Evaluating Geoscience Students' Spatial Thinking Skills in a Multi-Institutional Classroom Study

Carol J. Ormand,[1,a] Cathryn Manduca,[1] Thomas F. Shipley,[2] Basil Tikoff,[3] Cara L. Harwood,[4] Kinnari Atit,[2] and Alexander P. Boone[2]

ABSTRACT

Spatial thinking skills are critical to success in many subdisciplines of the geosciences. We tested students' spatial skills in geoscience courses at three institutions (a public research university, a comprehensive university, and a liberal arts college, all in the midwest) over a two-year period. We administered standard psychometric tests of spatial skills to students in introductory geology, mineralogy, sedimentology and stratigraphy, hydrogeology, structural geology, and tectonics courses. In addition, in some courses we administered a related spatial skills test with geoscience content. In both introductory and upper level undergraduate geology courses, students' skills vary enormously as measured by several spatial thinking instruments. Additionally, students' spatial skills generally improve only slightly during one academic term, in both introductory and advanced geoscience classes. More unexpectedly, while there was a tendency for high-performing students to be adept at multiple spatial skills, many individual students showed strong performance on tests of one spatial skill (e.g., rotation) but not on others (e.g., penetrative thinking). This result supports the contention that spatial problem solving requires a suite of spatial skills, and no single test is a good predictor of "spatial thinking." © *2014 National Association of Geoscience Teachers*. [DOI: 10.5408/13-027.1]

Key words: spatial skills, spatial thinking, spatial learning, mental rotation, penetrative thinking, disembedding

INTRODUCTION

Geoscientists are quick to describe the important role that spatial skills play in their work, from making observations in the field to interpreting abstract spatial representations of multivariate data (e.g., Libarkin and Brick, 2002; Titus and Horsman, 2009; Piburn et al., 2011; Liben and Titus, 2012; Manduca and Kastens, 2012). The ability to visualize spatial relations—such as object shapes, relative locations, and how these change over time—is a fundamental skill necessary to understand and reason about geoscience concepts. This skill is also necessary to communicate effectively with diagrams that are used pervasively in geoscience and other STEM disciplines. This conclusion comes from both long-term longitudinal studies (e.g., Shea et al., 2001) and small-scale laboratory studies (e.g., Hegarty et al., 2009). Faculty members frequently describe students' difficulty with spatial visualization as one of the barriers to success in geoscience courses (e.g., Reynolds et al., 2006; Rapp et al., 2007; Riggs and Balliet, 2009; Titus and Horsman, 2009). Research in Engineering (Sorby, 2009) shows that curriculum aimed at helping students improve their spatial skills can have a dramatic effect in improving success in courses and in retaining students in the major. There is also some evidence that suggests that students need to attain a threshold level of competence—but not mastery—in spatial thinking in order to succeed in undergraduate STEM programs (Uttal and Cohen, 2012). Thus, it is critically important to understand what spatial skills are fundamental to the geosciences and how best to develop those skills in our students. This research is aimed at the first step: developing our understanding of the role of spatial thinking in geoscience education.

Mental rotation has received significant attention in the cognitive science literature since Shepard and Metzler's (1971) study laying out the argument for an analog-like mental rotation process. In this study, subjects were asked whether two images represented the same object, with one rotated relative to the other, or mirror-image objects. The authors found that the time it took subjects to confirm that two objects were the same increased linearly with the angular difference between the objects, thus suggesting that subjects were solving each problem by rotating a representation of the object in the diagram. Subsequent studies have investigated the effect of gender and age differences in mental rotation (e.g., Vandenburg and Kuse 1978; Jansen and Heil, 2010), learning effects on mental rotation (e.g., Newcombe et al., 1983; Uttal et al., 2013), and the neural basis of mental rotation (e.g., Zacks, 2008).

Perhaps as a result of this attention, mental rotation tests have commonly been used as proxies for spatial reasoning ability. Yet spatial reasoning is not a single ability. Converging recent findings in cognitive science—from cognitive psychology, linguistic psychology, and neuropsychology—argue that a significantly more diverse skill set is required to cover the breadth of spatial thinking. Chatterjee (2008), for example, proposes a basic typology of four classes of spatial visualization skills. Briefly, these four classes involve spatial relations within objects (e.g., the orientation of the c-axis within a quartz crystal or the slope of a cross-

Received 5 April 2013; revised 11 October 2013 and 25 November 2013; accepted 26 November 2013; published online 26 February 2014.

[1]Science Education Resource Center, Carleton College, Northfield, Minnesota 55057, USA
[2]Department of Psychology, Temple University, Philadelphia, Pennsylvania 19122, USA
[3]Department of Geosciences, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA
[4]Geology Department, University of California-Davis, Davis, California 95616, USA
[a]Author to whom correspondence should be addressed. Electronic mail: cormand@carleton.edu

TABLE I: Number of study participants from each course.

	Liberal Arts College	Comprehensive University	Research University
Introductory Geology	32 (Spring 2010)		130 (Spring 2010)
	9 (Winter 2009)		
Hydrogeology	8 (Winter 2009)		
Mineralogy	19 (Winter 2009)		
Sedimentology & Stratigraphy		12 (Spring 2010)	
Structural Geology	21 (Winter 2009)		17 (Spring 2010)
Tectonics	15 (Winter 2010)		

bed) and relations between objects (e.g., the relative locations of outcrops or the orientation of bedding relative to metamorphic foliation), with static and dynamic versions of each of those categories. As a result of the research emphasis on rotation, the majority of the research on spatial skills in the context of STEM education has focused on 2D to 3D visualization and mental transformations (rotation and folding). Only a small body of work in cognitive science of education has studied any of the other geoscience-relevant spatial skills (e.g., Kastens and Ishikawa, 2006).

One example of a spatial skill that is used widely in geology is visualizing penetrative relations, such as imagining the interior of an object. Research on individual differences in ability to visualize penetrative relations has found a broad range of skills across individuals and a consistent effect of gender (Kali and Orion, 1996; Hegarty et al., 2009). On average, males outperform females in measures of penetrative thinking. Hegarty et al. (2009) report effect sizes of 0.5 and 0.7 in their studies (males are, on average, one-half to seven-tenths of a standard deviation better than females). To put this result in perspective, the difference is comparable to the most robust gender effects previously reported for spatial skills, which are well established for mental rotation (Newcombe et al., 1983). Furthermore, there is a pronounced effect of age on some spatial abilities; mental rotation begins to decrease, dramatically, around the age of 30 (Vandenberg and Kuse, 1978). In addition, while there appears to be some shared variance, penetrative thinking is not the same as mental rotation and may require different cognitive processes. Measures of the two skills correlate only 0.5 overall, and only 0.3 when shared variance associated with reasoning ability is factored out (Hegarty et al., 2009). Thus, a student may excel at mental rotation but still struggle with other spatial tasks.

Relatively little work has yet been done quantifying geoscience students' spatial skills and the impact of geoscience courses on those skills. Our goals for this study were to determine what spatial skill levels students bring to undergraduate geoscience classes, how instruction in geoscience courses affects students' spatial skills, to what extent the different components of spatial thinking correlate (e.g., if a student excels at mental rotation, how likely is it that she will excel at penetrative thinking?), and to what extent spatial skills correlate with success in geoscience courses.

Spatial Skills and Tests

We have focused on three types of spatial thinking skills for this study: mental rotation, penetrative thinking, and disembedding. While these are not the only important spatial skills in the geosciences, we do see pervasive applications of these skills in the geoscience curriculum. For example,

- Mental rotation (visualizing the effect of rotating an object) is essential for understanding crystal symmetry, the use of stereonets in structural geology, and the motions of tectonic plates around Euler poles.
- Penetrative thinking (visualizing spatial relations inside an object) is key to visualizing a slice through any object at any scale. This skill is essential to understanding such diverse topics as mineral dislocations, sedimentary deposits, groundwater flow, structural cross-sections, ocean circulation patterns, and mantle tomography.
- Disembedding (isolating and attending to one aspect of a complex display or scene) is essential any time one needs to find patterns in noisy data, such as when interpreting seismic reflection profiles, stratigraphic sections, or paleoclimate data. However, it can also be critical in tasks as simple as attending to the geologically important features in an outcrop while ignoring nongeologic features.

Study Populations and Settings

We tested students' spatial skills in Introductory and Structural Geology classes at a top-tier public research university; in Introductory Geology, Hydrogeology, Mineralogy, Structural Geology, and Tectonics classes at a private liberal arts college; and in a Sedimentology and Stratigraphy class at a private comprehensive university, all in the midwest. The numbers of participants in this study from each course are shown in Table I.

In most of the analyses that follow, the students in upper-level courses at the liberal arts college are considered as a single population. This simplification was necessary because: (1) Many of the students in upper-level courses were cross-enrolled in another of these courses; and (2) the student population in each of these courses includes a range of experience levels, from sophomores to seniors.

Although overall the number of male and female participants in the study was nearly equal, they were not evenly distributed in each classroom. Table II shows the gender distributions of the study participants in each course. All participants were between 18 and 30 years old. We did not collect data on participants' race or ethnicity.

METHODS

Data Collection

We administered various measures of spatial thinking skills as pre- and post-tests in each classroom participating

TABLE II: Gender demographics.

	Liberal Arts College	Comprehensive University	Research University
Introductory Geology	67% female, 33% male		45% female, 55% male
Hydrogeology, Mineralogy, and Structural Geology	61% female, 39% male		
Sedimentology & Stratigraphy		58% female, 42% male	
Structural Geology			29% female, 71% male
Tectonics	67% female, 33% male		

in the study. Pretests were administered during the first week of classes and post-tests were administered during the last week of classes. Institutional Review Boards approved our study at all three institutions; only students who signed informed consent forms took the tests. We also asked the participants for permission to request their course grade and cumulative GPAs from the registrar, to analyze the relationship between spatial skills, overall academic success, and success in geoscience courses.

For comparison purposes, we administered the same tests in a laboratory setting, with a 3 to 4-week interval between pretest and post-test, an interval dictated by the need to have participants return to the laboratory during the same semester they took the pretest. Participants in this group were 27 students enrolled in an undergraduate psychology course at a research university. None of them had prior experience in geology. This population is not a control group for the students in our study, per se. Rather, we used their paired scores on pre- and post-tests to assess the test–retest effect for each instrument. That is, we measured how much improvement can be expected from pretest to post-test simply from taking the test twice, with no instructional intervention. This is important because significant improvement occurs on some tests simply from taking the test multiple times (Uttal et al., 2013).

To provide a baseline for comparison, we gave the Purdue Visualization of Rotations Test (Guay, 1976) to every participant in the study. Items in this test consist of line drawings of geometric figures, in logic statements of the form, "(First object) is to (first object, rotated) as (second object) is to . . . ." All five of the possible answers are diagrams of the second object, rotated. The test-taker is to select the letter corresponding to the second object that has been rotated *in the same manner* as the first object. There are ten items on the test, and each item is worth one point, with no penalty for incorrect answers. See Fig. 1(a) for an example question from the Purdue Visualization of Rotations Test (PVRT).

In 2010, we also administered the Educational Testing Service (ETS) Hidden Figures test (Ekstrom et al., 1976), a test of disembedding skills, thus providing a second point of comparison for students in those courses. This test requires the participant to identify which of five geometric figures is hidden within each diagram of horizontal, vertical, and diagonal lines. There are sixteen items on the test, and each item is worth one point, with no penalty for incorrect answers. Performance on tests of this skill is correlated with persistence in the sciences, including interest in STEM careers, choice of a STEM major in college, completion of a degree in a STEM discipline, and choice of a career in a STEM field (see Witkin et al. (1977) for a review). See Fig. 1(b) for an example of the type of question in the ETS Hidden Figures test. Table III shows which additional measures we used in each course; all of these measures are described and discussed below.

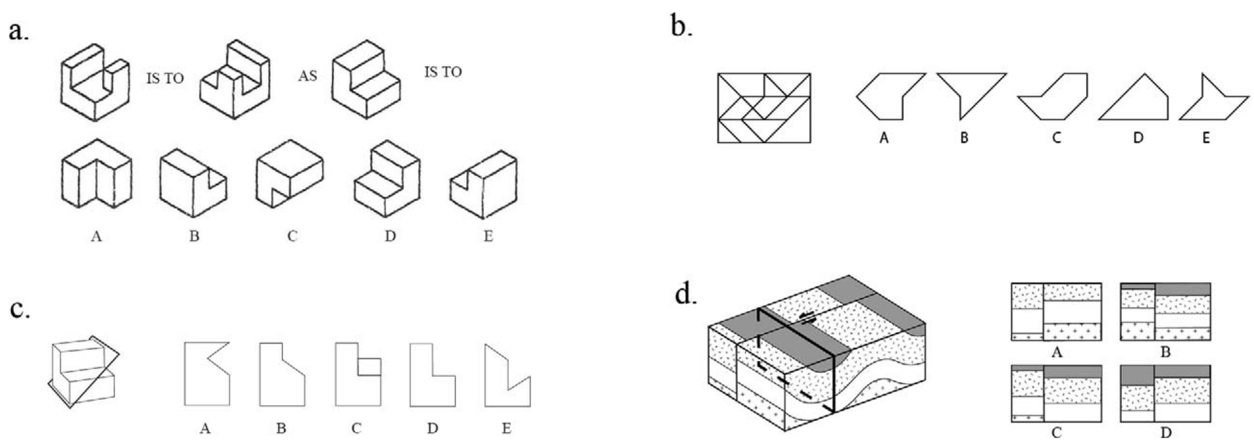

FIGURE 1: Examples of the types of questions from the spatial skills tests used in this study. (a) The Purdue Visualization of Rotations Test. Subjects are asked to identify what the object at the top right would look like if rotated in the same fashion as the first object. (b) Question in the style of the ETS Hidden Figures test. (The actual test questions are copyrighted; this example was drafted by the first author.) Participants are asked to identify which of the five shapes, A through E, can be found in the figure on the left. (c) The Planes of Reference test. Subjects are asked to identify the correct shape of the intersection of the plane and the object. (d) Our Geologic Block Cross-sectioning Test. Subjects are asked to identify the correct cross-section.

TABLE III: Spatial thinking measures administered in each course.

Term	Courses	Liberal Arts College	Comprehensive University	Research University
Winter, 2009	Introductory Geology, Hydrogeology, Mineralogy, and Structural Geology	Purdue Visualization of Rotations		
Winter and Spring, 2010	Introductory Geology	Purdue Visualization of Rotations, ETS Hidden Figures		Purdue Visualization of Rotations, ETS Hidden Figures
	Sedimentology & Stratigraphy		Purdue Visualization of Rotations, ETS Hidden Figures	
	Structural Geology			Purdue Visualization of Rotations, ETS Hidden Figures, Planes of Reference, Block diagrams
	Tectonics	Purdue Visualization of Rotations, ETS Hidden Figures, Planes of Reference, Block diagrams		

To test penetrative thinking ability, we used the Planes of Reference test (Titus and Horsman, 2009). This test consists of items from Crawford and Burnham (1946), Myers (1953), and Titus and Horsman (2009). In this test, participants are asked to choose the shape of intersection of a slicing plane with a geometric solid. There are 15 items on the test, and each item is worth one point, with no penalty for incorrect answers. Although not as widely used to study spatial thinking, this test has been used in prior studies of spatial thinking in the geosciences (e.g., Titus and Horsman, 2009) and has obvious surface validity as a measure of skill in visualizing the shape of a slice through a solid. However, it does not measure the ability to visualize the interior of the slice. Therefore, we also developed a geoscience-specific test of penetrative thinking to use in parallel with the Planes of Reference test. We refer to this as the "Geologic Block Cross-sectioning Test," and it consists of a multiple-choice test of the subject's ability to recognize the correct vertical cross-section through a geologic block diagram. This test is inspired and informed by the work of Kali and Orion (1996), who explored high school students' abilities to visualize 3D structures via open-ended block diagrams. Many of the wrong answers in our multiple choice block diagram test are based on the kinds of mistakes Kali and Orion (1996) observed. There are 14 items on this test, and each item is worth one point, with no penalty for incorrect answers. See Figs. 1(c) and (d) for example questions from the Planes of Reference and Geologic Block Cross-sectioning Tests.

Data Analysis

We conducted standard statistical analyses of our data to answer our research questions:

1. What spatial skill levels do students bring to undergraduate geoscience classes?
2. How does taking geoscience courses affect students' spatial skills?
3. To what extent do different components of spatial thinking correlate?
4. To what extent do spatial skills correlate with success in geoscience courses?

For each course and each corresponding set of study participants (the students in that course who took both the pre- and post-test), and for each test administered to that group, we have calculated the

- Average score and standard deviation, pre- and post-,
- Average improvement over the course of the semester,
- *p* values, using a paired, 2-tailed t-test of pre- and post-test scores,
- Effect sizes, using Cohen's d,
- Pearson correlation coefficients for each pair of tests,
- Pearson correlation coefficients for each test and the students' geology course grades, and
- Pearson correlation coefficients for each test and the students' cumulative GPAs.

We also calculated the average improvement over a 3 to 4-week period, on the same pre- and post-tests, for students at a research university who were not enrolled in a geoscience course (and are not geoscience majors). This allowed us to evaluate the test-retest effect for each of these spatial thinking tests, providing a measure of how much improvement could be expected on each test from taking it twice, without any geological instruction between test administrations.

RESULTS

Students' Spatial Skills and Improvements

Table IV shows pre- and post-test averages and standard deviations (normalized as percentages) for all of the classes in our study, while Fig. 2 shows a few representative distributions of pre- and post-test scores.

TABLE IV: Normalized spatial skills average test scores and gains.

Spatial Skill Test	Institution[1] and Course(s)	*n*	Pretest Score (Standard Deviation)	Post-test Score (Standard Deviation)	Gain	*p*-value	Cohen's d
PVRT	LAC: intro geology	41	41.5 (21.2)	50.2 (21.3)	8.8	<0.001	0.41
	LAC: mineralogy, hydrogeology, structure, tectonics	63	60.2 (22.1)	69.7 (22.1)	9.5	<0.001	0.43
	RU: intro geology	130	49.2 (24.0)	56.1 (24.1)	6.9	<0.001	0.29
	RU: structure	17	60.0 (20.9)	63.5 (18.7)	3.5	0.48	
	CU: sed/strat	12	37.5 (11.4)	50.8 (15.6)	13.3	<0.01	1.02
ETS Hidden Figures	LAC: intro geology	41	44.8 (26.8)	51.2 (28.6)	6.4	0.12	
	LAC: tectonics	15	59.2 (28.3)	65.8 (28.1)	6.7	0.11	
	RU: intro geology	130	41.4 (20.6)	54.9 (23.8)	13.6	<0.001	0.61
	RU: structure	17	50.0 (19.8)	58.1 (17.8)	8.1	0.12	
	CU: sed/strat	12	54.2 (26.6)	46.4 (31.0)	−7.8	0.13	
Planes of Reference	LAC: tectonics	15	59.5 (18.7)	69.3 (22.4)	9.8	<0.01	0.49
	RU: structure	17	57.7 (21.3)	67.5 (14.7)	9.8	0.03	0.55
Block diagrams	LAC: tectonics	15	56.6 (17.8)	64.3 (24.1)	7.6	0.07	
	RU: structure	17	73.1 (15.3)	74.4 (17.1)	1.3	0.77	

[1]LAC = liberal arts college; CU = comprehensive university; RU = research university.

In every class involved in our study, students' spatial abilities vary from zero or near zero to perfect or near-perfect scores on a variety of measures, with averages of ~40%–70% and standard deviations on the order of 15%–30% (Fig. 2). Comparison of standard deviations for the pre- and post-tests indicates that the *range* in students' abilities does not systematically change over the course of an academic term (see Table IV). Moreover, there is an equally wide distribution of skill levels in both introductory courses and advanced courses within the geoscience major. Thus, even though advanced undergraduate majors have stronger spatial skills, *on average*, than the less advanced students, a significant portion of majors in advanced courses have weak spatial skills.

Pre- to post-test comparisons show, on average, modest gains on most measures of spatial abilities, where gain is simply the student's post-test score minus their pretest score. For example, average class gains on the PVRT range from 3.5%–13.3% (see Table IV). The only exception to this trend is the Sedimentolgy & Stratigraphy class at the comprehensive university, which showed modest losses on the disembedding test. However, with only 12 students from that course participating in this study, that result is not statistically significant.

We administered the same tests in a laboratory setting at a different research university, with a 3 to 4-week interval between pretest and post-test, to students not enrolled in any geoscience courses. Under those conditions, test-retest gains on the ETS Hidden Figures test and on the Planes of Reference test are comparable to the gains we see in these classroom experiments, and are statistically significant. However, no test–retest effect is found on the Purdue Visualization of Rotations Test or on the Geologic Block Cross-sectioning Test (Table V).

For each combination of institution, course, and spatial skills test, we calculated the probability that students' test scores would show the measured gains, using a paired, two-tailed *t*-test to calculate *p* values (a measure of the probability of obtaining results at least as extreme as those observed). While many of the class sizes are rather small to draw conclusions about the statistical significance of these gains, half of the *p* values are less than 0.05, and most of these are less than 0.01.

Because *p* values are influenced by sample size, we also calculated the effect sizes using Cohen's *d*, a ratio of average improvement to variability in the sample. While *p* values tell us about the *likelihood* of a particular outcome, effect sizes tell us about the *magnitude* of the experimental effect. In general, a Cohen's *d* value of 0.20 is considered small, 0.50 is medium, and 0.80 is considered to be a large effect (Cohen, 1992). Thus, with the exception of the Sedimentolgy & Stratigraphy class at the comprehensive university, which had only 12 participants, our calculated Cohen's *d* values tell us that (where gains are statistically significant) students are making small to medium improvements on these tests. Fig. 2 illustrates how these gains are distributed across an individual class. In general, there is an upward shift in the class distribution of test scores, although in some classes a few individual students earned lower scores on the post-test than on the pretest (for example, see Fig. 3).

Correlations

One advantage of administering multiple spatial thinking tests to a sample of students is that it allows us to determine whether and to what extent these skills are correlated. Statistical analyses reveal moderate to strong correlations between some of the spatial thinking skills we tested. For example, we calculate a Pearson correlation coefficient (R) of 0.56 for post-test scores between the Purdue Visualization of Rotations Test and the Planes of Reference test ($n = 89$; Fig. 4 and Table VI), consistent with previous findings that mental rotation and penetrative thinking are related, but different, skills (Hegarty et al., 2009). However, some spatial skills test scores correlate very

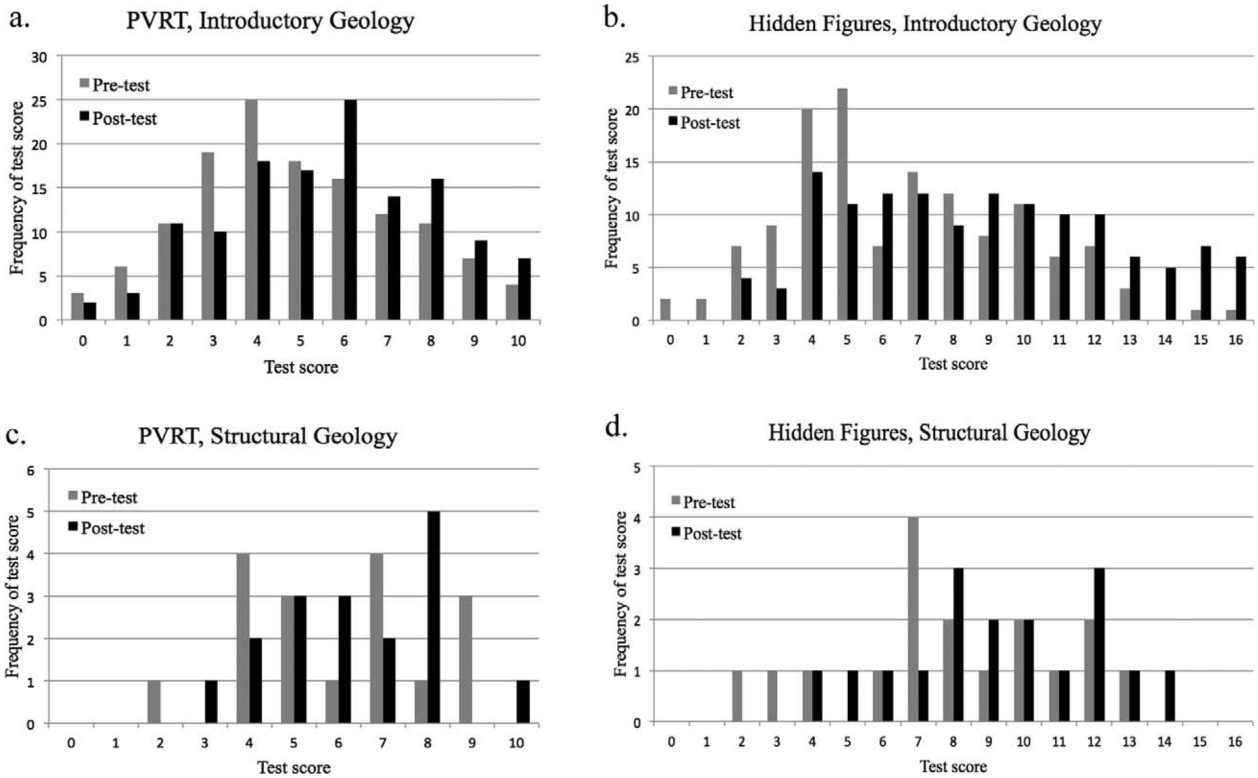

FIGURE 2: Examples of the distributions of student scores on the spatial skills tests used in this study. All data are from classes at the research university. The ***x*****-axis shows the number of questions answered correctly, and the** ***y*****-axis shows the numbers of students in each class getting that score. The left to right shift in distributions of scores from pretest to post-test indicates the improvement in that particular spatial thinking skill, for that set of students. The extremely wide range of spatial skill levels in each class creates a large overlap of pre- and post-test scores. While these distributions are from classes at the research university, they are typical for introductory and upper-level classes in our study. (a) Purdue Visualization of Rotations Test (PVRT), introductory geology class. (b) Educational Testing Service (ETS) Hidden Figures test, introductory geology class. (c) PVRT, structural geology class. (d) ETS Hidden Figures test, structural geology class.**

weakly: the Planes of Reference test and the ETS Hidden Figures test, for example, have a Pearson correlation coefficient of only 0.16 (with $n = 32$; Table VI). This result indicates that penetrative thinking and disembedding abilities are fundamentally different cognitive skills. Consistent with this result, previous research also indicates that spatial visualization skills, such as penetrative thinking, are unrelated to object visualization skills, such as disembedding (Kozhevnikov et al., 2005).

We also found a moderately strong correlation ($R = 0.55$) between post-test scores on the Planes of Reference test and our Geologic Block Cross-sectioning Test ($n = 32$; Fig. 4 and Table VI). Since both of these tests measure students' penetrative thinking skills, one might expect an even higher correlation. However, the Planes of Reference test is a measure of the use of penetrative thinking to imagine the shape of intersection of a plane with a geometric solid, while our Geologic Block Cross-sectioning Test is a measure of the use of penetrative thinking skills to imagine the internal details of a slice through the interior of an object. Thus, we infer that these tests are measuring related but fundamentally different skills. Furthermore, the Planes of Reference test is domain-general; that is, it does not rely on knowledge specific to any field of study. The Geologic Block Cross-sectioning Test, however, is domain-specific, containing geoscience contextual information. Some geoscience students may be able to apply their knowledge of geologic structures and past experience with similar diagrams to deduce the correct answers without mentally visualizing the correct answer. Thus, they may not necessarily be using

TABLE V: Normalized laboratory test-retest average scores and gains.

	n	Test (Std Dev)	Retest (Std Dev)	Gain	p-value
PVRT	27	38.5 (19.0)	44.0 (22.2)	5.5 (18.9)	0.14
ETS Hidden Figures	27	20.7 (25.6)	36.3 (33.4)	15.6 (24.2)	<0.01
Planes of Reference	27	38.0 (16.5)	46.7 (22.6)	8.6 (20.0)	0.03
Block diagrams	27	29.6 (16.8)	32.0 (13.6)	2.4 (17.0)	0.47

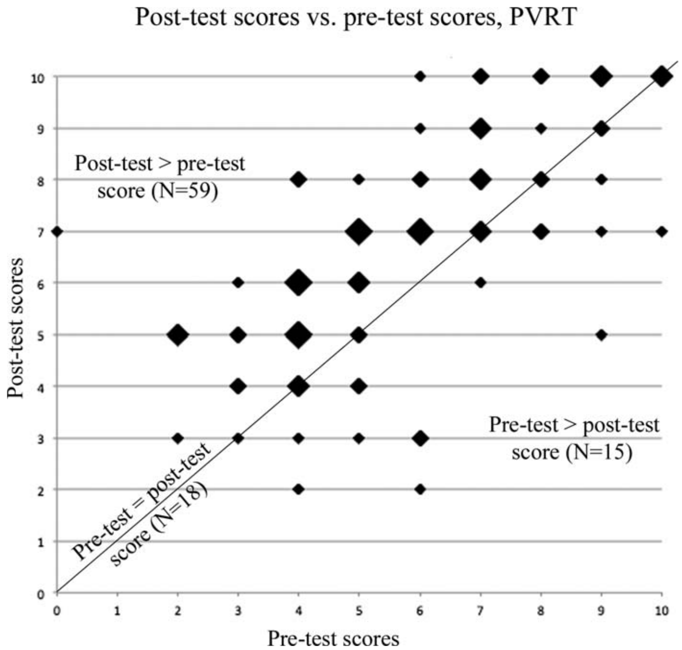

FIGURE 3: Post-test vs. pretest scores on the PVRT for all upper-level geology students participating in our study. The size of the point on the graph indicates the number of students with that pair of pre- and post-test scores. Smallest points represent individual students, slightly larger points represent two students, larger points represent three or four students, and the largest points represent five or six students. The vast majority of students score higher on the post-test than on the pretest ($n = 59$), a few students score the same on the post-test as on the pretest ($n = 18$), and fewer still score lower on the post-test than on the pretest ($n = 15$). The $y = x$ line on the graph separates students who show improvement on the post-test from those who do not.

penetrative thinking skills for this exercise. Use of this domain-specific knowledge may also be contributing to the lack of a stronger correlation between the Planes of Reference and Geologic Block Cross-sectioning Tests.

Finally, we also compared students' spatial thinking skills with their course grades and cumulative grade point averages (Table VI). To our surprise, there are no significant correlations of spatial thinking skills to these measures of academic success. The strongest correlation is a modest correlation between scores on the Geologic Block Cross-sectioning Test and course grade (0.28, $n = 32$). These findings appear to contradict the general conclusion that spatial skills correlate with success in the STEM disciplines (e.g., Shea et al., 2001). There are, however, two possible explanations for this. First, as Shea et al. (2001) point out, course grades depend on a wide array of factors. While spatial thinking is an important component of many geoscience courses, it may be that students with weak spatial skills are compensating by performing well in other aspects of those courses. Second, as suggested in a review of spatial learning in STEM (Uttal and Cohen, 2012), students may require a threshold level of spatial reasoning skill; once above that threshold, other factors, such as working memory capacity and motivation, are more important for success.

DISCUSSION AND IMPLICATIONS

The classroom studies described here demonstrate that students arrive in undergraduate geoscience classrooms with a wide range of spatial thinking skills, from very weak to quite strong. This variation in skill level is not surprising, since the skills that make up spatial reasoning are not explicitly taught in current curricula (NRC, 2006). Class scores in our study *average* in the 40–70% range on a wide variety of instruments, and standard deviations are on the order of 15%–30% (see Table IV). This is true for several different kinds of spatial skills, for students at a variety of institutions, in both introductory and upper-level courses.

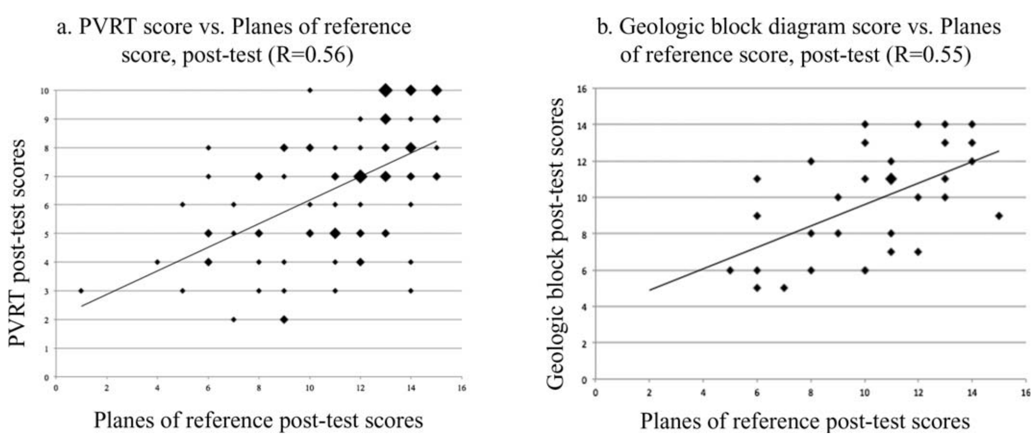

FIGURE 4: (a) Graph of post-test scores on the Purdue Visualization of Rotations Test vs. Planes of Reference test for all students in our study who took both tests ($n = 89$). Although $R = 0.56$, indicating a statistically significant correlation of these two skills, note that some students who excel at one of these skills are very weak in the other. (b) Graph of post-test scores on the Geologic Block Cross-sectioning Test vs. the Planes of Reference test for all students in our study who took both tests ($n = 32$). With $R = 0.55$, these skills are also moderately strongly correlated, with similar scatter. Point size conventions are the same as in Figure 3; the smallest points represent individual students, while each of the largest points represent five or six students.

TABLE VI: Correlations (Pearson's r) between spatial skills post-test scores and measures of student academic success.

	PVRT	ETS Hidden Figures	Planes of Reference	Block Diagrams	Course Grade
ETS Hidden Figures	0.36*; $n = 207$				
Planes of Reference	0.56*; $n = 89$	0.16; $n = 32$			
Block diagrams	0.45*; $n = 32$	0.09; $n = 32$	0.55*; $n = 32$		
Course grade	0.00; $n = 213$	0.12; $n = 177$	0.18; $n = 69$	0.28; $n = 32$	
Cumulative GPA	−0.12; $n = 178$	0.07; $n = 178$	0.10; $n = 32$	0.23; $n = 32$	0.68*; $n = 177$

*$p < 0.02$

This variation in student skill levels presents quite a challenge for geoscience instructors.

On average, students make small to medium gains on these measures over the course of the semester, with an overall average gain of 10%, where gains are statistically significant. However, it is likely that some of these apparent gains are the combined result of improvement in the skill being measured and students taking a similar or the same test a second time (the test-retest effect). In laboratory conditions, with a 3 to 4-week testing interval, test-retest gains on the ETS Hidden Figures test and on the Planes of Reference test are comparable to the gains we see in these classroom experiments, while the Purdue Visualization of Rotations Test and the Geologic Block Cross-sectioning Test show no test-retest effect (Table V). Therefore, actual gains in our classrooms are as measured on the Purdue Visualization of Rotations Test and on the Geologic Block Cross-sectioning Test, but may be smaller than they appear on the ETS Hidden Figures and Planes of Reference tests. It is worth noting, however, that the time interval between testing and re-testing in our classroom studies is typically 2 to 3 months, while the interval between testing in the laboratory conditions is 3 to 4 weeks.

Not every student improves from pretest to post-test; some make no gains and a few perform worse on the post-test than on the pretest (see Fig. 3). These individual "losses" may be attributable to luckier random guessing on the pretests or to students simply having a bad day on the day of the post-test. Indeed, scores on post-tests given during the last week of the semester may be conflated by end of term stress levels and fatigue. In that case, student performance on the post-test may not reflect the strength of their spatial skills, and actual gains may be greater than measured. Pre- and post-test scores on these instruments show that, in general, undergraduate geoscience students' spatial skills have considerable room for improvement and are not strongly affected by geoscience coursework.

One might wonder why geoscience courses do not have a greater impact on students' spatial skills. However, the improvement of spatial thinking skills was neither an explicit learning goal nor an implicit focus for any of the courses involved in this study. This is in contrast, for example, to previous studies of the impact of spatial skills training in geoscience courses, where significant improvement in spatial thinking has been observed (e.g., Reynolds et al., 2006; Titus and Horsman, 2009). Indeed, the cohorts of students in different courses in our study showed different average gains on each of the spatial skills measures. We interpret this as reflecting different emphases on spatial topics and spatial tasks within those courses.

The range of correlations between the various spatial thinking instruments that we used in this study confirms that spatial thinking is multi-faceted. While much of the research literature has focused on mental rotation, one cannot generalize from an individual's mental rotation ability to his or her overall ability to think spatially. Even though various spatial skills do correlate with each other statistically, an individual student may (for example) excel at mental rotation but be unable to imagine what a slice through the interior of an object would look like, or vice versa (Fig. 4). This variation within individual students' spatial skills also presents challenges to instructors.

Analyses of large-scale data sets of spatial skills measured in high school show that performance on standardized psychometric measures of spatial skills predicts success in STEM outcomes: success in STEM majors in college and professional entry into a STEM field (Shea et al., 2001; Wai et al., 2009; Webb et al., 2007). In addition, prior studies have shown that poor spatial skills can be a barrier to learning geoscience (e.g., Rapp et al., 2007; Riggs and Balliet, 2009; Titus and Horsman, 2009). It may be, however, that only a threshold level of spatial competence may be necessary for success (Uttal and Cohen, 2012). Although some of the students in our study are succeeding at the undergraduate level without strong spatial skills, we wonder whether they will be able to continue to do so at the graduate school or professional level. An assessment of spatial thinking skills among geoscience graduate students, while beyond the scope of this study, would be a valuable pursuit.

In contrast to prior studies, our data do not show a correlation between spatial skills and success in geoscience courses. While we do not know why, we can speculate about some possible reasons. Success in geoscience courses depends on many factors, thus confounding any correlation between spatial skills and success. For example, a student with weak spatial skills may nonetheless earn a decent course grade through hard work, strong writing skills, and effective study habits. Likewise, there are many ways to fail (or perform poorly) in a geoscience course. So a student with strong spatial skills may earn a poor course grade through failure to apply him or herself to the coursework, weak communication skills, or poor study habits. In order to disentangle these effects, it would be informative to compare students' scores on the spatial skills tests to their performance on specific, spatially-demanding geoscience tasks.

CONCLUSIONS

Based on the data presented above, we draw the following conclusions:

1. There is a wide range of spatial ability, even for geology majors in upper-level courses.
2. Spatial skills cannot be measured with a single test; a suite of tests is necessary to characterize an individual's spatial skills, and an individual may excel at some spatial thinking skills while struggling with others.
3. Spatial thinking improves with practice.

Undergraduate geoscience education would benefit from identifying the full range of spatial skills involved in learning and doing geoscience (we suspect that we have not tested all key dimensions) and then developing effective teaching materials and strategies for improving those skills in our students. This course of action has the potential to increase the pool of students who are likely to choose to major in geoscience and to strengthen the abilities of those students to think like geoscientists.

Acknowledgments

We are grateful to Bryn Benford, Clint Cowan, Cam Davidson, Laurel Goodwin, Tom Hickson, Jim Ludois, Stephen Meyers, Paul Riley, Josh Roberts, Mary Savina, and Sarah Titus for allowing us to test their students; all of the students who agreed to participate in this study; two anonymous reviewers for thoughtful and helpful comments; and the National Science Foundation for funding this work (SBE #0541957).

REFERENCES

Chatterjee, A. 2008. The neural organization of spatial thought and language. *Seminars in Speech and Language*, 29:226–238.

Cohen, J. 1992. A power primer. *Psychological Bulletin*, 112:155–159.

Crawford, A.B., and Burnham, P.S. 1946. Forecasting college achievement: A survey of aptitude tests for higher education. New Haven, CT: Yale University Press. 291 pp.

Ekstrom, R.B., French, J.W., Harman, H.H., and Dermen, D. 1976. Manual for kit of factor-referenced cognitive tests. Princeton, NJ: Educational Testing Service.

Guay, R. 1976. Purdue spatial visualization test. West Lafayette, Indiana: Purdue Research Foundation.

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

Jansen, P. and Heil, M. 2010. Gender differences in mental rotation across adulthood. *Experimental Aging Research*, 36:94–104.

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

Kastens, K.A., and Ishikawa, T. 2006. Spatial thinking in the geosciences and cognitive sciences: A cross-disciplinary look at the intersection of the two fields. *In* Manduca, C.A. and Mogk, D.W., eds., Earth and mind: How geologists think and learn about the Earth. Boulder, CO: Geological Society of America Special Paper 413, p. 53–76.

Kozhevnikov, M., Kosslyn, S., and Shephard, J. 2005. Spatial versus object visualizers: A new characterization of visual cognitive style. *Memory & Cognition*, 33:710–726.

Libarkin, J.C. and Brick, C. 2002. Research methodologies in science education: Visualization and the geosciences. *Journal of Geoscience Education*, 22:449–455.

Liben, L.S. and Titus, S.J. 2012. The importance of spatial thinking for geoscience education: Insights from the crossroads of geoscience and cognitive science. *In* Kastens, K.A. and Manduca, C.A., eds., Earth and mind II: A synthesis of research on thinking and learning in the geosciences. Boulder, CO: Geological Society of America Special Paper 486, p. 51–70.

Manduca, C.A. and Kastens, K.A. 2012. Geoscience and geoscientists: Uniquely equipped to study Earth. *In* Kastens, K.A. and Manduca, C.A., eds., Earth and mind II: A synthesis of research on thinking and learning in the geosciences. Boulder, CO: Geological Society of America Special Paper 486, p. 1–12.

Myers, C.T. 1953. Validity study of six spatial relations tests. Princeton, NJ: Educational Testing Service; RB-53-15.

Newcombe, N., Bandura, M.M., and Taylor, D.G. 1983. Sex differences in spatial ability and spatial activities. *Sex Roles* 9:377–386.

[NRC] National Research Council. 2006. Learning to think spatially—GIS as a support system in the K–12 curriculum. Washington DC: National Academies Press, 313 pp.

Piburn, M.D., Kraft, K. vdH., and Pacheco, H. 2011. A new century for geoscience education research. Prepared for the National Academies Board on Science Education Committee on the Status, Contributions, and Future Directions of Discipline-Based Education Research.

Rapp, D., Culpepper, S.A., Kirkby, K., and Morin, M. 2007. Fostering students' comprehension of topographic maps. *Journal of Geoscience Education*, 55:5–16.

Reynolds, S.J., Piburn, M.D., Leedy, D.E., McAuliffe, C.M., Birk, J.P., and Johnson, J.K. 2006. The Hidden Earth — Interactive computer-based modules for geoscience learning. *In* Manduca, C.A. and Mogk, D.W., eds., Earth and mind: How geologists think and learn about the Earth. Boulder, CO: Geological Society of America Special Paper 413, p. 171–186.

Riggs, E.M. and Balliet, R.N. 2009. A role for mental rotations in field-based problem solving. Geological Society of America *Abstracts with Programs*, 41(7):195.

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

Shepard, R.N. and Metzler, J.T. 1971. Mental rotation of three dimensional objects. *Science*, 171:701–703.

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

Titus, S.J., and Horsman, E. 2009. Characterizing and improving spatial visualization skills. *Journal of Geoscience Education*, 57:242–254.

Uttal, D.H. and Cohen, C.A. 2012. Spatial thinking and STEM education: When, why and how? *In* Ross, B., ed., Psychology of learning and motivation. Oxford: Academic Press, Vol. 57, p. 147–181.

Uttal, D.H., Meadow, N.G., Tipton, E., Hand, L.L., Warren, C., and Newcombe, N.S. 2013. The malleability of spatial skills: A meta-analysis of training studies. *Psychological Bulletin*, 139:352–402.

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

Wai, J., Lubinski, D., and Benbow, C.P. 2009. Spatial ability for STEM domains: Aligning over fifty years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101:817–835.

Webb, R.M., Lubinski, D., and Benbow, C.P. 2007. Spatial ability: A neglected dimension in talent searches for intellectually precocious youth. *Journal of Educational Psychology*, 99:397–420.

Witkin, H.A., Moore, C.A., Goodenough, D.R., and Cox, P.W. 1977. Field-dependent and field-independent cognitive styles and their educational implications. *Review of Educational Research*, 47:1–64.

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