Spatial Thinking in Physics: Longitudinal Impacts of 3-D Spatial Training

David I. Miller (David.Miller@berkeley.edu)
Graduate School of Education, University of California – Berkeley, Berkeley, CA 94720 USA

Diane F. Halpern (Diane.Halpern@cmc.edu)
Department of Psychology, Claremont McKenna College, Claremont, CA 91711 USA

Abstract

In previous research, we found that twelve hours of 3-D spatial training, compared to a randomized control condition, improved the spatial skills and physics exam scores of gifted science, technology, engineering, and mathematics (STEM) undergraduates (n = 55) directly after training. This paper reports on longitudinal findings of this training study. After eight months, training differences did not exist for spatial skills, physics grades, or physics self-efficacy. Large gender differences, favoring males, existed for some spatial skills, physics self-efficacy, and physics grades. Correlational analyses found that mental rotation performance, not spatial working memory, predicted physics self-efficacy and some physics learning outcomes. These results suggest that sustained exposure to spatially enriching activities over several semesters or years may be necessary to address concerning gender gaps in spatial skills among those most likely to pursue advanced educational and occupational positions in physics.

Keywords: spatial thinking; physics problem-solving; skill acquisition; individual differences

Introduction

Although frequently neglected in traditional education, 3-D spatial skills are critical to success in science, technology, engineering, and mathematics (STEM) fields (Wai, Lubinski, & Benbow, 2009) especially physics (Hake, 2002; Kozhevnikov, Motes, & Hegarty, 2007). Based on decades of evidence, a recent National Science Board (NSB, 2010) report concluded that individuals skilled in spatial thinking are “an untapped pool of talent critical for our highly technological society” (p. 20). Men consistently outperform women on many spatial tasks especially mental rotation (Silverman, Choi, & Peters, 2007; Voyer, Voyer, & Bryden, 1995); these results present alarming concerns to STEM educators and policy makers who aim to increase the proportion of women in STEM fields (Halpern et al., 2007).

Fortunately, recent research has found that spatial experience such as action video games (Feng, Spence, & Pratt, 2007) and formal spatial coursework (Sorby, 2009) can robustly improve spatial skills. However, most prior research has failed to investigate how long these effects last and how spatial training can improve outcomes for students majoring in STEM fields (Uttal et al., under review).

With several self-selected cohorts of undergraduate engineering students, Sorby (2009) found that spatial training was associated with higher engineering retention rates for women and higher grades in future STEM courses including engineering graphics, calculus, and physics. However, Sorby’s studies were mostly quasi-experimental and therefore confounded the effects of self-selection (and hence also motivation, diligence, etc.). Similar differences in GPA and retention rates were found for randomized studies although as Sorby noted, “sample sizes for the randomly selected groups were generally too small to infer statistical significance” (p. 476).

Therefore, in Miller and Halpern (2010), we randomly assigned gifted STEM undergraduates from a highly selective science and engineering college to either a training condition that completed twelve hours of spatial training or a control condition. Results indicated that that twelve hours of spatial training (1) improved the skills to mentally rotate and visualize cross-sections of 3-D objects, (2) narrowed gender differences in spatial skills perhaps because of ceiling effects, and (3) improved examination scores in introductory calculus-based Newtonian physics by nearly one-third of a letter grade (d = 0.38) but not for other STEM courses. This paper reports on longitudinal research that investigated (1) whether these training effects are stable across eight months and (2) how to further explain the improvements in physics examination scores.

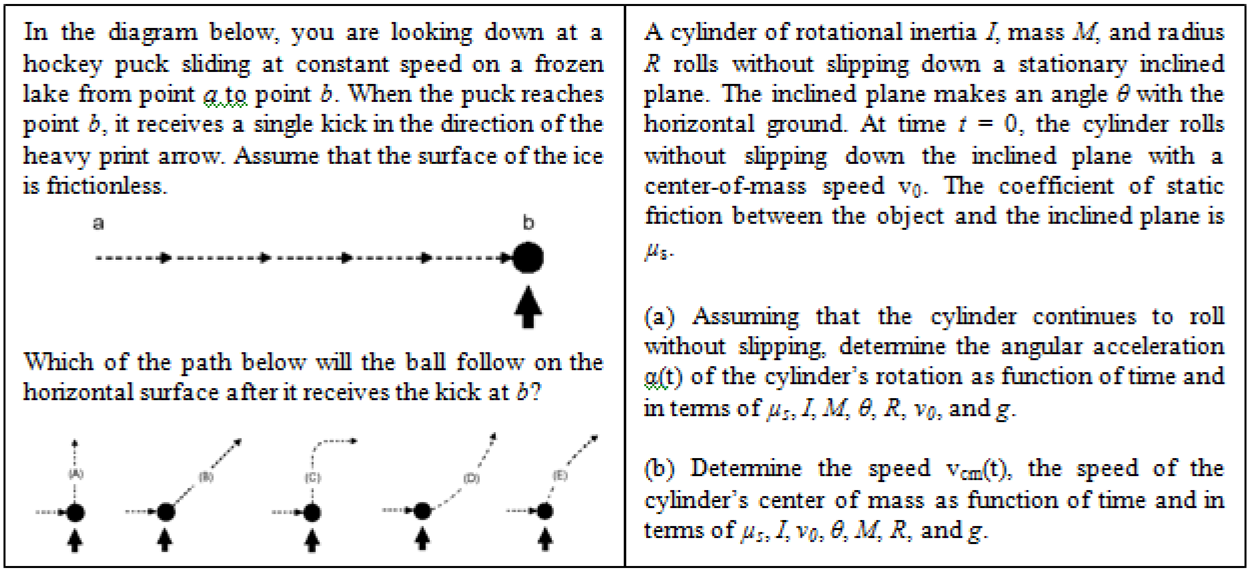
As Uttal et al. (under review) noted, demonstrating stable, durable training effects is critical for designing effective spatial curriculum and educational policies. Out of the 217 research studies reviewed by Uttal et al., only three studies measured spatial performance more than one month after training. These studies are two randomized studies with non-STEM undergraduates (Feng, Spence, & Pratt, 2007; Terlecki, Newcombe, & Little, 2008) and one within-subjects study (with no control condition) with middle school and high school atmospheric science students (Hedley, 2008). These three studies found large, durable training improvements (average d = 0.67) for three to five months after training. Furthermore, these longitudinal studies found little decrement in spatial skills between the immediate and delayed spatial post-tests suggesting these effects could persist for even longer than five months. Perhaps most encouragingly, Terlecki et al. (2008) found durable transfer to untrained spatial tasks (e.g., the stimuli in the transfer test bore little resemblance to the stimuli in the training tasks). These longitudinal results strongly suggest that improvements in spatial skills can be long-lasting up to at least five months, although it remains unclear whether these effects would also generalize to gifted STEM undergraduates. We contribute to the durability literature by measuring spatial performance eight months after training and measuring physics learning outcomes ten months after training.

In this paper, we also aimed to further understand the improvements that we previously found in physics course grades. Kozhevnikov and colleagues (Kozhevnikov, Hegarty, & Mayer, 2002; Kozhevnikov, Motes, Hegarty, 2007; Kozhevnikov & Thornton, 2006) have provided insightful data regarding the role of spatial thinking in physics learning. With quantitative, protocol analysis, and eye fixation data, Kozhevnikov et al. (2007) argued that, "multidimensional physics problems and spatial visualization tasks require the problem solver to simultaneously process multiple pieces of spatial information that tax the supplies of visual/spatial working memory resources" (p. 576). In explaining the correlation between spatial skills and physics problem solving, Kozhevnikov et al. downplayed the importance of the unique skills required for individual spatial skills tests and instead focused their theoretical explanations on the variance that general spatial working memory shares with spatial skills tests and physics problem solving. We call this explanation the working memory hypothesis. Although this remains an interesting hypothesis, Kozhevnikov et al. did not measure spatial working memory and, to our knowledge, no prior study has directly investigated the relationship between spatial working memory and physics problem solving.

Alternatively, the correlation between spatial skills and physics problem solving could be in part explained by common specific strategies and cognitive skills. We call this explanation the specific spatial skills hypothesis. In this vein, the choice of spatial tests and physics outcomes matters greatly when investigating the role of spatial thinking in physics. In Miller and Halpern (2010), we found an interesting result regarding the type of physics outcome we analyzed: for the training group, we found improvements in physics course exam performance, but no improvements on the Force Concept Inventory – a commonly used measure of qualitative, conceptual understanding of physics (Hestenes, Wells, & Swackhamer, 1992). These two types of assessments require substantially different processes for applying physics knowledge; for instance, the course examination questions required extensive application of mathematical problem solving skills including calculus (see Figure 1). In contrast, the conceptual questions assessed students' qualitative, not mathematical, understanding of physics principles. With regards to spatial thinking, the most important differences between the two assessments are: the course examination problems often provided no visual-spatial diagrams of the physical situations (students would have to generate their own diagrams, or attempt to solve the problems without such diagrams) and typically involved more complex motion (including rotational and sometimes 3-D motion) compared to the conceptual questions.

Furthermore, the role of spatial thinking in physics might be highly dependent on specific physics courses and topics. All of the research discussed previously investigated the role of spatial thinking in specifically kinematics which is the physics of how objects move in space over time. Research on other physics topics have suggested that specifically mental rotation may play a small and sometimes nonsignificant role in learning introductory electricity and magnetism (Saglam, & Millar, 2006; Watkins, Dowd, & Mazur, 2010).

This current research investigated the longitudinal impacts of spatial training. We hypothesized that, after eight months, the training group would continue to have higher spatial skills and higher physics grades compared to the



In the diagram below, you are looking down at a hockey puck sliding at constant speed on a frozen lake from point *a* to point *b*. When the puck reaches point *b*, it receives a single kick in the direction of the heavy print arrow. Assume that the surface of the ice is frictionless.

Which of the path below will the ball follow on the horizontal surface after it receives the kick at *b*?

A cylinder of rotational inertia I, mass M, and radius R rolls without slipping down a stationary inclined plane. The inclined plane makes an angle θ with the horizontal ground. At time $t = 0$, the cylinder rolls without slipping down the inclined plane with a center-of-mass speed v_0. The coefficient of static friction between the object and the inclined plane is μ_s.

(a) Assuming that the cylinder continues to roll without slipping, determine the angular acceleration $\alpha(t)$ of the cylinder's rotation as function of time and in terms of μ_s, I, M, θ, R, v_0, and g.

(b) Determine the speed $v_{cm}(t)$, the speed of the cylinder's center of mass as function of time and in terms of μ_s, I, v_0, θ, M, R, and g.

Figure 1: Example qualitative, conceptual kinematics problem (left – from Kozhevnikov, Motes, & Hegarty, 2007) and a typical mathematical, physics examination problem (right).

control group. We collected grades data for introductory electricity and magnetism - the course directly following introductory Newtonian physics. Furthermore, to help better understand the improvements in physics, we hypothesized that the training group would have higher self-efficacy for solving highly spatial physics problems. For instance, it is possible that spatial training could have improved students' self-efficacy for applying spatial strategies on highly spatial physics problems which could have improved physics exam scores (Bandura, 1997; Cooke-Simpson & Voyer, 2007). Finally, to test the working memory and specific skills hypotheses, we had students complete a measure of spatial working memory. The working memory hypothesis would predict that spatial working memory and physics learning outcomes would share unique variance, and the specific skills hypothesis would predict that individual measures of spatial skills and physics learning outcomes would share unique variance.

Method

Participants

STEM undergraduate majors (22 women, 33 men) were recruited during their first-year at a small, highly selective liberal arts college with a strong STEM focus. Forty-nine percent of students' mothers and 60% of students' fathers had received an advanced graduate or professional degree. All participants were either 19 years old (n = 45) or 20 years old (n = 10) at the time of this longitudinal assessment. SAT – Mathematics (M = 762, SD = 39), SAT – Critical Reading (M = 728, SD = 49), and SAT – Writing (M = 710, SD = 57) scores indicated exceptionally high academic aptitude. Furthermore, pre-test scores on standardized measures of spatial skills indicated substantially higher initial spatial skills compared to more average populations (Miller & Halpern, 2010). We choose to focus on such an extremely gifted STEM population because such undergraduates are disproportionately more likely to become future STEM innovators (NSB, 2010). For instance, Wai, Lubinski, and Benbow (2009) found that 45% of all STEM PhDs in their longitudinal study (n = 400,000) were within the top 4% of spatial skills in high school. Our longitudinal subsample represented 71% of the original pre-test sample (n = 77) and missing data analyses indicated that retention rates did not significantly differ in terms of experimental assignment ($\chi_2(1) = 1.17$, $p = .280$), gender ($\chi_2(1) = 1.10$, $p = .294$), initial spatial skills (all F's < 1), or SAT scores (all F's < 1). With a fixed significance level of 0.05, power analyses showed that 23 students per condition yield a statistical power of 80% or greater for detecting an effect size of d = 0.74 or greater for a one-tailed independent samples t-test. Analyses also revealed 72% power for detecting the average effect size of d=0.67 found in previous longitudinal studies.

Materials and Procedure

Students were randomly assigned to a training condition (12 female, 18 male) in which they completed six two-hour spatial training sessions (one session per week), or a control condition (10 female, 15 male) in which they did not. The spatial training heavily emphasized developing spatial skills by sketching 3-D objects (see Figure 2); these materials were developed and tested by Sorby (2009). Sorby has found large improvements in spatial skills with engineering undergraduates with initial low spatial skills. Students completed measures of spatial skills prior to training (pretest), one week after the last spatial training session (immediate posttest), and eight months after training (delayed posttest). This paper reports on the delayed posttest data.

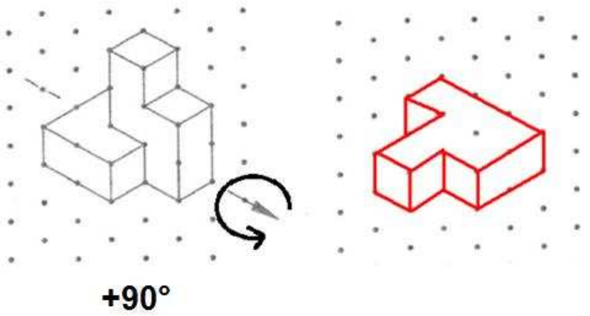


Figure 2: Sample workbook problem from the spatial training. On 2-D sketch paper, students are asked to mentally rotate the left 3-D object 90 degrees around the indicated axis and sketch the correct rotation (shown in red) on the dot paper to the right (from Sorby & Wysocki, 2003).

Measures

Spatial skills – Mental rotation. The Mental Rotation Test (Peters et al., 1995, Form C) measured students' skills to mentally rotate 3-D objects about two or more axes. Students completed Form A of the Mental Rotation Test during pre- and post-testing and hence have not encountered the specific stimuli on Form C before. For the immediate posttest, we previously found greater improvements on Form A for the training group. The test consisted of 24 problems.

Spatial skills – Mental cutting. The Mental Cutting Test (CEEB, 1939) and Novel Cross-Sections Test (Hegarty et al., 2009) measured students' skills to visualize cross-sections of 3-D objects cut by a specified 2-D plane. Students completed the Mental Cutting Test during pre-testing and post-testing and hence were familiar with the specific stimuli on that test, but not familiar with the Novel Cross-Sections Test. For the immediate posttest, we had found greater improvements on the Mental Cutting Test for the training group. We included the Novel Cross-Sections Test to test whether spatial training improved the construct of mental cutting, not such test performance on one specific test. Each test consisted of 10 problems.

Spatial working memory. The Spatial Working Memory test (Kane et al., 2004, complex rotation span) measured participants' cognitive capacity to simultaneously process and store novel spatial information. On computers, students judged whether a set of individually presented letters were normal or mirror-imaged (processing task) while simultaneously remembering the locations of a sequence of short and long arrows radiating from the center of a computer screen (storage task). At the end of a trial, the students recalled the positions of the arrows in the order they were presented. Set sizes ranged from two to six letter-arrow displays per trial (with 3 trials per set size for 15 trials total). We scored the recall data using the partial credit procedure advocated by Conway et al. (2005).

Physics outcomes. Final grades in introductory electricity & magnetism were converted to numerical scores by assigning "A" = 4.0, "A-" = 3.667, and so on; these grades reflect student work completed six to ten months after training. From the previous semester, we also had physics examination scores and pre/post measures of physics conceptual understanding (Hestenes, Wells, & Swackhamer, 1992, Force Concept Inventory) for introductory Newtonian physics/kinematics. For this previous semester, physics examination scores contributed 80% of the final mechanics course grades; these mechanics scores reflect student work completed during training to two months after training.

Physics self-efficacy. Three different Likert scales measured student's self-efficacy for solving physics problems. Two of these scales asked for students' strength of agreement with different statements such as, "If I get stuck on a physics problem on my first try, I usually try to figure out a different way that works" (Adams et al., 2006, Colorado Learning Attitudes about Science Survey – Problem Solving Confidence Subscale, 4 items; Çalişkan, Selçuk, & Erol, 2007, Physics Self-Efficacy Scale – Solving Physics Problems Subscale, 10 items). Furthermore using recommendations specified by Bandura (1997), we constructed a physics problem solving self-efficacy scale by presenting students with four math-intensive kinematics problems (the problem on the right of Figure 1 is one of those problems) and asked students to rate on a 1-10 scale their confidence in "correctly solving the above problem during a physics examination in which an equation sheet is provided." All the physics self-efficacy scales showed good internal consistency (α's = .74 to .93) and were highly correlated with one another (r's = .64 to .87) demonstrating convergent validity. All of the scales significantly correlated with SAT – Mathematics scores (r's = .34 to .41, all p's < .01) which is expected since physics problem-solving is often highly mathematical. However, none of the scales significantly correlated with SAT – Critical Reading and SAT – Writing scores (r's = -.01 to .23, all p's > .05) demonstrating discriminate validity. Furthermore, all self-efficacy scales highly correlated with the physics outcomes described above (r's = .49 to .71) demonstrating criterion validity. Because of these solid psychometric properties, we computed a composite physics self-efficacy scale by summing the standardized z-scores for each scale.

Results

Table 1 contains intercorrelations, internal consistencies (Cronbach's α), and descriptive statistics for all four spatial measures. Notice the low internal consistencies for the Mental Cutting Test and Novel Cross-Sections Test. Furthermore, those two scales were only modestly correlated with one another (r = .30) although they aim to measure the same construct. For these reasons, we interpret results with those two measures cautiously.

Measure	1	2	3	4
1. MRT				
2. MCT	.41**			
3. NCST	.41**	.30*		
4. SWM	.48**	.10	.44**	
α	.81	.57	.61	.78
M	72.1	86.0	70.2	57.7
SD	13.0	15.3	21.3	13.9

Table 1: Intercorrelations, internal consistencies (Cronbach's α), and descriptive statistics for all four spatial measures. MRT = Mental Rotation Test, MCT = Mental Cutting Test, NCS = Novel Cross-Sections Test, SWM = Spatial Working Memory.

*p<.05 (one-tailed), **p<.01 (one-tailed). All scores have been normalized to maximum score of 100.

The four spatial dependent measures were analyzed with a two-factor between subjects 2 × 2 (Assignment [control, training] × Gender [men, women]) multiple analysis of variance (MANOVA). Results indicated a main effect of gender (F(4, 48) = 2.74, p = .039) but no main effect of assignment (F < 1) or interaction Assignment × Gender (F < 1). For individual measures, the effect sizes for training differences were generally small: Mental Rotation Test (d = -0.09), Mental Cutting Test (d = 0.08), Novel Cross-Sections Test (d = 0.05), and Spatial Working Memory test (d = -0.37). A positive effect size indicates an advantage for the training group. Since the main multivariate effect of gender was significant, we conducted a set of one-factor ANOVAs to analyze the main effects of gender across the four tests. Results indicated that men outperformed women on the mental rotation test (F(1, 53) = 6.40, p = .014, d = 0.71) and mental cutting test (F(1, 53) = 8.46, p = .005, d = 0.82) but not on the novel cross-sections test (F(1, 53) = 1.49, p = .228, d = 0.34) or spatial working memory test (F < 1).

A similar set of Assignment × Gender ANOVAs indicated no main effect of assignment or interaction Assignment × Gender on either physics self-efficacy or electricity and magnetism grades. However, results indicated a main effect of gender on physics self-efficacy (F(1, 51) = 5.59, p = .022, d = .70) and electricity and magnetism grades (F(1, 50) = 8.69, p = .005, d = .85).

Physics Measure	Spatial Measure	
	SWM	MRT
Force Concept Inventory – Pre	.16	.55**
Force Concept Inventory – Post	.02	.53**
Newtonian Physics Exam Score	.17	.42**
Physics Self-Efficacy	.11	.42**
Electricity & Magnetism Grades	.05	.20

Table 2: Correlations between physics outcome measures and spatial working memory (SWM) and mental rotation (MRT). *p<.05 (one-tailed), **p<.01 (one-tailed).

To investigate the working memory and specific skills hypotheses, hypotheses, we correlated physics outcomes with working memory and spatial skills measures (see Table 2). For these analyzes, we only used mental rotation as an indicator of spatial skills because of the questionable construct validity of the two mental cutting measures. As shown in Table 2, spatial working memory correlated with none of the physics outcomes, in direct opposition with the working memory hypothesis. However, mental rotation correlated with all physics outcome measures except for electricity and magnetism grades. Except for introductory Newtonian physics grades, correlations between mental rotation and physics outcome measures remained significant after controlling for SAT – Mathematics, SAT – Critical Reading, and SAT – Writing scores. Hence, the significant correlations between mental rotation and some physics outcomes were not because of general academic aptitude, bolstering the specific skills hypothesis.

Discussion

Although the training group, compared to the control group, had higher spatial skills and physics examination scores directly following training (Miller & Halpern, 2010), these training effects did not persist after eight months. However, after eight months, men substantially outperformed women on some spatial measures, had greater physics problem-solving self-efficacy, and achieved higher grades in electricity and magnetism. These results match other studies that have found particularly large gender differences in physics learning (AAUW, 2010, p. 9), although some empirically validated approaches can narrow these gender differences (Miyake et al., 2010). This study adds to this literature by finding that gender differences in mental rotation shares variance with gender differences in physics self-efficacy and some measures of physics learning (although perhaps not for electricity and magnetism grades). However since we did not find long-term improvements in spatial skills, these data are correlational and therefore we cannot make strong conclusions regarding the causal relationship between mental rotation and physics outcomes.

This study has important theoretical implications for the role of spatial thinking in physics and possibly for other STEM fields. This study's results suggest that specific spatial skills, not general spatial working memory, may affect physics self-efficacy and some physics learning outcomes. This result opens up a wide body of research to investigate the relationships between specific spatial skills measures and specific physics topics. However, we are cautious in generalizing our results to other populations because of this study's group of extraordinarily gifted STEM undergraduates; results may differ for more average populations.

Why did we find no evidence of lasting improvements when other researchers have? For example, past longitudinal research with about ten hours of training (Feng et al., 2007; Terlecki et al., 2008) has found little decrement in spatial skills after four months; this suggests that the improvements were stable and likely to last for eight months as well. Furthermore, Sorby (2009) found long-term improvements in STEM course grades and engineering retention rates, although those results could be because of self-selection effects. Our extraordinarily talented STEM population may help explain this divergence from past research. Our sample had extremely high initial spatial skills and students with more average spatial skills may have benefited more from the same amount of spatial experience.

One general limitation of this study is its small sample size. Hence, the null results for non-physics courses and the null results for the 8–10 month longitudinal data could perhaps be explained by a lack of statistical power. However, we note that the effect size magnitudes for these null results were typically small (most $d < 0.20$) and varied in direction (e.g., the control group sometimes nonsignificantly outperformed the training group), suggesting a lack of statistical trends for these null results. Furthermore, power analyses indicated that our statistical tests were sufficiently powered to detect the large average effect size of $d = 0.67$ found in previous longitudinal training studies (Feng et al., 2007; Hedley, 2008; Terlecki et al., 2008).

Interestingly, in online surveys, participants reported that the training exercises were challenging and appropriate to their learning needs (Miller & Halpern, 2010); this study finds that the activities were challenging enough to produce short-term, but not long-term, improvements. This result suggests that sustained exposure to spatially enriching activities over several semesters or years may be necessary to address concerning gender gaps in spatial skills among gifted STEM populations. A promising direction for future research is to investigate how spatial enriching activities like sketching can be integrated into existing STEM courses like physics; such an approach could help learners systematically improve their spatial skills over an extended period of time and also help improve STEM learning outcomes (National Research Council, 2006; Newcombe, 2010). This study suggests that such a sustained and systematic approach may be necessary to improve long-term outcomes and narrow gender gaps among students who are the most likely to pursue advanced STEM educational degrees and occupational positions.

Acknowledgments. This research was funded by the Shanahan Student-Directed Research Funds at Harvey Mudd College. We also thank Peter N. Saeta for his helpful advice and mentoring, Wendy Menefee-Libey for her support from Harvey Mudd's Academic Excellence program, Joseph J. Williams for comments on a draft of this paper, and Alyssa Dray, Brendan Folie, and Neal Pisenti for their time in teaching the spatial training program.

References

Adams W. K., Perkins K. K., Dubson M., Finkelstein N.D., & Wieman C.E. (2006). New instrument for measuring student beliefs about physics and learning physics: the Colorado learning attitudes about science survey. Physical Review Special Topics - Physics Education Research, 2, 1–14.

American Association of University Women (2010). Why So Few? Women in Science, Technology, Engineering, and Mathematics. Washington, DC.

Bandura, A. (1997). Self-efficacy: The exercise of control. W. H. Freeman: New York.

Çalişkan, S., Selçuk, G. S., & Erol, M. (2007). Development of physics self-efficacy scale. Proceedings of the 2007 American Institute of Physics Conference, Vol. 899, pp. 483 - 484.

CEEB Special Aptitude Test in Spatial Relations, developed by the College Entrance Examination Board, USA, (1939).

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. Psychonomic Bulletin & Review, 12(5), 769–786.

Cooke-Simpson, A., & Voyer, D. (2007). Confidence and Gender Differences on the Mental Rotations Test. Learning and Individual Differences, 17, 181-186.

Feng, J., Spence, I., Pratt, J. (2007). Playing an action video game reduces gender differences in spatial cognition. Psychological Science, 18, 850–855.

Halpern. D. F., Aronson, J., Reimer, N., Simpkins, S., Start, S., & Wentzel, K. (2007). Encouraging girls in math and science. Washington, DC: U. S Department of Education. NCER 2007-2003.

Hedley, M. (2008). The use of geospatial technologies to increase students' spatial abilities and knowledge of certain atmospheric science content. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database. (UMI No. 3313371)

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.

Hestenes, D., Wells, M. & Swackhamer, G. (1992). Force Concept Inventory. The Physics Teacher, 30, 141-151.

Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W. & Engle, R. W. (2004) The generality of working memory capacity: A latent variable approach to verbal and visuospatial memory span and reasoning. Journal of Experimental Psychology: General, 133(2), 189–217.

Kozhevnikov, M., Hegarty, M. & Mayer, R. (2002). Visual/spatial abilities in problem solving in physics. In M. Anderson, B. Meyer & P. Olivier (Eds), Diagrammatic Representation and Reasoning, pp. 155-173. Springer-Verlag.

Kozhevnikov, M., Motes, M. A., & Hegarty, M. (2007). Spatial visualization in physics problem solving. Cognitive Science, 31, 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, 113-134.

Kustusch, M. B., & Beichner, R. J. (2010). Student difficulties with right hand rules. Poster presented at the 2010 Physics Education Research Conference, Portland, OR.

Miller, D. I., Halpern, D. F. (2010). Improving the 3-D spatial skills of gifted STEM undergraduates. The General Psychologist, 45(2), 42.

Miyake, A., Kost-Smith, L. E., Finkelstein, N. D., Pollock, S. J., Cohen, G. L., & Ito, T. A. (2010). Reducing the gender achievement gap in college science: A classroom study of values affirmation. Science, 330, 1234-1237.

National Science Board (2010). Preparing the next generation of *STEM innovators: Identifying and developing our nation's* human capital. Arlington, VA: National Science Foundation.

National Research Council (2006). Learning to think spatially. Washington D. C.: The National Academies Press.

Newcombe, N.S. (2010). Picture this: Increasing math and science learning by improving spatial thinking. American Educator: Summer 2010, 29-43.

Peters, M., Laeng, B., Latham, K., Johnson, M., Zaiyouna, R., & Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. Brain and Cognition, 28, 39-58.

Saglam, M., & Millar, R. (2006). Upper high school students' understanding of electromagnetism. International Journal of Science Education, 28(5), 543–566.

Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. Archives of Sexual Behavior, 36, 261–268.

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

Sorby, S.A., & Wysocki, A.F. (2003). Introduction to 3-D spatial visualization: An active approach. Clifton Park, New York: Thomson-Delmar Learning.

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

Uttal, D.H., Liu, L.L., Marulis, L.M., Lewis, A.R., Warren, C.M. (in press). Training spatial skills: What works, for whom, and for how long? Psychological Bulletin.

Voyer, D., Voyer S., Bryden, M.P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. Psychological Bulletin, 117, 250–270.

Wai, J., Lubinski, D., Benbow, C. P. (2009). Spatial Ability for STEM Domains: Aligning Over 50 Years of Cumulative Psychological Knowledge Solidifies Its Importance. Journal of Educational Psychology, 101, 817-835.

Watkins, J., Dowd, J., & Mazur, E. (2010). Gender, Mental Rotations, and Introductory Physics. Poster presented at the 2010 Physics Education Research Conference, Portland, OR.