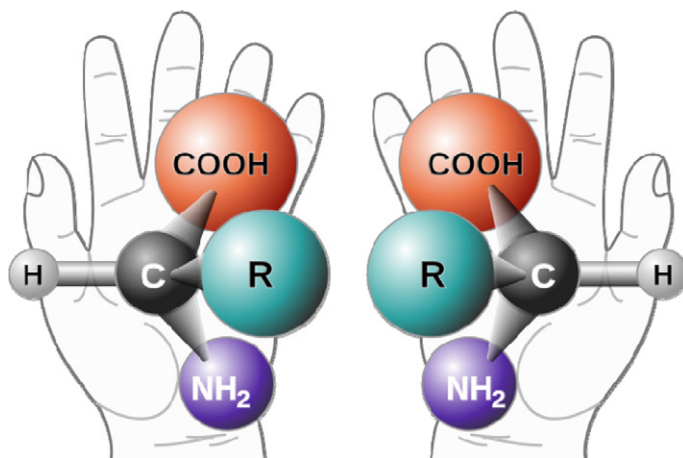
Figure 2 Chirality. Although the two molecules above have the same set of spatial relations, it is not possible to transform one molecule into the other through spatial transformations such as rotation, translation or scaling. The same property holds true for the relation between our two hands. (Image is in the public domain.) (For color version of this figure, the reader is referred to the web version of this book.)

and the same set of relations among these objects, it is not possible to superimpose the left hand onto the right hand. Chemists and physicists have adopted this embodied metaphor, often referring to left- and right-hand configurations of molecules.

Chirality matters greatly because although enantiomers share the same atoms, their spatial differences greatly affect how the isomers behave in chemical reactions. A classic example was the failure to distinguish between enantiomers of the Thalidomide molecule. One version of this drug acted as an effective treatment for morning sickness, and was prescribed in the early 1960s to many thousands of pregnant women. Unfortunately, its enantiomer caused very serious birth defects. Chemists and pharmacists did not realize that this spatial, but not structural, difference was important until it was too late (Fabro, Smith, & Williams, 1967; See Leffingwell, 2003 for other examples). Both forms were included in the dispensed drug, which led to notoriously severe birth defects.

As in our discussion of geology, this is but one of a great number of spatial relations that are critically important in chemistry. As many researchers (and students) have noted, learning to understand systems of spatial relations among molecules, and the representations of these molecules pictorially or with physical blocks, is one of the central challenges in learning chemistry.

2. STEM LEARNING AND SPATIAL TRAINING: A SKEPTICAL FIRST LOOK

The spatial demands of STEM learning and practice raise intriguing questions: can teaching people to think spatially lead to improvements in STEM education? Should spatial training be added to the arsenal of tools and techniques that educators, researchers, businesses, and the military are using to try to increase competence in STEM-relevant thinking? There is growing enthusiasm about the promise of training spatial thinking, and some researchers and educators have developed and refined spatial training programs that are specifically designed to enhance spatial thinking and prevent dropout from STEM fields. For example, Sorby & Baartmans (1996, 2000) developed a ten-week course to train spatial thinking skills that are important early in the college engineering curriculum. The program has been very successful, leading to large and substantial gains not only in engineering retention but also in psychometrically-assessed spatial ability.

However, before embarking on a large-scale program of spatial training, we need to think very carefully and skeptically about how and why spatial thinking is, and is not, related to STEM achievement. We want educational interventions to be based on the strongest possible evidence. Is the existing evidence strong enough to support the recommendation that spatial training should be instituted to raise the number of STEM-qualified workers and students? The many reported correlations between STEM achievement and spatial ability are a necessary first step, but simple correlations are obviously not enough to justify the implementation of large-scale implementations. Our skepticism is also justified by preliminary empirical findings. For example, the results of several studies indicate that the relation between spatial skills and STEM achievement grows *smaller* as expertise in a STEM field increases.

Our primary goal therefore is to review and synthesize the existing evidence regarding the relation between spatial skills and STEM achievement. We take a hard look at the evidence, and we also consider when, why, and how spatial abilities do and do not relate to STEM learning and practice, both at the expert and novel levels. In addition to its practical importance, the questions we raise here have important implications for cognitive psychology. For example, we discuss what happens at the level of cognitive representation and processing when one becomes an expert in a spatially-rich STEM domain. Our discussion sheds substantial light not only on the role of spatial reasoning in STEM but also on the characterization of expert knowledge in spatially-rich or demanding content domains.

We begin by discussing what spatial thinking is and how it has been defined. We then consider the existing evidence that spatial ability and STEM performance are related. This review indicates that spatial abilities *do* predict both entrance into STEM occupations and performance on STEM-related tasks in novices. However, the evidence for a relationship between spatial skills and STEM occupations and performance is weaker and less consistent in for STEM experts. For example, whether expert geologists succeed or fail on an authentic geology task seems to have little to do with their level of spatial skill (Hambrick et al., 2011). We then consider possible causes of this surprising, perhaps even paradoxical, novice-expert difference. We conclude that much of the difference stems from how experts represent and process domain-specific knowledge. As domain-specific knowledge increases, the need for the abilities measured by typical spatial abilities tests goes down.

This pattern of results suggests a specific role for spatial training in STEM education: spatial training may help novices because they rely more on de-contextualized spatial abilities than experts do. Therefore, spatial training might help to prevent a consistent problem in STEM education: Frequent dropout of students who enter STEM disciplines (but fail to complete their degrees and often go into non–STEM fields). We then consider research on the effectiveness of spatial training, including a recent meta-analysis (Uttal et al., (Manuscript accepted for publication)) that has shown that spatial skills are quite malleable, and that the effects of training can endure over time and can transfer to other, untrained tasks. We conclude by making specific recommendations about when, whether, and why spatial training could enhance STEM attainment. We also point the way to the next steps in research that will be needed to fully realize the potential of spatial training.

3. WHAT IS SPATIAL THINKING?

Any discussion of a psychological construct such as spatial thinking should begin with a clear definition of what it is. Unfortunately, providing a good definition is not nearly as easy as one would hope or expect. It is easy enough to offer a general definition of spatial thinking, as we already did above. However, it turns how to be much harder to answer questions such as the following: is there one spatial ability, or are there many? If there are many kinds of spatial abilities, how do they relate to one another? Can we speak about how spatial information is represented and processed independent of other abilities (Gershmehl & Gershmehl, 2007).

Many factor-analytic studies have addressed these sorts of questions. However, these studies have not yielded consistent results, in part because the resulting factors are greatly affected by the tests that are used, regardless

of what the researcher intended the test to measure (Linn & Peterson, 1985; Hegarty & Waller, 2005). Theoretical analyses, based on the cognitive processes that are involved, have proved somewhat more promising, although there is still no consensus as to what does and does not count as spatial thinking (Hegarty & Waller, 2005).

Generally speaking, most of the research linking spatial abilities and STEM education has focused on what Carroll (1993) termed *spatial visualization,* which is the processes of apprehending, encoding, and mentally manipulating three-dimensional spatial forms. Some spatial visualization tasks involve relating two-dimensional representations to three-dimensional representations, and vice versa. Spatial visualization is a sub-factor that is relevant to thinking in many disciplines of science, including biology (Rochford, 1985; Russell-Gebbett, 1985), geology (Eley, 1983; Kali & Orion, 1996; Orion, Ben-Chaim, & Kali, 1997), chemistry (Small & Morton, 1983; Talley, 1973; Wu & Shah, 2004), and physics (Kozhevnikov, Motes, & Hegarty, 2007; Pallrand & Seeber, 1984). As applied to particular domains of science, spatial visualization tasks involve imagining the shape and structure of two-dimensional sections, or cross sections, of three-dimensional objects or structures. Mental rotation is sometimes considered to be a form of spatial visualization, although other researchers consider it to be a separate factor or skill (Linn & Peterson, 1985).

Although it is not always possible to be as specific as we would like about the definition of spatial skills, it *is* possible to be clearer about what psychometric tests do not measure: complex, expert reasoning in scientific domains. By definition, most spatial abilities tests are designed to isolate specific skills or, at most, small sets of spatial skills. They therefore are usually deliberately de-contextualized; they follow the traditional IQ testing model of attempting to study psychological abilities independent of the material on which they are used. For example, at least in theory, a test of mental rotation is supposed to measure one's ability to rotate stimuli in general. As we discuss below, the kinds of knowledge that psychometric tests typically measure may therefore become *less* important as novices advance toward becoming experts. We therefore need to be very careful about assuming that complex spatial problems in STEM domains are necessarily solved using the kinds of cognitive skills that psychometric tests tap.

4. Relations between Spatial Thinking and STEM Achievement and Attainment

Many studies have shown that there are moderate-to-strong correlations between various measures of spatial skills and performance in particular STEM disciplines. For example, a variety of spatial skills

are positively correlated with success on three-dimensional biology prob-
lems (Russell-Gebbett, 1985). Rochford (1985) found that students who
had difficulty in spatial processes such as sectioning, translating, rotating
and visualizing shapes also had difficulty in practical anatomy classes.
Hegarty, Keehner, Cohen, Montello, and Lippa (2007) established that
the ability to infer and comprehend cross sections is an important skill
in comprehending and using medical images such as x-ray and
magnetic resonance images. The ability to imagine cross sections,
including the internal structure of 3-D forms is also central to geology,
where it has been referred to as "visual penetration ability" (Kali &
Orion, 1996; Orion, Ben-Chaim, & Kali, 1997). Understanding the
cross-sectional structure of materials is a fundamental skill of engineering
(Duesbury & O'Neil, 1996; Gerson, Sorby, Wysocki, & Baartmans,
2001; Hsi, Linn, & Bell, 1997; Lajoie, 2003). These and many similar
findings led Gardner (1993) to conclude that "it is skill in spatial ability
which determines how far one will progress in the science" (p. 192).
(See Shea, Lubinski, & Benbow, 2001, for additional examples).

Thus, there is little doubt that zero-order correlations between various
spatial measures and STEM outcomes are significant and often quite strong.
But there is an obvious limitation with relying on these simple correlations:
the third variable problem. Although spatial intelligence is usually the first
division in most hierarchical theories of intelligence, it is obviously corre-
lated with other forms of intelligence. People who score highly on tests
of spatial ability also tend to score at least reasonably well on tests of other
forms of intelligence, such as verbal ability. For example, although current
chemistry professors may have performed exceptionally well on spatial
ability tests, they are likely as well to have performed reasonably well on
the verbal portion of the SAT, a college admissions test that is used widely
in the United States. The observed correlations between spatial ability and
achievement therefore must be taken with a grain of salt because of the
strong possibility that their correlations are due to unidentified variables.

4.1. Moving Beyond Zero-Order Correlations

Fortunately, some studies have controlled more precisely for several other
variables, using multiple regression techniques. For example, Lubinski,
Benbow and colleagues (e.g., Shea et al., 2001; Wai, Lubinski,
& Benbow, 2009) have demonstrated a unique predictive role for spatial
skills in understanding STEM achievement and attainment. These
researchers used large-scale datasets that often included tens of thousands
of participants. In general, the original goal of the research was not
(specifically) to investigate the relation between spatial skills and STEM,
but the original researchers did include enough measures to allow future
researchers to investigate these relations.

Benbow and Stanley (1982) studied the predictive value of spatial abilities among gifted and talented youth enrolled in the Study of Mathematically Precocious Youth. To enter the study, students took several tests in middle school, including both the SAT Verbal and the SAT Math. Students also completed two measures of spatial ability, the Space Relations and Mechanical Reasoning subtests of the Differential Aptitude Test. In many cases, the original participants have been followed for thirty years or more, allowing the researchers to assess the long-term predictive validity of spatial tests on (eventual) STEM achievement and attainment.

This work showed that psychometrically-assessed spatial skills are a strong predictor of STEM attainment. The dependent variable here is the career that participants eventually took up. Even after holding constant the contribution of verbal and mathematics SAT, spatial skills contributed greatly to the prediction of outcomes in engineering, chemistry, and other STEM disciplines. These studies clearly establish a unique role of spatial skills in predicting STEM achievement.

However, one potential limitation is that they were initially based on a sample that is not representative of the general U.S. population. As its name implies, the Study of Mathematically Precocious Youth is not a representative sample of American youth. To be admitted to the study, youth had to be (a) identified in a talent search as being among the top 3% in mathematics, and then (b) score 500 or better on both the Verbal and Mathematics SAT at 12- to14-years of age. In combination, these selection criteria resulted in a sample that represented the upper 0.5% of American youth at the time of testing (1976—78) (Benbow & Stanley, 1982).

It is reasonable to ask whether the results are limited to this highly selected sample (Wai et al., 2009). If so, they would not provide a solid foundation for a program of spatial training to facilitate STEM learning among more typical students. For these reasons, Wai et al. extended their work to more diverse samples. They used the Project Talent database, which is a nationally representative sample of over 400,000 American high school students, approximately equally distributed across grades 9—12. The participants were followed for 19 years, again allowing the researchers to predict ultimate career choices. The results in the more representative sample were quite similar to those of the project talent dataset, and hence it seems quite likely that spatial skills indeed are a unique, specific predictor of who goes into STEM.

Figure 3 provides a visual summary of Wai et al's findings on the relations between cognitive abilities assessed in high school and future career choice. The figure includes three axes, representing Verbal, Mathematical and Spatial ability on the X, Y, and Z axes, respectively. The scores are expressed as z-scores; the numbers on the axes represent deviations from zero expressed in standard deviation units. The X and Y axes are easy to understand. For example, the 23 participants who ended up in science
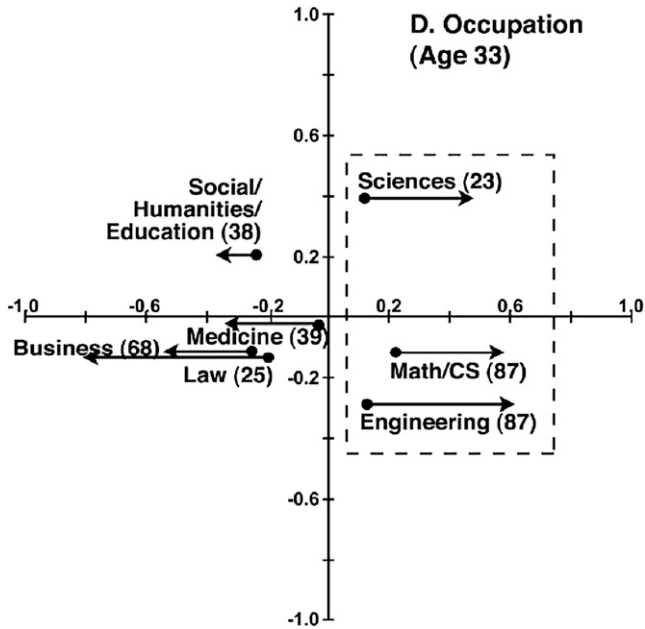
Figure 3 Results from Wai, Lubinski, and Benbow (2009). The X axis represents Math SAT, and the Y axis represents Verbal SAT, expressed in standard deviation units. The arrows are a third, or Z, dimension. The length of the arrow represents the unique contribution of the spatial ability test to predicting eventual career. (Reprinted with permission of the American Psychological Association.)

occupations scored about 0 .40 *SD* above the mean on the SAT Math. The Z axis is represented by the length of the vectors extending from the point representing the intersection of the X and Y axis. The length of each vector can be construed as the value-added of knowing the spatial score in predicting entry into the particular career. Note that the vectors are long and in the positive direction for all STEM fields. Moreover, spatial ability also strongly predicts entry into business, law, and medicine, but in the negative direction. Clearly, if one wants to predict (and perhaps ultimately affect) what careers students are likely to choose, knowing their level of spatial skills is critically important (Wai et al., 2009).

Moreover, there appears to be no upper limit on the relation between spatial skills and STEM thinking. The relation between spatial skill and STEM attainment held even several standard deviations from the mean; the most spatially talented youth were the most likely to go into STEM fields, even at the very upper ends of the distribution of the spatial abilities test.

In summary, psychometrically-assessed spatial ability strongly predicts who does and does not enter STEM fields. Moreover, this relation holds true even after accounting for other variables, such as Mathematics and

Verbal Aptitude. In fact, in some fields, spatial ability contributes more unique variance than SAT scores do to the prediction of STEM achievement and attainment. Wai et al. (2009) noted that the evidence relating spatial ability and future STEM attainment is exceptionally strong, covering 50 years of research with more than 400,000 participants, with multiple datasets converging on very similar conclusions.

5. Spatial Cognition and Expert Performance in STEM Disciplines

The results presented thus far make a strong case for the importance of spatial reasoning in *predicting* who goes into STEM fields and who stays in STEM. But why is this true? At first glance, the answer seems obvious: STEM fields are very spatially demanding. Consequently, those who have higher spatial abilities are more able to perform the complex spatial reasoning that STEM requires. It makes sense that no upper limit on the relation has been identified; the better one is at spatial skills, the better one is at STEM. On this view, there is a strong relation between spatial ability and STEM performance, at all levels of expertise because spatial abilities either limit or enhance whether a person is able to perform the kinds of spatial thinking that seem to characterize STEM thinking (See Stieff, 2004, 2007 for a more detailed account and critique of this explanation).

But this seemingly simple answer turns out not to be so simple. In this section we present a seeming paradox: even though spatial abilities are highly correlated with entry into a STEM field, they actually tend to become *less* important as a student progresses to mastery and ultimately expertise. Despite the well-replicated correlations between spatial abilities and choosing a STEM career, experts seem to rely surprisingly little on the kinds of spatial abilities that are tested in spatial ability tests. In the next section we consider the literature that supports these claims.

We note at the outset of this discussion that research on the spatial abilities and their role in STEM expertise is rather limited. Although there are many studies of spatial ability in STEM learners, many fewer have investigated the role of spatial ability in expert performance. Thus we are limited to some extent in judging the replicability and generalizability of the findings we report. Moreover, our choice of which disciplines to discuss is limited by the availability of research on expertise in the STEM disciplines.

5.1. Spatial Cognition and Expert Performance in Geology

Perhaps the best examples come from geology. As we have already noted, structural geology is basically a science of spatial and temporal

transformations, so if one were looking for relations between spatial ability and expert performance, this field would seem to be a good place to start. Hambrick et al. (2011) investigated the role of psychometrically-assessed spatial ability in expert and novice performance in a real-world geosciences task, bedrock mapping. Starting with a blank map, geologists or geology students were asked to map out the underlying structures in a given area, based on the observable surface features. This task would seem to require domain-specific knowledge about the kinds of rocks that might be found in given geological areas or are associated with given structures. At the same time, it would seem to require spatial reasoning, as the geologist must make inferences about how forces transformed underlying rock beds to produce the observed structured.

The study was conducted as part of a geology research and training camp, in the Tobacco Mountains of Montana. On Day 1, participants took several tests of both geospatial knowledge and cognitive ability, including spatial skills. On Day 2, participants were driven to four different areas and heard descriptions of the rock structures found there. They were then asked to complete the bedrock mapping task for that area. Each map was compared to a correct map that was generated by two experts. Scores were derived by comparing the participant's drawn map to a computerized, digital version of the correct map. This method resulted in a very reliable deviation score, which was then converted to a map accuracy percentage.

The primary results are presented in Figure 4, which is adapted from Hambrick et al. (2011). The dependent variable (shown on the Y axis) was average map accuracy. As the graph indicates, there was a significant interaction between visuospatial ability and geospatial knowledge. The graph is based on median splits of the two independent variables. For those
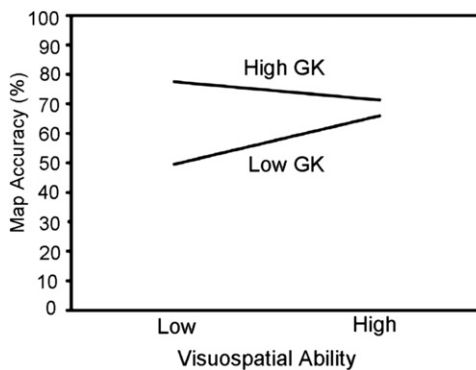

Figure 4 Results from Hambrick et al. (2011) spatial ability and expert geology performance. "GK" refers to geology knowledge.

with high geospatial knowledge, visuospatial ability did not affect performance on the bedrock mapping task. However, there was a significant effect of visual spatial ability in the low geospatial-knowledge group: those with high visual spatial ability performed well; their performance nearly matched that of the high geospatial knowledge group. However, individuals who had both low visuospatial ability and low geospatial knowledge performed much worse. Although not shown in the figure, the standard deviations in the two groups were nearly identical, suggesting that the lack of correlation between spatial skills and performance in the experts was not due to restriction of range. One might assume that the geology experts would all have high spatial skills and thus there would be little or no variance, but this turned out not to be true.

These results support the conclusion that visual spatial ability does not seem to predict performance among experts; those with high levels of geospatial knowledge performed very well on the task, regardless of their level of visual–spatial ability. Hambrick et al. (2011) concluded, "Visuospatial ability appears to matter for bedrock mapping, but only for novices," (p. 5).

Hambrick et al., (2011) (see also Hambrick & Meinz, 2011) coined the phrase the "circumvention-of-limits" hypothesis, suggesting that the acquisition of domain-specific knowledge eventually reduces or even eliminates the effects of individual differences in cognitive abilities. Their hypothesis is consistent with earlier work on skill acquisition (e.g., Ackerman, 1988) that showed that individual differences in general intelligence strongly predict performance early in the acquisition of new skills but have less predictive validity.

5.2. Spatial Cognition and Expert Performance in Medicine and Dentistry

Medical domains offer rich opportunities for studying the contribution of spatial abilities to performance. Medical professionals often need to infer the spatial properties of visible or obscured anatomical structures, including their relative locations with respect to each other. Spatial cognition would also seem, at least ostensibly, to be centrally important to understanding medical images, including those produced by CT, MRI, X-ray and ultrasound.

Hegarty, Keehner, Khooshabeh, and Montello (2009) explored the interaction between spatial ability and training by asking two complementary questions: does spatial ability predict performance in dentistry? Does dental education improve spatial ability?

To investigate the first question, Hegarty et al. investigated if spatial and general reasoning measures predicted performance in anatomy and restorative dental classes among first- and fourth-year dental students. First-

year dental students were tested at the beginning and end of the school year, and psychology undergraduates served as a control on the spatial measures. Two of the spatial ability measures were widely-used psychometric tests: a classic mental rotation test and a test of the ability to imagine a view of a given abstract object from a different perspective. The remaining two spatial tests measured the ability to infer cross sections of three-dimensional objects. The stimulus object in the first test was something the participants had never encountered in the natural world: an egg-shaped form with a visible internal structure of tree-like branches. The stimulus figure in the second test was a tooth with visible internal roots. Additional data was collected from the dental students' scores on the Perceptual Ability Test (PAT), a battery of domain-general spatial tests that is used to screen applicants for dental schools. The three groups were matched on abstract reasoning ability.

The spatial ability tests did not predict performance in anatomy classes for either group of dental students. There were modest correlations between performance in restorative dentistry and the investigator-administered spatial ability tests, and these correlations remained after controlling for general reasoning ability. The PAT was a better predictor of dental school performance than any single spatial measure considered alone. However, the contribution of spatial ability to performance in this study is nuanced, as we'll discuss below.

The second research question was addressed by comparing performances on both cross-section measures for all participants, and across test administrations. At the end of one year of study, first-year dental students showed significant improvement in their ability to identify cross-sections of teeth, but not in their ability to infer cross-sections of the egg-like figure. Fourth-year dental students outperformed first-year dental students (on their first attempt) and psychology students on the tooth cross-section test. Together, these results suggest that dental training enabled novice and more experienced students to develop, and refine, mental models of domain-specific objects, rather than to improve general spatial ability. At the same time, the results also provide evidence that spatial ability does not always become irrelevant. Furthermore, spatial ability, as measured by performance on the domain-general spatial tests, predicted performance on the tooth test for all participants, including fourth-year students. Thus, there is evidence that spatial ability *did* enable students to develop the mental models of the spatial characteristics of teeth.

5.3. Spatial Cognition and Expert Performance in Chemistry

Stieff (2004, 2007) investigated expert and novice chemists' performances on a classic visual–spatial task, the mental rotation of three-dimensional figures. He used the classic Shepard and Metzler (1971) figures, which

resemble three-dimensional blocks arranged in different positions. The participant's task is to decide whether a given block is a rotated version of a target. In addition, Stieff also included representations of three-dimensional chemical molecules. These were chemistry diagrams that are commonly taught in first- or second-year college chemistry classes.

There was a fascinating interaction between level of experience and the kinds of stimuli tested. Novice and expert chemists performed nearly identically on the Shepard and Metzler figures. In both groups, there was a strong, linear relation between degree of angular disparity and reaction time. This result is often taken as evidence for mental rotation; it takes more time to turn a stimulus that is rotated a great deal relative to the target than a stimulus that is rotated only slightly.

However, there was a strong expert-novice difference for the representations of three-dimensional symmetric chemistry molecules. The novices again showed the same relation between angular disparity and reaction time; the more the stimulus was rotated, the longer it took them to answer "same or different." In contrast, the function relating angular disparity to reaction time was essentially flat in the data for the experts; the correlation was nearly zero. Experts apparently used a very different mental process to make judgments about the meaningful (to them) representations of real chemical molecules and about the meaningless Shepard and Metzler figures. We discuss what this difference may be in the next section.

5.4. Spatial Cognition and Expert Performance in Physics

Several studies have found correlations between spatial abilities and performance in physics. In fact, in this domain researchers have been quite specific about when and why (e.g., Kozhevnikov, Hegarty, & Mayer, 2002). However, there have been only a few studies of the role of spatial abilities in physics problem-solving at the expert level. It is interesting to note, however, that in one study, spatial ability predicted performance at pre-test, before instruction, but not after instruction (Kozhevnikov & Thornton, 2006). The students in this study were not experts, either before or after instruction. Nevertheless, the results do provide evidence that is consistent with the claim that spatial abilities become less important as knowledge increases.

5.5. Interim Summary

The previous two sections raise a seeming paradox. On the one hand, research clearly demonstrates that spatial cognition is a strong and independent predictor of STEM achievement and attainment. On the other hand, at least at the expert level, spatial abilities do not seem to consistently predict performance. In the next section, we attempt to resolve this

seeming paradox by considering what it means, at the representational and processing level, to be an expert in a spatially-demanding STEM field. Addressing this question turns out to provide important insights into the nature of expert performance in STEM disciplines and the role of spatial cognition in that expertise.

6. The Nature of Expertise in Spatially Demanding STEM Disciplines

To understand why spatial skills seem not to predict performance at the expert level, we need to examine the nature of expertise in spatially-demanding fields. First, we note that *STEM* practice is often highly domain-specific, depending a great deal on knowledge that is accumulated slowly over years of learning and experience. What a chemist does in his or her work, and how he or she uses spatial representations and processes to accomplish it, is not the same as what an expert geoscientist or an expert engineer might do.

Second, we suggest that the nature of domain-specific knowledge is perhaps the primary characteristic of expertise in various STEM fields. Expertise in STEM reasoning is best characterized as a complex interplay between spatial and semantic knowledge. Semantic knowledge helps to constrain the demands of spatial reasoning, or allows it to be leveraged and used to perform specific kinds of tasks that are not easily answered by known facts. In what follows we discuss three specific examples of the nature of expert knowledge in several STEM fields. However, we begin with expertise in a non-STEM field, chess. It turns out that many of the findings and debates regarding the nature of chess expertise are also relevant to understanding STEM expertise in a variety of disciplines. In the case of chess, psychologists have provided quite specific and precise models of expert performance, and we consider whether, and how, these models could help us understand expertise and the role of spatial ability in STEM fields.

6.1. Mental Representations that Support Chess Expertise

Research on chess expertise (e.g., Chase & Simon, 1973) was the vanguard for the intense interest in expertise in cognitive science. Nevertheless, it remains an active area of investigation, and there are still important debates regarding precisely what happens when one becomes expert. A detailed account of these debates is well beyond the scope of this chapter, but a brief consideration of the nature of spatial representations in chess may shed important light on the nature of expertise in STEM fields.

Chess seems, at least ostensibly, to be a very spatially-demanding activity, for the same reasons that STEM fields seem to be. Playing chess seems to require keeping track of the locations, and potential locations, of a large number of pieces. However, just as in the case of STEM fields, psychometric spatial abilities do not consistently predict levels of chess performance (e.g., Holding, 1985; Waters, Gobet, & Leyden, 2002). Moreover, the spatial knowledge that characterizes chess expertise is very different from the kinds of spatial information that are required on spatial ability tests.

Most researchers agree that chess knowledge allows experts to represent larger "chunks" of information, but there is still substantial debate regarding what chunks are. Originally, Chase and Simon proposed that chunks consisted of thousands of possible arrangements or templates for pattern matching. On this view, at least part of the expertise is spatial in nature, in that knowledge allows the expert to encode more spatial information—the locations of multiple pieces—and hence recall more at testing. The specific effect of expertise is that it gives the expert many thousands of possible visual matches to which to assimilate locational information.

However, several researchers have challenged this traditional definition of chunking, stressing instead the organization of pieces in terms of higher-order semantic knowledge that ultimately drives perception and pattern matching. On this view, the "chunk" is not defined specifically by any one pattern of the location of chess pieces on the board. Instead, it is organized around chess-related themes and knowledge, such as patterns of attack and defense, number of moves to checkmate, or even previously studied matches (e.g., McGregor & Howes, 2002). Linhares and Brum's (2007) results highlight well the differences between the two models of chess expertise. They asked chess experts to classify various boards as the same or different. In some cases, experts often labeled two configurations that differed dramatically in the number of pieces as "the same". For example, a configuration that contained four pieces might be labeled "the same" as one that contained nine pieces. This result strongly suggests that the nature of the expertise cannot be based purely on spatial template matching, as it is very difficult to explain how chess arrangements that vary dramatically in so many ways could be included in a template that is defined at least in part on the basis of specific spatial locations on the board. Instead, the effect of the expertise seems to be at a much higher level, and is spatial only in the sense that each piece plays a role in an evolving, dynamic pattern of attack or defense (McGregor & Howes, 2002).

Given this analysis, it should no longer be surprising that de-contextualized spatial abilities do not predict level of expertise in chess. Becoming an expert in chess involves learning thousands (or more) different patterns of attack and defense at different stages of the game. The ability to mentally rotate a meaningless figure bears little relation to what is required to play chess at an expert level.

We are making an analogous claim for the nature of reasoning and problem-solving in expert STEM practice. Experts typically have a great deal of semantic knowledge, and this knowledge influences all aspects of the cognitive-processing chain, from basic visual attention to higher-level reasoning. It affects what they attend to, what they expect to see (hear, smell, etc.), and what they will think about when solving a problem. Memory and problem-solving are tied to the use of this higher-order knowledge, and consequently, lower-order (and more general) spatial abilities become substantially less important as expertise increases. We now discuss research that supports our claims regarding the (lack of) relation between spatial abilities and STEM performance at the expert level.

6.2. Mental Representations that Support Chemistry Expertise

As discussed above, chemistry experts do not seem to use mental rotation to solve problems regarding the configuration of a group of atom in a molecule. In some cases, factual or semantic knowledge will allow the STEM expert to avoid the use of spatial strategies. For example, Stieff's (2007) work on novice-expert differences in spatial ability reveals that experts relied substantially on semantic knowledge in a mental rotation task. The lack of correlation between angular disparity and experts' reaction time suggest that they may have already known the answers to the questions. For example, knowing properties of molecules (e.g., that one molecule is an isomer of another molecule) would allow them to make the "same-different" judgment without need to try to mentally align the molecule with its enantiomer. Stieff (2004, 2007) confirmed this hypothesis in a series of protocol analyses of experts' problem-solving. Semantic knowledge of chemical molecules allowed the experts to forego mental rotation.

6.3. Mental Representations that Support Expertise in Geometry

Koedinger and Anderson (1990) investigated the mental representations and cognitive processes that underlie expertise in geometry. They found that experts organized their knowledge around perceptual chunks that cued abstract semantic knowledge. For example, seeing a particular shape might prime the expert's knowledge of relevant theorems, which in turn would facilitate completing a proof. Thus, even in a STEM field that is explicitly about space, higher-order semantic knowledge guided the perception and organization of the relevant information. Although there are not, to our knowledge, specific studies linking psychometrically-assessed spatial ability with expertise in geometry, Koedinger and

Anderson's results suggest that it would not be surprising to find that spatial ability would not predict performance in advanced geometers.

6.4. Mental Representations that Support Expertise in Radiology

Medical decision-making has been the subject of many computer expert systems that match or exceed clinical judgment in predicting mortality after admission to an Intensive Care Unit. However, relatively few studies have focused specifically on the spatial basis of diagnosis. One important exception to this general claim is work on the development of expertise in radiology: the reading and interpretation of images of parts of the body that are not normally visible.

There have been many studies of the expertise that is involved in radiology practice (e.g., Lesgold et al., 1988). Although an extensive review of this work is beyond the scope of this paper, one consistent finding deserves mention because it again highlights the diminishing role of de-contextualized spatial knowledge and the increasing role of domain-specific knowledge. In comparing radiology students and radiology experts (who had read perhaps as many as 500,000 radiological images in their years of practice), Lesgold et al. (1988) noted that the description of locations and anomalies shifted with experience from one based on locations on the X-ray (e.g., in the upper-left half of the display), to one based on a constructed, mental model of the patient's anatomy (e.g., "there is a well-defined mass in the upper portion of the left lung"). Lesgold et al. (1988) suggested that expert radiologists begin by (a) constructing a mental representation of the patient's anatomy, and (b) coming up with and testing hypotheses of diseases processes and how they would affect the anatomy and hence the displayed image. Wood (1999), a radiologist herself, has described the interaction between spatial and semantic knowledge in the interpretation of radiologic images: "When we examine a radiograph, we recognize normal anatomy, variations in anatomy, and anatomic aberrations." These visual data constitute a stimulus that initiates a recalled generalization of meaning. Linkage of visual patterns to appropriate information is dependent on experience more than on spatial abilities.

Interestingly, the experienced radiologists used *fewer* spatial words in their descriptions of X-rays than the less experienced radiologists did. As in chess, the novice representation includes more information about locations in Euclidean space, and the expert's representation is more based on higher-level, relational knowledge of patterns of attack and defense in the case of chess and the relation between anatomy and disease processes in the case of radiology. Although, to our knowledge, no one has examined the role of psychometrically-assessed spatial skills in expert radiology

practice, we would again predict that their contribution would diminish as experience (and hence domain-specific knowledge) grows.

6.5. When Might Spatial Abilities Matter in Expert Performance?

Of course, it is certainly possible that psychometric spatial abilities may play an important role in other sciences, or in solving different kinds of problems. For example, it seems possible that de-contextualized spatial knowledge might play more of a role during critical new insights. Scientific problem-solving is often described as a moment of spatial insight (for further discussion, see Miller, 1984).

One famous example of insight and discovery of spatial structures is the work of James Watson and Francis Crick, who along with Rosalind Franklin and Maurice Wilkins, discovered the structure of the DNA molecule. This discovery involved a great deal of spatial insight. The data that they worked from were two-dimensional pictures generated from X-ray diffraction, which involves the analysis of patterns created when X-rays bounce off different kinds of crystals. Working from these patterns, Watson and Crick (1953) came to the conclusion that the (three-dimensional) double-helix structure could generate the patterns of two-dimensional photographs from which they worked. They studied other proposed structures but eventually rejected them as insufficient to account for the data. They then wrote, "We wish to put forward a radically different structure for the salt of deoxyribonucleic acid." (1953, p. 737). This radically different structure was the double-helix. We speculate that at moments of insight into "radically different structures", spatial ability may again become important. When there is no semantic knowledge to rely on, a scientist making a new discovery may have to revert to the same processes that novices use (e.g., Miller, 1984).

Some STEM disciplines besides STEM that may require spatial insight at more advanced levels of expertise, perhaps because they frequently require the design of new structures or insights. For example, various domains of engineering require that expert practitioners create new designs. The allied field of architecture also demands high levels of spatial thinking ability at all levels of expertise. But it is possible that spatially-intensive arts expertise, such as that required in architecture, may depend more on de-contextualized spatial abilities that are measured by spatial ability tests. This suggestion is obviously speculative, but it is interesting to note that we are not the only ones to make it. For example, scholars at the Rhode Island School of Design have proposed that the acronym STEM be expanded to STEAM, with the additional "A" representing Art (www.stemintosteam.org), in part to encourage more creative approaches to problem solving in STEM.

6.6. A Foil: Expertise in Scrabble

It may seem odd to finish a section on expertise in STEM practice with a discussion of expertise in Scrabble, a popular board game involving the construction of words on a board, using individual tiles for each letter. However, comparing the importance of de-contextualized spatial skills in STEM, Chess, and Scrabble affords what Markman and Gentner (1993) have termed an "alignable difference"—comparing the similarities and differences in the role of psychometric spatial abilities in Scrabble and in the previously reviewed fields makes clearer when and why spatial abilities matter in expertise.

Halpern and Wai (2007) investigated the relation between a variety of psychometric measures and expert performance in Scrabble. It is important to note that expert-level Scrabble differs substantially from the Scrabble that most of us have played at home or online. For example, in competitions, experts play the game under severe time pressure.

Two skills seem to predict expert-level performance in Scrabble: the ability to memorize a great number of words, and the ability to quickly mentally transform spatial configurations of words to find possible ways to spell. In contrast to chess, there are no specific patterns of attack and defense in Scrabble; experts need to be able to mentally rotate or otherwise transform existing board configurations to anticipate where they might be able to place the letters in their rack. Chess experts spend a great deal of time studying prior matches, but Scrabble experts do not. Spatial abilities matter, even at the level of a national champion, because players must be able to mentally transform emerging patterns to find places where the letters in their rack could make new, high-scoring words.

These examples illustrate a general point about when and why spatial abilities. The question should not be only, "Do spatial abilities matter?" but also, when, why, and how they matter. Spatial abilities are one important part of the cognitive architecture, but in real-life they are rarely used out of context or in isolation from other cognitive abilities. Although cognitive psychology textbooks may divide up semantic and spatial knowledge, the two are intimately intertwined in normal, everyday cognitive processing. Knowledge can often point people to the correct answers to spatial questions and hence reduces the need to rely on more general spatial skills. Nevertheless, there also situations in which psychometrically-assessed spatial skills will remain critically important.

6.7. Interim Summary

In summary, expertise in STEM fields bears some important similarities to expertise in chess: Although judgments are often made that involve information about the locations of items in space, these decisions are often made

in ways that differ fundamentally from the kinds of spatial skills that spatial ability tests measure. Experts' spatial knowledge is intimately embedded with their semantic knowledge of chess. The differences in representations and process help to explain why spatial ability usually does not predict performance at the expert level. However, the question of when spatial ability might matter to experts remains an important and open question.

7. THE ROLE OF SPATIAL ABILITIES IN EARLY STEM LEARNING

The results discussed thus far indicate that spatial abilities do predict STEM career choice, but that spatial abilities matter less as expertise increases. We suggest that spatial skills may be a gatekeeper or barrier for success early on in STEM majors, when (a) classes are particularly challenging, and (b) students do not yet have the necessary content knowledge that will allow them to circumvent the limits that spatial ability imposes. Early on, some students may face a Catch-22: they do not yet have the knowledge that would allow them to succeed despite relatively low spatial skills, and they can't get that knowledge without getting through the early classes where students must rely on their spatial abilities. This explanation would also account for the strong correlations between spatial abilities and STEM attainment that have been consistently documented in multiple, large-scale datasets (e.g., Wai et al., 2009). On our view, spatial skills correlate positively with persistence and attainment in STEM because those with low spatial abilities either do not go into STEM majors or dropout soon after they begin.

An examination of the pattern of dropout and persistence in STEM majors is consistent with our claims. Many students who declare STEM majors fail to complete them, and dropout appears to be greatest relatively early in the academic career. For example, in a study of over 140,000 students at Ohio Universities, Price (2010) found that more than 40% did not complete the STEM major and either dropped out of college all together or switched to non-STEM majors (and completed them). Moreover, a survival curve analysis of dropout and persistence in engineering indicates that dropout is most likely to occur in or around the third semester (Min et al., 2011). We hypothesize that students with low spatial skills initially do poorly but often persist for a semester or two, hoping that the situation will improve. However, after a semester or two, they come to conclude that they should leave the STEM major.

These data are obviously only correlational and certainly do not prove that low spatial abilities are a frequent cause of dropout in STEM fields. Certainly there are many other possible causes, ranging from the harsher

grading practices in STEM fields to the lack of availability of role models (e.g., Price, 2010). We claim only (a) that the observed data are quite consistent with our model of when and why spatial skills matter, and (b) that the influence of spatial skills on the pattern of STEM success and failure merits closer attention and additional research.

We have now made the case for when and why spatial training could help improve STEM learning and retention. We are now ready to address the next logical question: does spatial training really work, and if so, how and why? Why have prior researchers reached such differing conclusions regarding the effectiveness of spatial training?

8. The Malleability of Spatial Thinking

The assumption that spatial training could improve STEM attainment is predicated upon the assumption that spatial skills are, in fact, malleable. This issue also turns out to be a contentious one. Therefore, before concluding that spatial training could facilitate STEM attainment, we need to make sure that training actually works—that it leads to meaningful and lasting improvements in spatial abilities.

Many studies have demonstrated that practice does improve spatial thinking considerably (e.g., Sorby & Baartmans, 1996, 2000; Wright et al., 2008). However, many researchers have questioned whether the observed gains are meaningful and useful for long-term educational training. For example, one potential limitation of spatial training is that it may not transfer to other kinds of experience. Does training gained in one context payoff in other contexts? If spatial training does not transfer, then general spatial training cannot be expected to lead to much improvement in STEM learning. In fact, a summary report of the National Academies of Science (2006) suggested that training of spatial skills was not likely to be a productive approach to enhancing spatial reasoning specifically because of the putatively low rates of transfer.

A second potential limitation of spatial training is the time course or duration of training. While it may be easy to show gains from training in a laboratory setting, these gains will have little, if any, real significance in STEM learning if they do not endure outside of the laboratory. Most lab studies of spatial training last for only a few hours at most, with many lasting less than an hour (e.g., the typical experiment in which an Introductory Psychology student participates). Thus, to claim that spatial training could improve learning in real STEM education, we need to know that it can endure, at least in some situations.

A third potential problem concerns whether and to what extent it is the training, per se, that produces the observed gains. Many training studies use

a pre-test/post-test design, in which subjects are measured before and after training. It is well known that simply taking a test two or more times will lead to improvement; psychologists call this the test-retest effect. Thus, observed effects of training could well be confounded with the improvement that might result from simply taking the test two or more times. Thus it is critically important to have rigorous control groups to which to compare the observed effects of training. At the very least, the control group needs to take the same tests as the treatment group, at least as often as the training group does. Some researchers (e.g., Sims and Mayer, 2002) have claimed that when these sorts of control are included, the effects of training fall to non-significant levels. These researchers included multiple forms of training but also multiple forms of repeated testing in the control group. Both the training and control groups improved substantially, with effect sizes of the training effects exceeding 1 standard deviation. However, these levels were observed both in the control and the treatment groups, and hence despite the large levels of improvement, the specific effect of training relative to the control group, was not statistically significant. In summary, test-retest effects are always an important consideration in any analysis of the effects of educational interventions but they may be particularly large in the area of spatial training. Hence any claims regarding the effectiveness of spatial training interventions need to include careful consideration of control groups, the type of control group used, and the magnitude of improvement in the control group.

8.1. Meta-Analysis of the Effects of Spatial Training

Against this backdrop, we began a systematic meta-analysis of the most recent 25 years of research on spatial training. The meta-analysis had three specific goals. The first was to identify the effectiveness, duration, and transfer of spatial training. The second was to try to shed light on the variation that has been reported in the literature. Why do some studies (e.g., Sorby et al.) claim large effects of training, while others (e.g., Sims and Mayer, 2002) claim that training effects are limited or even non-significant when compared to appropriate control groups. Third, we sought to identify which kinds of training, if any, might work best and might provide the foundation for more systematic investigations of effectiveness and, eventually, larger-scale interventions that ultimately could address spatial reasoning problems.

We note that there have been some prior meta-analyses of spatial training, although these are now rather dated and limited. For example, Baenninger and Newcombe (1989) investigated a more specific question, that is, whether training could reduce or eliminate sex differences in spatial performance. These researchers found that training did lead to

significant gains, but that these gains were largely parallel in the two sexes; men and women improved at about the same rate. Training therefore did not eliminate the male advantage in spatial performance, although it did lead to substantial improvement in both men and women.

We surveyed 25 years of published and unpublished literature from 1984 to 2009. These dates were selected in part because they start when Baenninger and Newcombe's meta-analysis was completed. There has been a tremendous increase in spatial training studies, and therefore a new meta-analysis was in order. Moreover, our goal was substantially broader than Baenninger and Newcombe's goal: we did not limit our literature search to the issue of sex differences and thus would include studies that either included only males or females or that did not report sex differences. Moreover, we specifically focused on transfer and duration of training.

8.1.1. Literature Selection and Selection Criteria

The quality and usefulness of the outcomes of any meta-analysis depends crucially upon the thoroughness of the literature search, and this must include a search for both published and unpublished work. The specific details of the search and analyses methods are beyond the scope of this paper; readers are encouraged to see Uttal et al. (Manuscript accepted for publication) for further information. In addition to searching common electronic databases, such as Google Scholar and PsychInfo, we also searched through the reference lists of each paper we found to identify other potentially relevant papers. Moreover, we contacted researchers in the field, asking them to send both published and unpublished work.

We used a multi-stage process to winnow the list of potentially relevant papers. We sought, at first, to cast a wide net, to avoid excluding relevant papers. At each stage of the process, we read increasing amounts of the article. One criterion for inclusion in the analysis was reference to spatial training, very broadly defined, and to some form of spatial outcome measure. We did studies that focused only on navigational measure. We did not consider studies of clinical populations (e.g., Alzheimer patients) or non-human species.

The first step of the literature search yielded a large number (several thousand) of hits, and it was at this point that human reading of the possible target articles began. At this second step, at least two authors of the paper read the abstract of the paper to determine if it might be relevant. The coders were again asked to be as liberal as possible to ensure that as few relevant articles were missed. If, after reading the abstract, any coder thought the paper might be relevant, then the article was read in its entirety.

In summary, this process yielded a total of 206 articles that were included in the meta-analysis. Approximately 25% of the articles were unpublished, with the majority of these coming from dissertations.

Dissertation abstracts international thus was an important source of unpublished papers (If the dissertation was eventually published, we used the published article and did not include the actual dissertation in the paper).

We then read each article and coded several characteristics, such as the kinds of measures used, the type and duration of training used, the age of the participants, and whether any transfer measures were included. There was substantial variety in the kinds of training that were used, with some studies using intensive, laboratory-based practices of tasks such as mental rotation, while others used more general classroom interventions or full-developed training programs.

We converted reported means and standard deviations to effect sizes, which provide standardized measures of change or improvement, usually relevant to a control group in a between-subjects design or a pre-test score in a within subjects design. Effect sizes compare these measures in terms of standard deviation units. For example, an effect size of 1.0 would mean that training led to an improvement of one standard deviation in the treatment group, relative to the control group. The effect sizes were weighted by the inverse of the number of participants, so that larger studies would have greater influence in calculating the mean effect size and smaller studies would have less influence (Lipsey and Wilson, 2001).

As is likely in any meta-analysis, there was some publication bias in our work; effect sizes from published articles were higher than those from unpublished articles. However, the difference was not large, and the distribution of effect sizes from both sources was reasonably well distributed.

8.1.2. Overall Results

The results of our meta-analysis indicate that spatial training was quite effective. The overall mean effect size was .47 ($SD = .04$), which is considered a moderate effect size. Thus spatial training led, on average, to an improvement that approached one-half a standard deviation. Moreover, some of the studies demonstrated quite substantial gains, with many exceeding effect sizes of 1.0. This meta-analysis thus clearly establishes that spatial skills are malleable and that training can be effective.

In addition, the meta-analysis also sheds substantial light on possible causes of the variability in prior studies of the effects of spatial training. Why have some studies claimed that spatial abilities are highly malleable, while others have claimed that training effects are either non-existent or at best fleeting? One factor that contributes substantially to variability in findings is the presence and type of control group that is used. Researchers used a variety of experimental designs; most used some form of a pre-test/post-test design, measuring spatial performance both before and after training. Many, but not all, of these studies also included some form of control group that did not receive training or received an alternate, non-spatial training (e.g., memorizing new vocabulary words). In some cases,

both the experimental and control groups received multiple spatial tests across the training period. In many cases, we were able to separate the effects of training on experimental and control groups and to analyze separately the profiles of score changes in the two groups.

Two important results emerged from this analysis. First, as expected, experimental groups improved substantially more than control groups did. Second, improvement in the control groups was often surprisingly high, often exceeding an effect size of .40. We believe that much of the improvement was due to the influence of taking spatial tests multiple times. Those control groups that received multiple tests performed significantly better than control groups that received only a pre-test and post-test measure. The magnitude of improvement in the control group often affected the overall effect size of the reported difference between experimental and control groups. For example, a strong effect of training might seem small if the control group also improved substantially. In contrast, a week control group, or no control group, could make relatively small effects of training look quite large. We concluded that the presence and kinds of control groups substantially influenced prior conclusions about the effectiveness of training. Only a systematic meta-analysis that separated experiment and control groups could shed light on this issue.

8.1.3. Duration of Effects

We coded the delay between training and subsequent measures of the effectiveness of training. We measured the length of the delay in days. The distribution of delays was far from normal; it was highly skewed toward studies that included no delays or very short delays, often less than one hour. Most studies had only a small delay, with a mean of one hour or less. However, some studies did include much longer delays, and in these selected studies, the effects of training persisted despite the delay. Of course, these studies may have used particularly intensive training because the researchers knew that the participants would be tested again after a long delay. Nevertheless, they do at least provide an existence proof that training can endure.

8.1.4. Transfer

The issue of transfer is critically important to understanding the value of spatial training for improving STEM education. Training that is limited only to specific tasks and does not generalize will be of little use in improving STEM education. We defined transfer as any task that differed from the training. We also coded the *degree* of transfer, that is, the extent to which the task differed from the original. However, those that did include transfer measures found significant evidence of transfer. Tasks that were very similar to the original (e.g., mental rotation with two- versus three-dimension figures) would be classified as near transfer, but those that

involved substantially different measures would be classified as farther transfers (see Barnett & Ceci, 2002, for further discussion of the definitions of range of transfer).

Although only a minority of studies included measures of transfer, those that did found strong effects of transfer. In fact, the overall effect size for transfer studies did not differ from the overall effect of training. That is, in those studies that did include measures of transfer, the transfer measures improved as much on average as the overall effect size for training. Of course, as in the analysis of the duration of training, we need to note that studies that test for transfer are a select group. Nevertheless, they clearly indicate that transfer of spatial training is possible.

8.2. Is Spatial Training Powerful Enough to Improve STEM Attainment?

Finally, we need to address one more challenging question: could spatial training make enough of a difference to justify its widespread use? We found that the average effect size was approximately .43, but it is important to point out that individuals who go into STEM fields often have spatial ability scores that are substantially greater than $+.43$ *SD*. Thus it seems unlikely that spatial training would make up *all* of the difference between, for example, engineers and students who go into less spatially-demanding fields.

We have several responses to this concern. The first is that educators would be unlikely to choose a training program with average effects. Instead, they would select those that have consistently better than average effects, and there were several with effect sizes approaching 1.0 or greater. Moreover, the type of training implemented would likely not simply be an off-the-shelf choice; developing and implanting effective at scale would be an iterative process, during which existing programs would be refined and improved.

Second, we note that deciding whether an effect size is "big enough" to make a practical difference is often more a question of educational policy and economics than about psychology. Some effect sizes are very small but have great practical importance. For example, taking aspirin to reduce the odds of having a heart attack is now a well-known and accepted intervention, and millions of Americans now follow the "aspirin regimen". But the effect size of the aspirin treatment, relative to placebo, is actually quite small, and in some studies is less than .10. For every 1000 people taking aspirin, only a few heart attacks are prevented. Simply looking at the effect size, one might conclude that taking aspirin just doesn't work. However, because small doses of aspirin are very safe, the benefits are substantially greater than the risks. When distributed across the millions of people who take aspirin, the very small effect size has resulted in the prevention

of thousands of heart attacks. Thus, while spatial training will not prevent all of the dropout from STEM majors, we believe that it will increase the odds of success enough to justify its full-scale implementation, particularly given the relatively low cost of many effective programs.

Relatedly, we can be precise in estimating how much of an improvement an effect size difference of .43 would make. Wai et al. (2010) have given us very precise information about how much those in STEM careers differ from the mean. Given the properties of normal distributions—that most individuals are found near the middle and relatively few are found at the extremes — even relatively modest changes can make a big difference. Implementing spatial training, and assuming our mean effect as the outcome of this implementation, we would shift the distribution of spatial skills in the population by .43 to the right (i.e., increase the z-score of the spatial abilities of the average American students from 0 to +.43). Using Wai, Lubinski, and Benbow's finding that engineers have on average a spatial z-score of approximately .60, we found that spatial training could more than double the number of American students who reach or exceed this level of spatial abilities. Although a spatial-training intervention certainly won't solve all of America's problems with STEM, our review and analyses do suggest it could make an important difference, by increasing the number of individuals who are cognitively able to succeed and reducing the number that dropout after they begin.

9. Models of Spatial Training for STEM

The meta-analysis clearly establishes that spatial training is possible, and that at least in some circumstances it can both endure and transfer to untrained tasks. However, very few of these studies included STEM outcomes, and thus we do not know what kinds of spatial training are most effective in promoting STEM learning. There are, however, a few spatial training programs that have specifically addressed the issue of transfer to STEM outcomes.

One example is Sheryl Sorby's training program. We have already mentioned this 10-week course as an example of effective training for a STEM outcome. Here we discuss it in a bit more detail because it is at least somewhat domain-general and because there has been at least some research on its effectiveness both in promoting spatial skills and in promoting STEM persistence.

After noticing that many freshmen students, particularly females, were deficient in spatial visualization ability, a team of professors at Michigan Technological University (MTU), developed a semester-long course

intended to improve spatial visualization ability. The course emphasized sketching and interacting with three-dimensional models of geometric forms (Sorby & Baartmans, 2000). The sequence of topics mirrored the trajectory of spatial development described by Piaget and Inhelder (1967), with exercises in topological relations (spatial relations between objects), preceding instruction in projections (imagining how objects appeared from different view perspectives) and measurement (Sorby & Baartmans, 1996).

In a pilot version of the course, entering freshmen were screened for spatial ability, then randomly assigned "low spatial" students to experimental and comparison conditions. While the experimental group completed a 10-week spatial visualization curriculum, the comparison group had no additional instruction. The experimental group showed significant pre-to-post instruction gains on a battery of psychometric spatial ability tests, and outperformed the comparison group on a number of other benchmarks (Sorby & Baartmans, 2000).

With evidence for the efficacy of the instruction, the spatial visualization training course became a standard offering at MTU. A longitudinal study describing six years of performance data reported nearly consistent pre-to-post instruction gains on psychometric spatial tests among students who completed the spatial visualization course. In addition, students who completed the spatial visualization course were more likely to remain in their original major and complete their degree in a shorter time than those who did not take the course (Sorby & Baartmans, 2000).

A consistent finding from the longitudinal work was that entering male students tended to outperform female students on the screening exam. Motivated by the idea that early spatial visualization training might bolster girl's skills and confidence in STEM material, Sorby investigated whether the spatial visualization course she developed for freshman engineering students would be appropriate for middle school students. In a three-year study, Sorby found that students who participated in the training activities had significantly higher gains in spatial skills compared to the students who did not undergo such training (Sorby, 2009). Girls who underwent the spatial skills training enrolled in more subsequent math and science courses than did girls in a similarly identified comparison group. In a separate study with high school girls, Sorby found no difference in subsequent STEM course enrollments among girls who had participated in spatial skills training compared to those who had not, suggesting that the optimal age for girls to participate in spatial skills training is likely in or around middle school.

Of course, there are many other kinds of spatial training. Some are much less formal than Sorby's program. For example, one potentially promising line of work is the positive influences of playing videogames on spatial abilities. Several studies have now shown that playing videogames

has a strong, positive effect on visual–spatial memory and attention, (e.g., Gee, 2007; Green & Bavelier, 2003, 2006, 2007). It is tempting to say that playing these videogames might potentially help students do better in their early college years, but of course such a conclusion would be premature without additional research.

10. Conclusions: Spatial Training Really Does Have the Potential to Improve STEM Learning

In this final section we review what we have learned and consider when and why spatial training is most likely to be helpful in improving STEM learning. Our conclusion is quite simple: The available evidence supports the claim that spatial training could improve STEM attainment, but not for the reasons that are commonly claimed. The reason spatial abilities matter early on is because they serve as a barrier; students who cannot think well spatially will have more trouble getting through the early, challenging courses that lead to dropout. Thus we think that an investment in spatial training may pay high dividends. At least some forms of spatial training are inexpensive and have enduring effects.

This analysis points clearly to the kinds of research that need to be done. First, and most importantly, we need well-controlled studies of the effectiveness of spatial training for improving STEM. Although there have been many studies of the effectiveness of spatial training on spatial reasoning, very few have looked at whether the training affects STEM achievement (although see Mix & Cheng, in press, for an interesting discussion of the effects of spatial experience on children's mathematics achievement). Ultimately, the most convincing evidence would come from a Randomized Control Trial, in which participants were assigned to receive spatial training or control intervention before beginning a STEM class.

Second, we would need to be sure of the mechanism by which spatial training caused the improvement. Did spatial training specifically work by boosting the performance of students with relatively low levels of spatial performance and thus preventing dropout? A detailed, mixed-method, longitudinal study of progress through a spatial training program and, ultimately of career placement, is critically important to understanding whether spatial training prevents dropout.

Third, and finally, we need to investigate the value of spatial training in younger students. Here we have focused largely on college students, in part because this age range has been the focus of most studies of spatial training.

However, there has also been work on spatial training in younger students, and if effective, starting training at a younger age could convey a substantial advantage.

In conclusion, this chapter has helped to specify and constrain the ways in which spatial thinking does and does not affect STEM achievement and attainment. Spatial abilities matter, but not simply because STEM is spatially demanding. The time is ripe to conduct the specific work that will be needed to determine precisely when, why and how spatial abilities matter in STEM learning and practice.

ACKNOWLEDGEMENTS

REFERENCES

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology*, *117*, 288−318.

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

Barnett, S. M., & Ceci, S. J. (2002). When and where do we apply what we learn?: A taxonomy for far transfer. *Psychological Bulletin*, *128*(4), 612−637.

Benbow, C., & Stanley, J. (1982). Intellectually talented boys and girls: Educational profiles. *Gifted Child Quarterly*, *26*, 82−88.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor analytic studies*. New York: Cambridge University Press Cambridge.

Chase, W., & Simon, H. (1973). Perception in chess. *Cognitive Psychology*, *4*, 55−81.

Duesbury, R., & O'Neil, H. (1996). Effect of type of practice in a computer-aided design environment in visualizing three-dimensional objects from two-dimensional orthographic projections. *Journal of Applied Psychology*, *81*(3), 249−260.

Eley, M. (1983). Representing the cross-sectional shapes of contour-mapped landforms. *Human Learning*, *2*, 279−294.

Fabro, S., Smith, R., & Williams, R. (1967). Toxicity and teratogenicity of optical isomers of thaidomide. *Nature*, *215*, 296.

Gardner, H. (1993). *Frames of mind: The theory of multiple intelligences* (Tenth-anniversary ed.). New York: Basic Books.

Gee, J. P. (2007). *What video games have to teach us about learning and literacy*. (2nd Edition). New York: Palgrave Macmillan.

Gershmehl, P. J., & Gershmehl, C. A. (2007). Spatial thinking by young children: Neurologic evidence for early development and "educability". *Journal of Geography*, *106*(5), 181−191.

Gerson, H., Sorby, S., Wysocki, A., & Baartmans, B. (2001). The development and assessment of multimedia software for improving 3-D spatial visualization skills. *Computer Applications in Engineering Education*, *9*(2), 105–113.

Green, C. S., & Bavelier, D. (2003). Action video game modifies visual selective attention. *Nature*, *423*, 534–537.

Green, C. S., & Bavelier, D. (2006). Enumeration versus multiple object tracking: The case of action video game players. *Cognition*, *101*, 217–245.

Green, C. S., & Bavelier, D. (2007). Action-Video-Game experience alters the spatial resolution of vision. *Psychological Science*, *18*, 88–94.

Halpern, D., & Wai, J. (2007). The world of competitive scrabble: Novice and expert differences in visuopatial and verbal abilities. *Journal of Experimental Psychology*, *13*, 79–94.

Hambrick, D. Z., Libarkin, J. C, Petcovic, H. L., Baker, K.M., Elkins, J., Callahan, C.N., Turner, S. P., Rench, T.A. & LaDue, N. D. (2011). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology, General,* doi: 10.1037/a0025927.

Hambrick, D., & Meinz, E. (2011). Limits on the predictive power of domain-specific knowledge and experience for complex cognition. *Current Directions in Psychological Science*.

Hegarty, M., & Waller, D. (2005). Individual differences in spatial abilities. In P. Shah, & A. Miyake (Eds.), *The Cambridge handbook of visuospatial thinking* (pp. 121–167). New York: Cambridge University Press.

Hegarty, M., Keehner, M., Cohen, C., Montello, D. R., & Lippa, Y. (2007). The role of spatial cognition in medicine: Applications for selecting and training professionals. In G. Allen (Ed.), *Applied Spatial Cognition*. Mahwah, NJ: Lawrence Erlbaum Associates.

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

Holding, D. (1985). *The psychology of chess skill*. New Jersey: L. Erlbaum Assoc.

Hsi, S., Linn, M., & Bell, J. (1997). The Role of spatial reasoning in engineering and the design of spatial instruction. *Journal of Engineering Education*, 151–158.

Kali, Y., & Orion, N. (1996). Spatial abilities of high-school students in the perception of geologic structures. *Journal of Research in Science Teaching*, *33*, 369–391.

Koedinger, K., & Anderson, J. (1990). Abstract planning and perceptual chunks: Elements of expertise in geometry. *Cognitive Science*, *14*, 511–550.

Kozhevnikov, M., Hegarty, M., & Mayer, R. (2002). Revising the visualizer-verbalizer dimension: Evidence for two types of visualizers. *Cognition and Instruction*, *20*, 47–77.

Kozhevnikov, M., Motes, M., & Hegarty, M. (2007). Spatial visualization in physics problem solving. *Cognitive Science*, *31*(4), 549–579.

Kozhevnikov, M., & Thornton, R. (2006). Real-time data display, spatial visualization ability, and learning force and motion concepts. *Journal of Science Education and Technology*, *15*, 111–132.

Kuenzi, J. J., Matthews, C. M., & Mangan, B. F. (2007). Science, technology, engineering, and mathematics (STEM) education issues and legislative options. *Progress in Education*, *14*, 161–189.

Lajoie, S. (2003). Individual differences in spatial ability: Developing technologies to increase strategy awareness and skills. *Educational Psychologist*, *38*(2), 115–125.

Leffingwell, B. (2003). Chirality & bioactivity 1: Pharmocology. *Leffingwell Reports*, *3*, 1–27.

Lesgold, A., Rubinson, H., Feltovich, P., Glaser, R., Klopfer, D., & Wang, Y. (1988). Expertise in a complex skill: Diagnosing X-ray pictures. In M. H. Chi, R. Glaser, & M. J. Farr (Eds.), *The Nature of Expertise* (pp. 311–342). Hillsdale, NJ: Erlbaum Associates, Inc.

Linhares, A., & Brum, P. (2007). Understanding our understanding of strategic scenarios: What role do chunks play? *Cognitive Science*, *31*, 989–1007.

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

Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

McGregor, S., & Howes, A. (2002). The role of attack and defense semantics in skilled players' memory for chess positions. *Memory and Cognition*, *30*, 707–717.

Markman, A., & Gentner, D. (1993). Structural alignment during similarity comparisons. *Cognitive Psychology*, *25*, 431–467.

Mayo, M. (2009). Video games: A route to large-scale STEM education? *Science*, *323*, 79–82.

Miller, A. I. (1984). *Imagery in scientific thought: Creating 20th century physics*. Boston: Birkhauser.

Min, Y., Zhang, G., Long, R., Anderson, T. J., Ohland, M. W. (2011). Nonparametric survival analysis of the loss rate of undergraduate engineering students. *Journal of Engineering Education, 2*, 349–373.

Mix, K. S. & Cheng, Y. L. (in press). The relation between space and math: Developmental and educational implications. To appear in J. Benson (Ed.). *Advances in Child Development and Behavior, Vol. 42*. Elsevier.

National Academy of Sciences. (2006). *Learning to think spatially*. Washington: DC: The National Academies Press.

Orion, N., Ben-Chaim, D., & Kali, Y. (1997). Relationship between earth science education and spatial visualization. *Journal of Geoscience Education*, *45*, 129–132.

Pallrand, G. J., & Seeber, F. (1984). Spatial ability and achievement in introductory physics. *Journal of Research in Science Teaching*, *21*, 507–516.

Piaget, J., & Inhelder, B. (1967). *The child's conception of space*. New York: W. W. Norton.

Price, J. (2010). The effect of instructor race and gender on student persistence in STEM fields. *Economics of Education Review*, *29*, 901–910.

Rochford, K. (1985). Spatial learning disabilities and underachievement among university anatomy students. *Medical Education*, *19*, 13–26.

Russell-Gebbett, J. (1985). Skills and strategies: pupils' approaches to three-dimensional problems in biology. *Journal of Biological Education*, *19*(4), 293–298.

Sanders, M. (2009). STEM, STEM education, STEMmania. *The Technology Teacher*, *20*, 20–26.

Shea, D., Lubinski, D., & Benbow, C. (2001). Importance of assessing spatial ability in intellectually talented young adolescents: A 20-year longitudinal study. *Journal of Educational Psychology*, *93*, 604–614.

Shepard, R. N., & Metzler, J. (1971). Mental rotation of three-dimensional objects. *Science*, *171*(3972), 701–703.

Sims, V. K., & Mayer, R. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology*, *16*(1), 97–115.

Small, M., & Morton, M. (1983). Research in college science teaching: Spatial visualization training improves performance in organic chemistry. *Journal of College Science Teaching*, *13*, 41–43.

Sorby, S., & Baartmans, B. (1996). A course for the development of 3-D spatial visualization skills. *Engineering Design Graphics Journal*, *60*(1), 13–20.

Sorby, S., & Baartmans, B. (2000). The development and assessment of a course for enhancing the 3-D spatial visualization skills of first-year engineering students. *Journal of Engineering Education*, 301–307.

Sorby, S. (2009). Developing spatial cognitive skills among middle school students. *Cognitive Processing*, *10*(Suppl. 2), 312–315.

Sorby, S. A. (2009). Educational Research in Developing 3-D Spatial Skills for Engineering Students. *International Journal Of Science Education*, *31*(3), 459–480.

Stieff, M. (2004). A localized model of spatial cognition in chemistry. (Doctoral dissertation). Northwestern University, Evanston, IL.

Stieff, M. (2007). Mental rotation and diagrammatic reasoning in science. *Learning and Instruction*, *17*, 219−234.

Talley, L. H. (1973). The use of three-dimensional visualization as a moderator in the higher cognitive learning of concepts in college level chemistry. *Journal of Research in Science Teaching*, *10*(3), 263−269.

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

Uttal, D., Meadow, N. G., Hand, L., Lewis, A., Warren, C., & Newcombe N. (Manuscript accepted for publication). The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*.

Wai, Lubinski, & Benbow. (2009). Spatial ability for STEM domains: aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, *101*, 817−835.

Wai, Lubinski, & Benbow. (2010). Accomplishment in science, technology, engineering, and mathematics (STEM) and its relation to STEM Educational Dose: a 25-year longitudinal study. *Journal of Educational Psychology*, *102*, 860−871.

Waters, A., Gobet, F., & Leyden, G. (2002). Visuospatial abilities of chess players. *British Journal of Psychology*, *93*, 257−265.

Watson, J., & Crick, F. (1953). Molecular structure of nucleic acids. *Nature*, *171*, 737−738.

Wood, B. (1999). Visual expertise. *Radiology*, *211*, 1−3.

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

Wu, H., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science Education*, *88*(3), 465−492.