



Efficacy of Various Spatial Visualization Implementation Approaches in a First-Year Engineering Projects Course

Dr. Jacob L. Segil, University of Colorado Boulder

Jacob L. Segil is an Instructor for General Engineering Plus and Mechanical Engineering degree programs at the University of Colorado Boulder. Jacob has a B.S. in Mechanical Engineering with a minor in Bioengineering from the University of Illinois at Urbana-Champaign, a M.S. in Mechanical Engineering with a Bioengineering focus from the University of Colorado Boulder, and a Ph.D. in Mechanical Engineering from the University of Colorado Boulder. Jacob researches brain-machine interfaces, neural prosthetic devices, and engineering education.

Beth A Myers, University of Colorado Boulder

Beth A. Myers is the engineering assessment specialist for the Integrated Teaching and Learning Program at the University of Colorado Boulder. She holds a BA in biochemistry, ME in engineering management and is currently a PhD candidate studying engineering education at the College of Engineering and Applied Science. She has worked for the University of Colorado in various capacities for 16 years, including as a program manager for a small medical research center and most recently as Director of Access and Recruiting for the College of Engineering and Applied Science. Her interests are in quantitative and qualitative research and data analysis.

Jacquelyn F. Sullivan Ph.D., University of Colorado, Boulder

Jacquelyn Sullivan is founding co-director of the Integrated Teaching and Learning Program and the General Engineering Plus degree program at the University of Colorado Boulder's College of Engineering and Applied Science. She has been intimately involved in the retention-promoting First Year Engineering Projects course since its inception, and has recently become a driver for spatial visualization skill building through the course. She is currently launching CU Teach Engineering, a unique initiative to produce secondary science or math teachers through a new design-based engineering degree, with the ultimate goal of broadening participation among those who attend engineering college. Sullivan was conferred as an ASEE Fellow in 2011 and was awarded NAE's 2008 Gordon Prize for Innovation in Engineering and Technology Education.

Dr. Derek T Reamon, University of Colorado, Boulder

Derek Reamon is the Co-director of the Integrated Teaching and Learning Program (ITLP) and the General Engineering Plus (GE+) degree program, and a Senior Instructor in the Department of Mechanical Engineering. As ITLP co-director, he coordinates 19-22 sections of First-year Engineering Projects, a course that has a proven benefit on retention within engineering and is also a nationally recognized model for freshman design courses. The GE+ program has created a flexible engineering degree and a pathway to secondary math and science teaching licensure, to increase the numbers of STEM teachers that have strong engineering design backgrounds. Derek is also an award-winning teacher and was most recently awarded the John and Mercedes Peebles Innovation in Education from CU's College of Engineering and Applied Science. Dr. Reamon received his PhD in engineering education from Stanford University in 1999. His dissertation was one the first in the nascent field of engineering education research.

Efficacy of Various Spatial Visualization Implementation Approaches in a First-Year Engineering Projects Course

Introduction

Spatial visualization (SV) skills are both learnable¹ and highly correlated with success in engineering. Convinced that improvement in our engineering students' spatial visualization skills would support improved retention in the College of Engineering and Applied Science at the University of Colorado Boulder—a highly research-active university, the first-year engineering projects course faculty team embarked on an evolving and escalating effort to cultivate students' spatial visualization (SV) abilities. Starting in the 2013 academic year, the SV skills of cohorts of entry-level engineering students were measured before and after their completion of a first-year engineering project design course. Students were assessed using the *Purdue Spatial Visualization Test: Visualization of Rotations*² (PVST:R) pre- and post-tests. In subsequent semesters of the same course, each student cohort was assessed before and after a specific SV implementation approach to see the impact of the addition of various formal curricular approaches to cultivate spatial visualization skills.

Our motivation to implement and study spatial visualization skills came primarily from the evidence in the literature concerning historic gender differences in SV ability and the effects of SV skills on retention in engineering. For decades, studies using the PVST:R and other spatial visualization tests have consistently shown gender differences^{3,4} in which male students significantly outperform female students. The source of this gender gap is under investigation; Yoon et al. recently confirmed that the PVST:R does not contain a test bias against gender⁵. Additionally, Sorby and Baartman performed a six-year longitudinal study that showed that a first-year SV intervention increased retention rates in engineering⁶.

Without a clear understanding of the source of the SV gender gap, but convinced that ample evidence existed to support SV intervention as a powerful retention tool with the potential to differentially impact female students, we began to incorporate SV skill-building curriculum into the college's *GEEN 1400: First-Year Engineering Projects* course. Over time, our results motivated us in an escalating fashion to intervene *more* with SV curriculum. This paper describes the various spatial visualization approaches and implementations across five semesters, and reports the resulting efficacy (or lack thereof) of each method.

Methods

To augment the first-year design experience of entry-level engineering students, varying approaches to the addition of formal spatial visualization curricula were implemented and tested during five semesters. We admittedly “failed often in order to succeed sooner” and iterated our implementation approach after each semester's SV growth results were assessed.

The PVST:R was consistently administered at the beginning and end of each implementation for those students who did not originally pass. Our implementation approaches to improve students' SV skills included:

- *Intervention 0*: No special training in SV outside of regular design coursework (that is, hope for the best); pre- and post-test data available as baseline
- *Intervention 1*: In-course SV curriculum and homework assignments, with the potential to earn extra credit
- *Intervention 2*: Voluntary out-of-class SV workshops with homework assignments
- *Intervention 3*: A mandatory out-of-class, four-part SV workshop series for students who did not initially “pass” the SV assessment.

Table 1. Details of the various spatial visualization implementation approaches employed.

Intervention Method & Semester	Institutional Investment			Student Accountability	Student Incentive	Curriculum Framework
	People	Places	Things			
0 Spring 2013	None	None	None	None	None	None
1 Fall 2013	Existing design course faculty (nine faculty members)	None beyond existing class	In-class SV lecture, online homework assignments	5 graded homework assignments	Extra credit awarded if >5 homework assignments completed	Computer-based online practice sets
2 Spring 2014	Graduate student TA (paid hourly)	Dedicated classroom for 6 hours weekly	In-class SV lecture and online homework assignments	Voluntary attendance at out-of-class SV workshops	Beverage gift cards and prize lottery available to students who attended SV workshops	Hands-on workshops and computer-based online practice sets
3 Fall 2014 and Spring 2015	Dedicated faculty member and undergrad TA (paid hourly)	Dedicated classroom for 6 hours weekly	In-class SV lecture, hands-on workshop material (workbooks, blocks, play dough, etc.)	Required attendance at workshops for students who did not pass SV assessment	5% of semester course grade when SV assessment target met	In-person, hands-on curriculum for 4 out-of-class SV workshops

A comparison of various lenses concerning the implementation approach is provided in Table 1. The *institutional investment* lens refers to the resources committed to the SV implementation, including faculty/student employees (*people*), the classroom time/space (*places*), and classroom resources (*things*). The *student accountability* lens details the method used by the faculty to encourage and quantify the student participation. The *student incentive* lens refers to the mechanisms used to motivate students to practice and build their SV skills. Lastly, the *curriculum framework* lens describes the method used to implement the SV curriculum.

To begin, in spring 2013, no special training in spatial visualization was provided to students as we explored whether the SV skills of the 205 engineering students who took the PVST:R pre- and post-assessment would be boosted through the design course itself, or through other factors in the first-year engineering curriculum. This implementation approach provided baseline SV data for a typical first-year cohort.

During the subsequent fall 2013 semester, the next cohort of 279 students was introduced to SV concepts in the form of an introductory SV lecture provided by the design course faculty. In addition, students completed five graded, online SV homework assignments—and could complete more for extra credit. All students were required to complete the assignments, regardless of PVST:R score.

The spring 2014 cohort of 305 engineering students was also introduced to SV concepts in an introductory lecture. Students who did not achieve a passing threshold of at least 20 (of a possible 30) on the PVST:R pretest were asked to participate in voluntary, out-of-class SV workshops led by a graduate student teaching assistant (TA). A classroom was dedicated for six hours weekly for the voluntary workshops, and the TA used online homework assignments and hands-on curriculum to teach SV skills. To incent student participation, a “coffee cart” beverage gift card (~\$5 value, good for any beverage available) was provided each week to student workshop attendees, and attendees who ultimately achieved a score of ≥ 20 on the SV assessment were entered into a lottery to win an iPad.

Finally, during the fall 2014 semester a cohort of 342 students participated in four, two-hour, out-of-class SV workshops, which were *required* for all students who did not achieve scores of at least 20 on the PVST:R pre-test. A dedicated faculty member, assisted by an undergraduate TA, held weekly out-of-class workshops for two four-week sequences. The two sessions gave students who did not achieve success with the first session the opportunity to repeat the SV workshop series. This approach provided much higher commitment levels to both student participation and passing the SV assessment, and put minimal burden on the course instructors. The SV workshops were presented in “Montessori style” with the classroom set up in stations through which students rotated during the two-hour workshop. The stations made use of hands-on materials including workbooks, blocks, play dough, pen and paper sketching, etc. For the first time, participation in the SV workshops was mandatory for non-passers, and they earned 5% of their semester grades once they passed the PVST:R test. The assessment was administered two additional times—after four weeks of workshops and again after eight weeks.

Given the strength of the fall 2014 outcomes shown in Table 2, a similar strategy for delivering the SV curriculum was implemented for the spring 2015 cohort of 316 students in two design-focused first-year engineering projects courses. The SV workshops were scaled from just over 53 students to 67, with the workshops once again being completed during the first half of the semester to give students the opportunity to apply their newly acquired SV skills to their design projects.

As seen in Table 1, the implementation approach escalated each semester across all lenses. The results are presented in Table 2 and discussed below.

Findings

Table 2. Pre- and post-test results from various implementations.

Intervention Method & Semester	# Students in Cohort	Pre-Test Passing Rate (%) (# of students)	Average Pre-Test Score of Workshoppers*	Average Post-Test Score of Workshoppers	Workshoppers Post-Test Passing Rate (%) (#)
0 Spring 2013	205	78% (160)	15.8	17.3	29% (13)
Men	160	81% (130)	16.2	18.0	33% (10)
Women	45	67% (30)	15.1	16.1	20% (3)
1 Fall 2013	279	65% (180)	15.6	18.0	37% (37)
Men	189	72% (137)	15.6	17.6	38% (20)
Women	90	48% (43)	15.7	18.5	36% (17)
2 Spring 2014	305	84% (256)	16.2	22.2	41% (20)
Men	217	91% (197)	16.6	21.6	30% (6)
Women	88	67% (59)	16.0	22.8	48% (14)
3 Fall 2014	342	85% (289)	15.4	23.6	94% (50)
Men	263	92% (242)	16.0	23.9	95% (20)
Women	79	59% (47)	15.0	23.5	94% (30)
3 (cont.) Spring 2015	316	79% (249)	15.3	21.6	82% (55)
Men	231	84% (194)	15.3	21.6	84% (31)
Women	85	65% (55)	15.4	22.0	80% (24)

**To clarify, students who initially did not pass are referred to as “Workshoppers”*

Even though cohort 0 of entry-level students was immersed in an academic engineering culture and the first-year engineering projects-based design course included a significant amount of drawing and visualization of three-dimensional designs, without an SV curricular intervention, only 29% of students who initially did not pass the pre-test increased their SV skills to pass the post-test by semester end. We were not satisfied with that outcome.

Intervention 1, in which all students (including those who initially passed the SV pre-test) completed online, computer-based practice sets as homework, saw only 37% of initial non-passers ultimately passing the post-test. Again, we were not satisfied with this outcome.

In Intervention 2, only the 49 students with scores below 20 (the “pass” threshold) were asked to participate in voluntary, out-of-class spatial visualization workshops led by a graduate teaching assistant. Employing a “voluntary, this is good for you” approach did not work: of the 49 students, 32 (65%) completed one or more of the homework assignments, and only 26 (53%) took the workshop post-test—with a disappointing 41% ultimately passing the post-test threshold.

While both Interventions 1 and 2 pre- to post-test PVST:R scores showed statistically significant improvement, the SV skill gains and participation rates fell far short of the course learning

outcome goals. Still believing that improved spatial visualization skills would lead to improved retention in the engineering program, a much more intensive *and intentional* implementation of the SV curriculum was designed for fall 2014 and spring 2015 (Intervention 3). Our goal became clearer: to aggressively develop SV skills among *all* ~600 students enrolled annually in the first-year engineering projects course. To achieve this outcome, we knew we needed a higher-stakes implementation approach, to which we dedicated more resources.

Thus, Intervention 3 required the most institutional investment in people, places and things—accompanied by much higher student accountability and incentives. As seen in Table 2, students achieved by far the best SV skill development gains. Ultimately, 94% and 81% of initial “non-passers” achieved the threshold of 20 by course end in fall 2014 and spring 2015 respectively, with a full 99% and 96% of the fall and spring cohorts in the first-year design course ultimately achieving the SV threshold of 20. Finally, we were satisfied that our model met our aggressive SV improvement goal—and hopefully these results are impactful enough to realize long-term benefits in retention and engineering success.

Gender Does Matter. Across all interventions and cohorts we found significant gender differences among the students with pre-test scores of less than 20. With Intervention 3, the SV performance gap between male and female students was closed during both the fall and spring semesters. With the Intervention 3 fall 2014 cohort, 60% of initial non-passers were women, even though only 23% of the students in the course were female. The passing rate for male and female students began at 92% and 59% respectively during fall 2014; this was a statistically significant difference in the passing rate ($p < 0.05$). The post-intervention passing rate was 99% and 98% for male and female students respectively—no longer statistically different ($p > 0.05$). For the spring 2015 cohort, the passing rate for male and female students was again statistically significantly different at 85% for males and 65% for females. The post-intervention passing rate was 98% and 94% for male and female students respectively, again no longer statistically different due to gender ($p > 0.05$).

Resource investment matters. We have found that *doing our best* with SV skill development requires institutional investment of about one course equivalent of faculty time spread throughout the year, accompanied by one undergraduate TA for every 50 students receiving the SV intervention. With this model, we believe we can impact about 250 engineering students annually. And, with approximately 20% of our incoming first-year engineering students needing SV skill building, we expect to be able to implement this out-of-class workshop model to an entering cohort of up to 1,250 students.

Further, our results show that both student and faculty accountability affect the outcomes. We employ a faculty matrix accountability model designed to require little time for, but lots of caring from, the course instructors. The instructors of the individual sections of the design course are continuously kept abreast of their students’ progress by the dedicated SV faculty member. Typically, four or five students in each section (of ~30) need the SV intervention. Accountability for students to participate in the workshops (and receive the 5% of their course grades) lies with the course instructors, not the SV instructor. Thus, communication and mutual agreement on accountability are necessary for achieving student success.

As we learn more about what materials help our students most effectively, we anticipate a refinement of the course materials and activities (see details in the Appendix). A post-workshop survey found that time spent with the TA, drawing, and the hands-on blocks were the most beneficial materials. The least beneficial materials included the online computer tutorials and quizzes. We will pursue more creative activities for each of the stations; for example, we have planned an inductive learning station that asks students to describe orthographic views of objects while blindfolded. As always, we plan to *fail often to succeed sooner* and continue to improve upon the lessons learned from the previous interventions.

Appendix

Details about the activities used during the various interventions: Generally, successful activities involved physical implementation of the SV curriculum, while unsuccessful activities involved virtual (computer) interfaces.

- Online homework assignments (*unsuccessful*, used in Interventions 1 and 2)—Students were assigned practice sets via the class website. Practice sets involved selecting the correct multiple choice answers without requiring students to show work and/or explain their responses. Then, correct answers were provided by the website.
- Block and Draw (*successful*, used in Intervention 3)—Students used square blocks to build objects. Next, they used pencil on isometric graph paper to draw the isometric view of the object. Then they traded objects with another student, and drew the other object.
- Workbook Sessions (*successful*, used in Intervention 3)—Students use a workbook¹ to practice applying various curricular topics like drawing orthographic views or defining the 2-axis rotation. Students were required to show their work, select multiple choice answers, and check their answers in the solutions manual. The SV faculty and/or TA discussed any troublesome problems and provided strategies for each curricular topic.

Acknowledgements

This material was motivated and partially supported by the ENGAGE project. In addition, the authors thank their research colleague Denise W. Carlson for her insights and critique.

References

1. Sorby, Sheryl A. and Anne F. Wysocki. "Introduction to 3D Spatial Visualization: An Active Approach." New York, NY: Thomson Delmar Learning, 2003.
2. Guay, Roland B. "Purdue Spatial Visualization Test: Rotations." West Lafayette, IN: Purdue Research Foundation, 1977.
3. Battista, Michael T. "Spatial Visualization and Gender Differences in High School Geometry." *Journal for Research in Mathematics Education*. 21.1 (1990): 47-60.
4. Sorby, Sheryl A. "A Course in Spatial Visualization and Its Impact on the Retention of Female Engineering Students." *Journal of Women and Minorities in Science and Engineering*. 7.2 (2001b): 153-172.
5. Yoon, So Yoon. "Psychometric Properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (the Revised PVST:R)." Dissertation, Purdue University, 2011.

6. Sorby, Sheryl A., and Beverly J. Baartmans. "The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students." *Journal of Engineering Education*. 89.3 (2000): 301-307.

Analysis of Multi-Modal Spatial Visualization Workshop Intervention across Gender, Nationality, and Other Engineering Student Demographics

Jacob L. Segil, Jacquelyn F. Sullivan, Beth A. Myers, Derek T. Reamon and Marissa H. Forbes

Engineering Plus Program

College of Engineering and Applied Science, University of Colorado Boulder

Boulder, CO, USA

jacob.segil@colorado.edu

Abstract—Spatial visualization (SV) skills are both learnable and linked to student success in engineering. Our work was motivated by previous research revealing a gender gap in SV skills that favors male engineering students. A multi-modal SV workshop intervention was incorporated into a first-year engineering projects course with the goal of fostering SV skills for all students. An analysis of SV skills, as measured by the Purdue Spatial Visualization Test (PSVT), preceding and following workshop participation is presented. SV workshop results were studied for five undergraduate engineering student demographic subsets: 1) gender, 2) first-generation status, 3) nationality, 4) socioeconomic status, and 5) underrepresented minority status. Both female and international students were found to arrive at engineering college with less-developed SV skills. Since both populations are rapidly growing in the nation's engineering colleges, closing the SV performance gap is compelling. The multi-modal SV workshop provided substantial support to all students, with a median 23% performance gain. And, workshop attendees improved their PSVT passing rate (achieving a 20-point threshold) from 0% to 92%. Results demonstrate that this replicable multi-modal SV workshop can close SV skills performance gaps among all engineering students.

Keywords—student retention; gender gap; hands-on; spatial visualization; first-year engineering

I. INTRODUCTION

Spatial visualization (SV), the ability to mentally manipulate two- and three- dimensional objects, is a foundational skill for students and professions in the science, technology, engineering and math (STEM) fields. Additionally, SV abilities have been linked to student retention in engineering and happily, they are learnable [1].

Previous research revealed a gender gap in SV skills that favors male engineering students [2]; the extent to which other student demographics or characteristics may be tied to differentiated SV abilities is unknown. Between 2004 and 2014, the national growth of undergraduate engineering enrollments for women, international, and underrepresented minority (URM) students was rapid (Table 1) [3]. Engineering colleges have an ethical imperative to facilitate the best possible opportunities for each of their students to succeed and

persist in engineering, regardless of student characteristics and demographics—including gender, nationality and others.

In 2014, faculty at the University of Colorado Boulder incorporated SV interventions into several first-year engineering projects courses with the goal of nurturing spatial visualization skills for all students. The faculty were motivated by Dr. Sheryl Sorby's work at Michigan Technological University and by the ENGAGE Engineering initiative which highlighted the ways that SV skills improve performance and retention in engineering. Many intervention designs were previously implemented at CU-Boulder, but with varied and less-than-desired, success [4]. The Purdue Spatial Visualization Test: Visualizations of Rotations (PSVT:R), a cognitive test designed to measure SV ability in 3-D mental rotation, was administered before and after each intervention [5, 6]. Largely unsatisfied with their early results, the faculty developed a required four-week multi-modal SV workshop series for students who did not initially pass the PSVT.

This paper introduces the multi-modal SV workshop and presents an analysis of student SV skills preceding and following workshop participation. SV skill results are explored by 1) gender, 2) first-generation status, 3) nationality, 4) socioeconomic status, and 5) underrepresented minority status in order to quantify the effect of the workshop on SV skills development for a broad range of undergraduate engineering student populations.

TABLE I. NATIONAL UNDERGRADUATE ENGINEERING ENROLLMENTS

	2004 - 2005	2013 - 2014	% Increase
Female	63,980	117,511	84%
International	18,954	53,368	182%
URM	54,813	92,488	69%

II. RESEARCH QUESTIONS

- Do differential SV skills exist upon entry to the first-year engineering projects class across 1) female and male students; 2) first-generation students and not; 3) international and domestic students; 4) low SES students and not; and 5) URM students and not?

- To what extent does the multi-modal SV workshop impact both student PSVT performance and passing rate?
- Does the multi-modal SV workshop support *all* populations of engineering students in developing their SV skills?

III. METHODS

A. Research Methods

During the first week in each of four semesters (fall 2014, spring 2015, fall 2015 and spring 2016), 1,521 students enrolled in first-year, hands-on, design-based engineering projects classes in the College of Engineering and Applied Science at the University of Colorado Boulder were administered the PSVT. Based on previous research, a score of 20 points or more (out of 30 possible) constituted a passing grade [7]. To promote accountability, a passing grade of ≥ 20 earned full credit for 5% of the 16-week semester course grade. The 251 students (17%) who did not pass the initial PSVT (scores < 20) were required to attend an out-of-class four-week series of spatial visualization workshops; these 251 students are referred to as “workshoppers” in this study.

B. Undergraduate Engineering Student Demographic Subsets

Undergraduate engineering students were studied based on five dichotomous demographic subsets, including 1) gender, 2) first-generation college status (first-gen), 3) nationality, 4) low socioeconomic status (low SES), and 5) whether their race or ethnicity is an underrepresented minority (URM) in engineering.

The percentages of each demographic subset in the workshopper population, compared to the engineering college as a whole, are presented in Table II. The percent of workshopppers is cumulative across the 2014-2015 and 2015-2016 academic years; the percent of college enrollment is the average of the 2014 and 2015 first-year students [8]. The difference is provided to indicate the over and/or underrepresentation of each demographic subset among the SV workshops. Results show that both female and international students were greatly overrepresented in the workshops by 72% and 155%, respectively. The overrepresentation of these students in the workshop population indicates that they had differentially low PSVT pre-test passing rates, and suggests that these student groups tended to arrive at the engineering college with differentially less-developed SV skills, as measured by the PSVT.

Of note is that URM students were *underrepresented* among the workshopppers. This finding is encouraging, but may be because many rising first-year URM students attended a one- or two-week summer bridge program on campus the summer before their first year of engineering college, during which all students took the SV test and participated in a shorter version of the SV workshop. Also contributing to the low number of URM workshopppers may be that the SV intervention was also done for students in a first-year design course in a pre-engineering program that has a lower representation of URM students than is found for the engineering college overall.

TABLE II. WORKSHOPPER ENROLLMENT COMPARED TO ENGINEERING COLLEGE ENROLLMENT

	<i>Percent of SV Workshopppers</i>	<i>Percent of Engineering College</i>	<i>Percent Difference</i>
Female	50%	29%	+72%
First-Generation	16%	15%	+7%
International	23%	9%	+155%
Low SES	20%	17.5%	+14%
URM	13%	16%	-19%

C. The Multi-Modal Spatial Visualization Workshop

The multi-modal SV workshop was designed to loosely mimic a Montessori learning approach, allowing students to explore and learn materials by self-paced discovery as opposed to direct instruction, providing a range of activities that students encounter in each two-hour workshop, and designing materials so that a range of senses are called into action during each session.

The SV workshop was organized into two, four-week sessions. As an example, the first session focused on four foundational SV skills: 1) isometric views, 2) orthographic drawing, 3) 1-axis rotations, and 4) 2-axis rotations. The second session addressed more specific challenges, such as inclined planes, curved surfaces and timed problem solving.

Each two-hour workshop began with a brief topical introduction (e.g., isometric views, orthographic drawings, 1-axis and 2-axis rotations, etc.), after which students rotated through four stations that featured different activities on the same topic. For example, the “Block and Draw” station required students to build an object (tactile sensation) and then draw the object from various isometric viewpoints (physical task). In comparison, the “Peer Teach” station challenged students to watch a peer build an object (visual sensation) and then describe to/teach him/her how to draw the isometric view of the object (verbal task). The multi-modal nature of the SV workshop series thus created opportunities for students to practice each SV skill by tapping into a variety of sensory pathways [9].

The SV workshop was offered in the early evening twice weekly for a four-week session; a makeup workshop was also offered each Sunday afternoon. Thus, a student had three different opportunities to attend each week’s two-hour workshop. Students were required to attend all four weeks of the session in order to retake the PSVT, which they were required to pass to earn 5% of their design course grade.

A second four-week session, with the same design but different content, was provided for students who did not pass the PSVT following the first session of workshops. This eight-week progression was offered during the first half of the semester in order to facilitate students in cultivating their SV skills before their course workloads heightened, and was designed so that students could immediately apply their newly developed SV skills that same semester in their design courses.

A single faculty member taught the eight weeks of workshops, managed the SV workshop across the multiple sections of three first-year project classes, and maintained communication with both the first-year project faculty team and students involved in the workshops. The SV workshop was considered a half teaching credit each semester for the faculty member (thus equivalent to one course over the entire academic year).

D. Spatial Visualization Skill Metrics

The effects of the multi-modal SV workshop were quantified by comparing three pre- and post-workshop metrics: 1) the percentage of students who scored equal to or greater than 20 points (passing rate, PR), 2) the median integer score out of 30 possible points (performance, PF), and 3) the median integer difference between the post-test and pre-test scores for workshopppers (gain, G). Each of the 1,521 first-year projects class students took a pre-test; only the 251 workshopppers took the post-test. For non-workshopppers, the pre-test score and post-test scores were made equal for statistical analysis. The median PF scores were reported instead of the mean scores since the PF score is an ordinal variable.

E. Statistical Analyses and Software

Fisher's Exact Tests (a two-sample exact test for proportions) were used to test for statistically significant differences between passing rates. Mann-Whitney U tests (a nonparametric test analogous to the t-test for continuous data) were used to detect statistically significant differences in the ordinal performance metric. In each case, statistical significance was determined using an $\alpha = 0.05$. Statistical analyses were performed using MATLAB (MathWorks, Inc., Natick, MA). Effect sizes were not reported since the statistical analyses were studying the entire population and not making inferences based on a sample population.

IV. FINDINGS

PSVT pre-test and post-test results for the entire cohort of first-year engineering projects students and workshopppers are presented, by demographic subset, in Tables III and IV, respectively. Shaded cells denote statistically significant differences between the demographically dichotomous groups, and p-values are provided as superscripts.

While the pre-test and post-test performance scores (PF) for the entire cohort ("All") did not change (median scores of 25), the passing rate (PR) for the pre-test and post-test increased from 82.8% to a remarkable 98%. In other words, *nearly all* of the 1,521 first-year students completed their engineering projects class with SV skills that *surpassed the 20-point threshold*.

Looking deeper into these results, both the pre-test PF and PR for the female and international student cohorts were of particular interest (Fig. 1). Echoing the findings of other researchers, data from this study also revealed a gap in SV skills with respect to gender. In the pre-test PR metric, female students (PR = 67.5%) passed the test at a considerably lower rate than their male peers (PR = 88.2%). A similar gap was found between international and domestic students, with

international students (PR = 61.4%) passing the test at noticeably lower rates than domestic students (PR = 85.2%). International students passed the SV test at the lowest rate compared to each of the other demographic populations, as measured by the PR metric. These results are not only indicative of a gap in SV skills with respect to gender, but also a gap in SV skills with respect to the nationality (international versus domestic) of first-year engineering students at our university. Also, no differences were found among the first generation status, socioeconomic status, nor URM status demographic subsets—which further highlights the differences between female and international students compared to their peers.

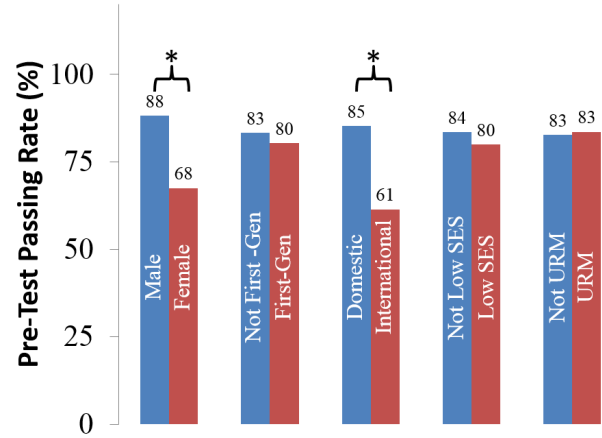


Fig. 1. Entire cohort pre-test passing rate. (n = 1,521). The asterisks indicate statistically significant differences between passing rates ($p < 0.001$).

For the 251 workshopppers (Table IV), the median pre-test PF score was 17, which improved to a median post-test PF score of 23. The improvement of all workshopppers from the pre- to the post-test was statically significant (p 's < 0.001) for all demographic groups (Fig. 2).

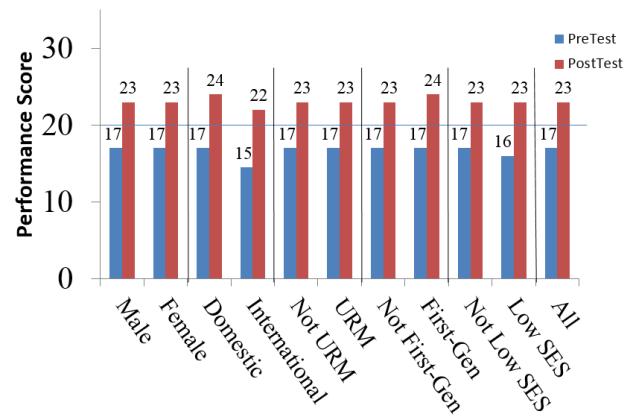


Fig. 2 Workshopper median pre- and post-test performance scores (n = 251). The horizontal line indicates the passing threshold of 20 points. All demographic subsets improved at a statistically significant level (p 's < 0.001).

The pre-test PR, by definition, was 0% for workshopppers and improved to a 91.6% PR. Nearly all students who completed the workshop series improved their SV skills and surpassed the standard threshold of 20. And, the median PSVT gain was 7 points (23% increase) for all workshopppers.

These findings demonstrate that the multi-modal SV workshop was successful in improving student SV skills for students from all demographic subsets, thus achieving our overarching goal of *equity in access* to an engineering education.

Of note, only the international workshopppers started and ended the workshop with significantly lower skills than the

alternative category, their domestic peers. However, the median international *and* domestic student gains were equivalent (7 points), indicating that both benefited comparably from the workshop.

No statistically significant differences were detected by gender, first-generation status, low SES status, or URM status for the workshopppers; for each of these demographic groups, students scored similarly on the pre-test and improved similarly on the post-test.

TABLE III. ENTIRE COHORT PRE-TEST AND POST-TEST PERFORMANCE AND PASSING RATES

Characteristic	Category	Student Count	Median Pre-Test PF (points)	Pre-Test PF Interquartile Range (points)	Pre-Test Passing Rate (%)	Median Post-Test PF (points)	Post-Test PF Interquartile Range (points)	Post-Test Passing Rate (%)
Gender	Male	1127	26	6	88.2	26	5	98.6
	Female	394	22 ^{.000}	6	67.5 ^{.000}	24 ^{.000}	4	96.2 ^{.005}
First-Gen	No	1287	25	6	83.3	25	5	98.3
	Yes	234	24 ^{.035}	6	80.3 ^{.300}	25 ^{.021}	5	96.1 ^{.043}
International	No	1368	25	6	85.2	25	5	98.6
	Yes	153	22 ^{.000}	9	61.4 ^{.000}	23 ^{.000}	6	92.2 ^{.000}
Low SES	No	1253	25	6	83.5	25	5	98.3
	Yes	268	24.5 ^{.165}	6.25	79.9 ^{.154}	25 ^{.284}	5	96.3 ^{.052}
URM	No	1289	25	6	82.7	25	5	98.1
	Yes	232	25 ^{.803}	6	83.6 ^{.777}	25 ^{.433}	4	97.0 ^{.307}
All		1521	25	6	82.8	25	5	98.0

Shaded cells denote statistically significant differences between the demographically dichotomous groups, and p-values are provided as superscripts.

TABLE IV. WORKSHOPPER PRE-TEST AND POST-TEST PERFORMANCE AND PASSING RATES

Characteristic	Category	Student Count	Median Pre-Test PF (points)	Pre-Test PF Interquartile Range (points)	Pre-Test Passing Rate (%)	Median Post-Test PF (points)	Post-Test PF Interquartile Range (points)	Post-Test Passing Rate (%)	Median Gain (points)	Gain Interquartile Range (points)
Gender	Male	125	17	4	0	23	4	93.6	7	4
	Female	126	17 ^{.779}	5	0	23 ^{.802}	5	89.7 ^{.362}	7 ^{.570}	6
First-Gen	No	210	17	4	0	23	5	91.9	7	4.5
	Yes	41	17 ^{.253}	5	0	24 ^{.934}	4	90.2 ^{.758}	7 ^{.708}	4
International	No	193	17	4	0	24	4	94.8	7	5
	Yes	58	14.5 ^{.002}	5	0	22 ^{.001}	4	81.0 ^{.002}	7 ^{.818}	3
Low SES	No	202	17	4	0	23	5	92.1	7	4.5
	Yes	49	16 ^{.614}	4.5	0	23 ^{.604}	4.25	89.8 ^{.572}	8 ^{.953}	5.75
URM	No	218	17	4	0	23	4	91.3	7	5
	Yes	33	17 ^{.205}	4	0	23 ^{.821}	4.25	93.9 ^{.999}	7 ^{.353}	5.25
All		251	17	4	0	23	4	91.6	7	5

Shaded cells denote statistically significant differences between the demographically dichotomous groups, and p-values are provided as superscripts.

V. FINAL THOUGHTS

This investigation analyzed the impact of CU-Boulder's multi-modal spatial visualization workshop on first-year engineering students' SV skills (as measured by the PSVT) across gender, nationality and other demographics for four semesters. The workshop resulted in comparable gains to all demographic subsets of students (median gain of 7 points) and SV workshop attendees improved their PSVT passing rate from 0% to 92% (based on a 20-point threshold).

Significant differences existed in the incoming spatial visualization skills of the studied first-year engineering students with respect to gender and nationality, with women and international students initially demonstrating less-developed SV skills (as measured by lower PSVT pre-test performance and passing rates). While the SV skills gender gap is well-established, in this work we also found a similar gap in SV skills between domestic and international engineering students.

Also of significance, this work found a replicable pathway for students to learn the SV skills necessary for student success by attending an eight-hour, out of class SV workshop series. The workshop design engaged one engineering faculty member to support many sections of three first-year engineering design courses, both negating the need for curriculum change within the design courses and for each faculty member to learn to teach SV skill development themselves.

Moving forward, we will assess the longitudinal effects of this multi-modal SV intervention as the students matriculate through the engineering college. We are primarily curious about whether correlations exist between SV scores and success and/or persistence in engineering.

Additionally, we hope to derive an optimal passing threshold for the PSVT based on empirical evidence. Although we have been using a 20-point passing threshold, workshopppers achieved a post-test median score of 23, which implies that students who originally scored 20, 21 or 22 points may have also benefitted from such a workshop. With student success as our primary goal, we question whether we are inadvertently leaving students behind when applying a 20 point threshold.

We found that this replicable, multi-modal SV workshop intervention can benefit each of the studied demographic subsets of students and is particularly important for two fast-growing populations of undergraduate engineering students: women and international students. This finding is of particular

importance as universities recruit and admit an increasing number of female and international students to colleges of engineering. In order to promote equity in student success, similar multi-modal SV workshops can be used to develop SV skills and help level the playing field for all students as they matriculate through their engineering degrees. At CU-Boulder, we will continue to develop the SV skills of our students to ensure equity in student success across gender, nationality and all other demographics.

VI. ACKNOWLEDGMENT

The authors express their appreciation to Dr. Sheryl Sorby for her extensive work on this topic, which provided a foundation on which we could build, and to Dr. P.K. Imbrie for early research guidance. The authors also thank WEPAN and the ENGAGE Engineering initiative for inspiration. Finally, the authors thank Denise W. Carlson for her insights and critique of the manuscript.

VII. REFERENCES

- [1] S. A. Sorby, "Educational research in developing 3-D spatial skills for engineering students," *Int. J. Sci. Educ.*, vol. 31, no. 3, pp. 459–480, Feb. 2009.
- [2] S. A. Sorby, "A course in spatial visualization and its impact on the retention of female engineering students," *J. Women Minor. Sci. Eng.*, vol. 7, no. 2, p. 20, 2001.
- [3] "College Profiles: American Society for Engineering Education." [Online]. Available: https://www.asee.org/papers-and-publications/publications/college-profiles#The_Profiles_and_Data_Book. [Accessed: 04-Apr-2016.]
- [4] J.L. Segil, B.A. Myers, J.F. Sullivan, and D.T. Reamon, "Efficacy of various spatial visualization implementation approaches in a first-year engineering projects course," *Proceedings, 2015 ASEE Annual Conference & Exposition*, Seattle, WA, June.
- [5] S. Y. Yoon, "Psychometric properties of the revised Purdue spatial visualization tests: visualization of rotations (the revised PVST:R)," dissertation, Purdue University, West Lafayette, IN, 2011.
- [6] R. B. Guay, "Spatial ability measurement: A critique and an alternative," presented at the *Annual Meeting of the American Educational Research Association*, West Lafayette, IN, April 3, 1980.
- [7] S. S. Metz, S. A. Sorby, T. S. Berry, C. C. Seepersad, A. M. Dison, Y. S. Allam, J. A. Merrill, W. Peters, E. Pfister-Altschul, S. C. Baxter, et al., "Implementing ENGAGE strategies to improve retention: focus on spatial skills engineering schools discuss successes and challenges," *Proceedings, 2011 ASEE Annual Conference & Exposition*, Vancouver, BC, June.
- [8] M. Steiner, "Undergraduate student statistics, College of Engineering and Applied Science, University of Colorado Boulder—2007 to 2015." Office of Dean, September 22, 2015.
- [9] J. L. Segil, B. Louie, J. F. Sullivan, and B.A. Myers, "Hands-on, minds-on spatial visualization workshops for first-year engineering students," presented at the 2015 *Women in Engineering ProActive Network - Change Leader Forum*, Denver, CO, 2015.

Investigation of Spatial Visualization Skills Across World Regions

Jacob L. Segil, Jacquelyn F. Sullivan, Janet Y. Tsai, Derek T. Reamon and Marissa H. Forbes

Engineering Plus Program

College of Engineering and Applied Science, University of Colorado Boulder

Boulder, CO, USA

jacob.segil@colorado.edu

Abstract—Spatial visualization (SV) skills contribute to success in engineering. However, ample research from American university settings indicates that various subsets of engineering students have significantly less-developed SV skills than those demonstrated by the majority male population. A multi-modal SV workshop intervention was provided within a first-year engineering projects design course in order to “close the SV gap” for all students. A nationality gap was identified in which international students scored dramatically lower on the Purdue Spatial Visualization Test (PSVT) compared to their domestic counterparts. In this paper, an analysis of SV skills segmented by world region is presented. No difference was found between East Asian and domestic students; however, Middle Eastern students scored dramatically lower on the pre-test than East Asian and domestic students across both a performance score (19 out of 30) and passing rate (46%). The Middle Eastern students improved significantly due to the workshop, with median SV gains of 4 points on the performance score and 38% on the passing rate. These results identify first-year Middle Eastern students as being potentially at risk within colleges of engineering due to less developed spatial visualization skills that can be strengthened through targeted, short-duration training and skill development interventions.

Keywords—student retention; international students; hands-on; spatial visualization; first-year engineering; Middle Eastern region

I. INTRODUCTION

Spatial visualization (SV) skills are integral to engineering education, and improve retention in engineering programs [1]. Faculty at a highly research-active university developed a multi-modal SV workshop for students in several first-year engineering projects design courses to improve SV skills.

Previously, we analyzed and disaggregated SV results by various student demographics, including gender, socio-economic status, underrepresented minority status, first-generation college status and nationality [2]. Results replicated the well-established SV skill gender gap, and exposed a surprising result: international students arrived at our engineering college with far less-developed SV skills than their domestic peers. This finding presents an equity issue for universities recruiting an increasing number of international engineering students. Nationally, the undergraduate engineering enrollments for international students increased by 182% between the 2004-2005 and 2013-2014 academic years [3]. Educators have an ethical responsibility to promote equity in

student success to an engineering education; our SV research findings identify a student population whose SV skills need significant and early bolstering to scaffold them for subsequent success in their engineering educations.

Other researchers have studied the differences in SV skills among international and domestic (US) students in various forms. Dr. Sheryl Sorby investigated the differences among domestic and international engineering students over many years (1996-2011) with a small population of international students (242 out of 11,441 total) enrolled at Michigan Technological University [4]. This study showed a statistically significant difference in Purdue Spatial Visualization Test (PSVT) scores among Middle Eastern, African, Chinese and Indian students compared to their domestic peers, while students from Far East Asia, South America and Canada showed no differences in SV performance. In addition, Sorby studied students from the United Arab Emirates and compared their PSVT scores to a domestic student population [5]; this work confirmed a gap between the spatial skills of the UAE students compared to domestic students. We were motivated by these findings to ascertain if our international engineering students performed differently on the PSVT than their domestic peers, whether students from various world regions demonstrated differential SV skills, and whether our multi-modal SV workshop intervention could improve student performance on the PSVT. This paper reviews the multi-modal SV workshop intervention and analyzes the pre- and post-test results for students across various world regions.

II. RESEARCH QUESTIONS

- Do differential SV skills exist upon entry to the first-year engineering projects design class across the following world regions of origin: 1) Americas (except domestic students); 2) Europe; 3) Africa; 4) Middle East; 5) South and East Asia; and 6) Australia and Pacific?
- To what extent does the multi-modal SV workshop differentially impact both student PSVT performance and passing rate for international students?
- Does the multi-modal SV workshop equitably support *all* international students in developing their SV skills?

III. METHODS

A. Research Methods

Students completed the Purdue Spatial Visualization Test (PSVT) during the first week of class for their first-year, hands-on, design-based engineering projects classes at a highly active-research public university. A total of 2,441 students enrolled in these courses over six semesters (Fall 2014 – Spring 2017). Students who scored above or equal to 20 points (out of 30 possible) passed the test according to a previously used passing threshold [6]. Students who did not initially pass the PSVT ($n = 388$ or 16% of all students) were required to attend the out-of-class spatial visualization workshop series; in this study, we refer to these students as “workshoppers.” Students earned 5% of their semester course grades for passing the PSVT test, which encouraged consistent SV workshop attendance and motivation to perform.

B. World Region Subsets

In the present study, we analyzed our international undergraduate student cohort by world region, using the world regions defined by the World Bank [7]. We identified six world regions from which our international students were citizens: 1) Americas (except domestic students); 2) Europe; 3) Africa; 4) Middle East; 5) South and East Asia; and 6) Australia and Pacific. Among the six world regions, only the Middle East and Asian regions had large enough student populations in our six-semester dataset to analyze statistically. We considered domestic students (USA nationals) a seventh world region for comparison. 236 international students (10% of all students) enrolled in the course and 84 international students participated in the SV workshops (22% of workshoppers). Table II depicts the workshop enrollment across world regions.

Note that Middle Eastern students are over-represented as workshoppers (14%) compared to their enrollment in the course overall (4%) while domestic students are underrepresented (78% of workshoppers versus 91% of the course).

TABLE I. WORKSHOPPER ENROLLMENT ACROSS WORLD REGIONS

	Number of Students	Percentage of Students in Course	Percentage of SV Workshoppers
Americas	12	< 1%	< 1%
Europe	8	< 1%	< 1%
Africa	2	< 1%	< 1%
Middle East	102	4%	14%
South and East Asia	111	5%	6 %
Australia and Pacific	1	< 1 %	0%
USA (Domestic)	2,205	91%	78%

C. The Multi-Modal Spatial Visualization Workshop

Workshoppers attended a four-week evening SV workshop during which they rotated through various hands-on activities in two-hour sessions to study SV topics such as isometric drawing, orthographic views, etc. The PSVT was re-administered

following workshop completion. If students did not pass the PSVT on the second attempt, they were required to participate in an additional four-week workshop series before being eligible to retake the PSVT a final time. This form of the SV intervention was administered over six semesters to more than 2,400 engineering students. Additional detail concerning prior SV interventions and results are available in a previous paper [8].

D. Spatial Visualization Skill Metrics

The effects of the SV workshop were quantified by comparing two pre- and post-test metrics: 1) the median integer score out of 30 possible points (performance, PF) and 2) the percentage of students who scored equal to or greater than 20 points (passing rate, PR). All students enrolled in the course completed a pre-test while only *workshoppers* completed a post-test. For non-workshoppers, the pre- and post-test scores were made equal for statistical analysis. Median (rather than mean) PF scores are reported since the PF score is an ordinal variable.

E. Statistical Analyses and Software

Fisher’s Exact Tests (a two-sample exact test for proportions) were used to test for statistically significant differences between passing rates. Mann-Whitney U tests (a nonparametric test analogous to the t-test for continuous data) were used to detect statistically significant differences in the ordinal performance metric. In each case, statistical significance was determined using $\alpha = 0.05$. Statistical analyses were performed using MATLAB (MathWorks, Inc., Natick, MA). Effect sizes were not reported since the statistical analyses studied the entire population instead of making inferences based on a sample population.

IV. FINDINGS

Table II presents the number of students from each world region (in total and by gender) as well as the performance score (PF), passing rates (PR), and interquartile ranges for the pre- and post-test. Only the Middle East, East Asia, and domestic world regions were studied in depth due to the limited number of enrolled engineering students from all other world regions. The number of students from the Middle East and East Asia regions were similar ($n = 102$ and $n = 111$, respectively) while the number of domestic students was an order of magnitude greater ($n = 2,205$).

Figures 1 and 2 depict the median PF and PR for the pre- and post-test across world regions. The Middle Eastern student pre-test PF and PR scores (19 points and 46%, respectively) were significantly lower than both the East Asian and domestic student scores. After the workshop series, the Middle Eastern student post-test PR and PF scores improved dramatically (to 23 points and 84%), but were still significantly lower than the post-test scores for both East Asian and domestic student populations. However, no difference was found between the East Asian and domestic students across both metrics for the pre- and post-test. In fact, East Asian and domestic students achieved the identical median pre- and post-test PR scores (25 and 26, respectively, out of 30).

TABLE II. ENTIRE COHORT PRE-TEST AND POST-TEST PERFORMANCE AND PASSING RATES

Characteristic	Category	Student Count	Median Pre-Test PF (points)	Pre-Test PF Interquartile Range (points)	Pre-Test Passing Rate (%)	Median Post-Test PF (points)	Post-Test PF Interquartile Range (points)	Post-Test Passing Rate (%)
Middle East	Total	102	19	10	46	23	5	84
	Male	79	20	10	52	23	5	91
	Female	23	14	10	26	22	5	61
South and East Asia	Total	111	25	8	79	26	6	98
	Male	84	26	7	83	26	6	99
	Female	27	21	7	67	23	6	96
USA (Domestic)	Total	2,205	25	5	86	26	5	98
	Male	1622	26	5	91	26	4	98
	Female	583	23	5	72	24	4	97
All Students	Total	2,441	25	6	84	25	5	98

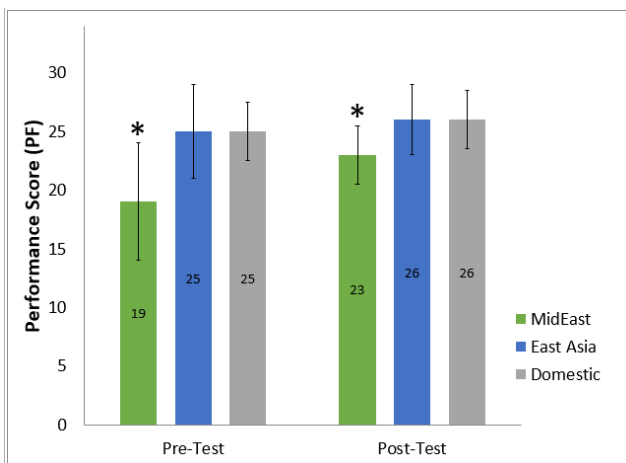


Fig. 1. Median pre- and post-test performance scores for Middle Eastern (MidEast), South and East Asia (East Asia), and domestic students. Error bars display interquartile range. Asterisks indicate statistically significant differences ($p < 0.01$).

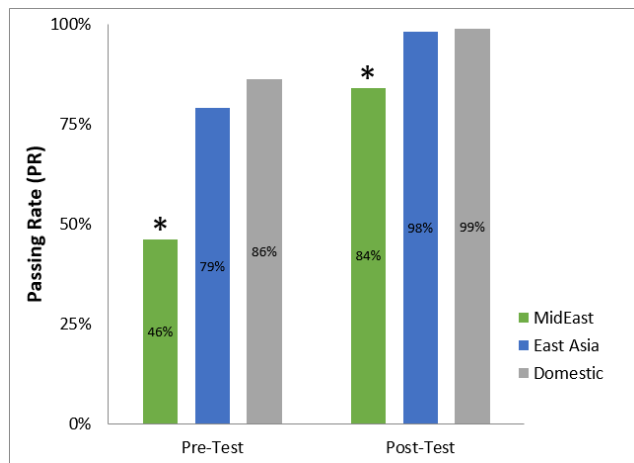


Fig. 2. Passing rates for Middle Eastern (MidEast), South and East Asia (East Asia), and domestic students. Asterisks indicate statistically significant differences ($p < 0.01$).

We studied the results from the Middle Eastern students in more depth by disaggregating the data by gender, as shown in Figures 3 and 4. These results reveal an exaggerated gap in SV skills between Middle Eastern female and male students. Just 26% of Middle Eastern women passed the SV pre-test, compared to 52% of Middle Eastern men (a 26% difference). And, the Middle Eastern women's pre-test PR scores were significantly lower than the pre-test PR scores for Middle Eastern men (14 to 20 points, a 6-point difference). This gender gap among Middle Eastern students is similar in magnitude compared to domestic students (a 3-point difference in PF and a 19% difference in PR). However, domestic women and men scored dramatically higher than their Middle Eastern counterparts on both the PR and PF metrics. In fact, the Middle Eastern women pre-test PF and PR were the lowest across all world regions and gender by a large margin. Across genders, the after-workshop post-test PF and PR scores were largely improved for all Middle Eastern students; male students improved their PF scores by 3 points and their PRs by 39% while female students improved their PF scores by 8 points and their PRs by 35%.

Looking deeper, a stark difference is observed in the post-test PRs between Middle Eastern women and all other students (Table II). All post-test PRs are greater than 90% for all tested student populations, except for Middle Eastern women who passed the PSVT at a rate of 61%. However, the Middle Eastern women are a small cohort of students ($n = 23$), and the PR metric can thereby be deceiving. A 61% post-test PR indicates that 14 Middle Eastern women passed the PSVT while nine women did not achieve a passing rate. The overall passing rate of 98% for the entire student cohort ($n = 2,441$) over six semesters likewise shows that only 49 students across the six semesters of workshop never passed the PSVT. To our dismay, Middle Eastern women make up 18% of all students who never passed the PSVT while they make up less than 1% of our student population. Middle Eastern women perform worse on the PSVT both before and after the workshop series compared to all other demographic and geographic cohorts. The SV instructor observed that, as a cohort, these students worked hard on their SV skill development during the workshops; their lower success rates did not appear to be associated with lack of motivation or commitment. He also observed that these students benefitted

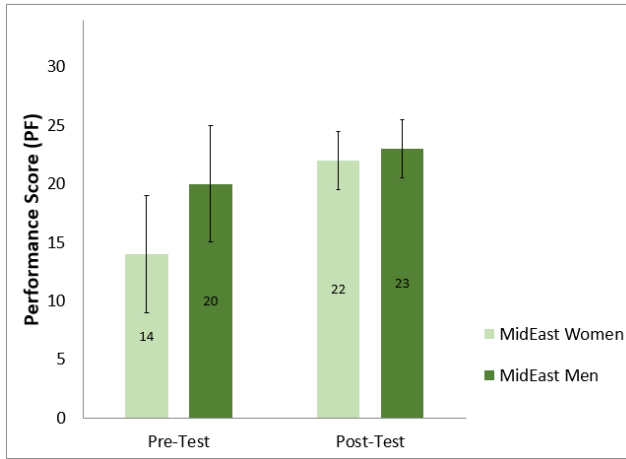


Fig. 3 Performance scores for Middle Eastern (MidEast) women and men for the pre- and post-test. Asterisks indicate statistically significant differences ($p < 0.01$).

from additional practice with skills like 3D sketching, the right-hand-rule, and others.

V. FINAL THOUGHTS

This study examined the differences in spatial visualization skills for students across world regions before and after a SV workshop series intervention. The results indicated that 1) Middle Eastern students achieve lower SV scores compared to their East Asian and domestic peers; 2) East Asian and domestic students perform at a statistically equivalent level; and 3) Middle Eastern women score lower than any demographic or geographic subset of the student population. These findings suggest that a concerted effort must be made to ensure equity in student success in engineering, especially for our (growing) Middle Eastern student population. Clearly, the SV workshop intervention is inadequate for this student population. Now that we know of this gap, our challenge is to find alternative approaches to aid this population and promote equity in student success.

Across both the PF and PR metrics, Middle Eastern students entered and left the SV intervention with lower skills compared to their East Asian and domestic peers. Middle Eastern students improved dramatically in their scores (PF of 19 to 23, PR of 46% to 84%) indicating that the SV workshop series was effective, but insufficient. However, as a group, the baseline SV skills of our incoming Middle Eastern engineering students are so far below our East Asian and domestic students that the four- (or eight-) week workshop series was not enough to “close the gap.”

Colleges of engineering may inadvertently admit students into programs in which they are already at a disadvantage since SV skills are not assessed or indicated during the admissions process. With the ample research concerning the gender gap in SV skills and now the growing research indicating that specific international student cohorts perform dramatically differently from their domestic counterparts, administrative reform is called for to ensure equity in student success within our colleges of engineering.

Finally, we conducted an exploratory qualitative investigation of Middle Eastern students’ experiences with SV

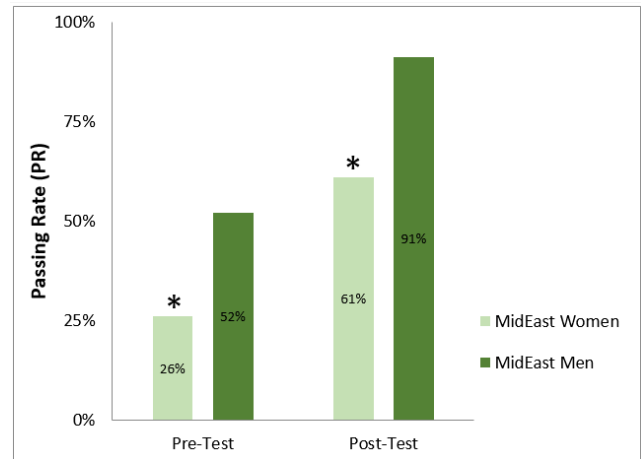


Fig. 4 Passing rates for Middle Eastern (MidEast) women and men for the pre- and post-test. Asterisks indicate statistically significant differences ($p < 0.01$).

skills development. We were curious about their perspectives on SV skills, their educational backgrounds, and any insights into which aspects of the SV intervention were most valuable. We interviewed two undergraduate engineering Middle Eastern students, Karim and Hana (pseudonyms). Karim is a junior civil engineering major who participated in the SV intervention in Spring 2015 (two years prior to the interview); he scored 17 on the PSVT pre-test and 23 on the post-test after a single four-week workshop series. Hana is a first-year architectural engineering major who participated in the SV intervention in Spring 2017 (same semester as the interview); she scored 13 on the PSVT pre-test, 17 after the first series of SV workshops, and a passing 24 after the second series of SV workshops. Both volunteered to participate in an hour-long interview that was approved by the university’s Institutional Review Board.

Karim is interested in architecture and discussed how he has always loved skyscrapers. He thought he was good at spatial visualization; however, he did not pass the PSVT on his first attempt. At first, Karim did not remember much about the SV workshops, but throughout the discussion more and more memories emerged. He recalls thinking that “I actually thought I would do good, but I did very bad,” and that “[SV] is something that I say is very easy for me, in my mind, to think about it.” He also remembers seeing his progress and knowing that he had improved during the workshop. “I’m pretty sure [I improved] because of the workshop because I thought I remember seeing the bottom, how it changed.” Karim described further how the SV workshop teaching assistant helped him with a strategy using the X-Y-Z coordinate system to rotate the object in his mind. While Karim did not provide details concerning his educational background in SV, his interview provided confirmation that students notice their improvement in SV skills during their time in the SV workshop series.

Hana provided a more in-depth description of her SV workshop experience likely because she had just finished the workshop series when the interview was conducted. She described how SV was a skill she had not thought about since playing with blocks in elementary school. She thought she would pass the SV test initially since she already had an interest

in architectural engineering, but was saddened when she scored 13 out of 30 on her pre-test. Initially she thought, “Oh, that will be easy, like, this is a children thing,” but then she noticed that “I got a really bad headache that I—like, I don’t know, I can flip it with my hand but I can’t see the final result in my eyes or in my brain.” While describing her experience in the SV workshops, Hana explained that “[the instructor] gave us really strict and straightforward rules, what we should look at when we see an object, what should we look [sic] first, and then second, and then third. And, I also learned the... about the components, the X, Y and Z. This is really tremendously helpful.” She reflected that “I really see a big and huge change from just the beginning of the class until today.” Hana confirmed that her educational background did not include practicing SV skills (at least since elementary school) and that she imagined herself to have strong SV skills when arriving at engineering college. She went on to detail many ways in which she looks at the world differently after practicing these SV skills in the workshop series.

These initial qualitative findings provide ample motivation to continue with our SV intervention and to investigate further—both quantitatively and qualitatively—the source of the dramatic differences in SV skills among our Middle Eastern students, and what can be done to provide equity in access to a US engineering education.

VI. ACKNOWLEDGMENT

The authors express their appreciation to Dr. Sheryl Sorby for her extensive work on this topic, which provided a foundation on which we could build, and to Dr. P.K. Imbrie for early research guidance. The authors also thank WEPAN and the ENGAGE Engineering initiative for inspiration. Finally, the authors thank Denise W. Carlson for her insights and critique of the manuscript.

VII. REFERENCES

- [1] S. A. Sorby, “Educational Research in Developing 3-D Spatial Skills for Engineering Students,” *Int. J. Sci. Educ.*, vol. 31, no. 3, pp. 459–480, Feb. 2009.
- [2] J. L. Segil, J. F. Sullivan, B. A. Myers, D. T. Reamon, and M. H. Forbes, “Analysis of multi-modal spatial visualization workshop intervention across gender, nationality, and other engineering student demographics,” in *Frontiers in Education Conference (FIE), 2016 IEEE*, 2016, pp. 1–5.
- [3] “College Profiles: American Society for Engineering Education.” [Online]. Available: https://www.asee.org/papers-and-publications/publications/college-profiles#The_Profiles_and_Data_Book. [Accessed: 04-Apr-2016].
- [4] S. A. Sorby and N. Veurink, “Spatial skills among minority and international engineering students,” *age*, vol. 25, p. 1.
- [5] S. A. Sorby, S. Cubero, N. Pasha-Zaidi, and H. Karki, “Spatial skills of engineering students in the United Arab Emirates,” *QScience Proc.*, vol. 2015, no. 4, p. 32, Jun. 2015.
- [6] S. S. Metz *et al.*, “Implementing ENGAGE Strategies to Improve Retention: Focus on Spatial Skills Engineering Schools Discuss Successes and Challenges,” in *American Society for Engineering Education*, Vancouver, BC, 2011.
- [7] “The World Bank - Countries by Region.” [Online]. Available: <http://www.worldbank.org/en/country>. [Accessed: 04-Apr-2017].
- [8] J. Segil, B. Myers, J. Sullivan, and D. Reamon, “Efficacy of Various Spatial Visualization Implementation Approaches in a First-year Engineering Projects Course,” presented at the ASEE Conferences, 2015, pp. 26.590.1–26.590.8.

The Development and Assessment of a Course for Enhancing the 3-D Spatial Visualization Skills of First Year Engineering Students

SHERYL A. SORBY

*Department of Civil and Environmental Engineering
Michigan Technological University*

BEVERLY J. BAARTMANS

*Department of Mathematical Sciences
Michigan Technological University*

ABSTRACT

In January 1993, we received NSF funding to develop a pre-graphics course for freshman engineering majors who are weak in 3-D spatial visualization skills. A text and computer lab exercises utilizing I-DEAS software were written specifically for this course. The course is 3-credits (quarter system) with two hours of lecture and two hours of computer lab each week. It was offered at Michigan Technological University (MTU) for the first time during the 1993 Fall term and has been offered each fall since that time. The objective of the course is to provide the prerequisite spatial skills needed by students to succeed in their subsequent engineering graphics courses. Assessment for the course has been continuous. Recently, a six-year longitudinal study was conducted to determine the overall success of this project. This paper will describe the project and the assessment findings from the longitudinal study.

I. INTRODUCTION

Visualization of problems is critical for success in engineering education. In most cases, it is an essential ingredient for student understanding. It is recognized that the ability to visualize is an important tool required of engineers in order to function effectively.¹⁻⁴ In addition to the traditional visualization tasks associated with engineering design, enhanced visualization skills are necessary to function in this new age of Computer Aided Design. In fact, Norman⁵ found that a person's spatial ability is the primary factor that explains differences in performance in fully utilizing computer-based technology. Unfortunately, at a time when visualization skills are increasingly important to students, engineering graphics (the primary course where students first learn visualization concepts) has been de-emphasized, and in many cases, dropped from engineering curricula altogether.⁶

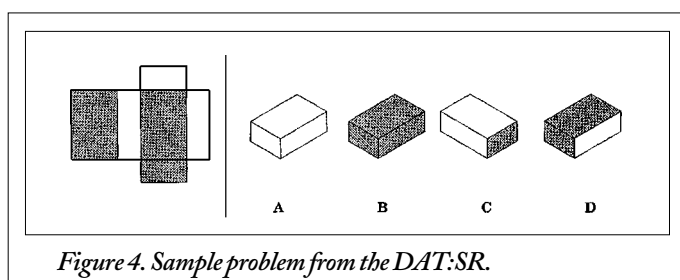
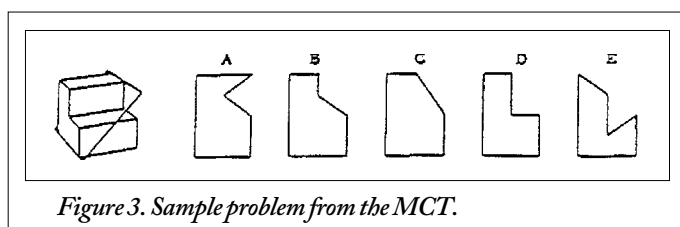
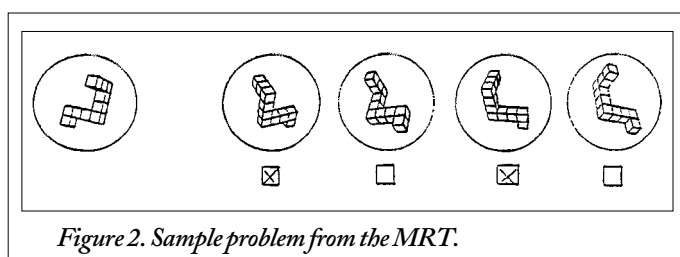
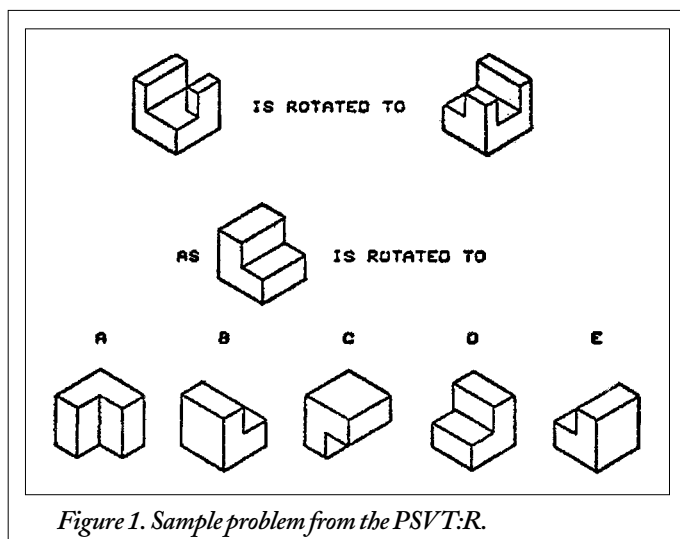
A. The Development of 3-D Spatial Ability

According to Piagetian theory, an individual acquires spatial visualization ability through three distinct stages of development.⁷ In the first stage, children learn topological spatial visualization where they are able to discern an object's topological relationship with other objects—i.e. how close the objects are to one another, an object's location within a group of objects, the object's isolation, etc. In the second stage of development, projective representation is acquired. At this stage, people are able to conceive what an object will look like from different perspectives. In the final stage of spatial visualization development, a person learns to combine projective abilities with the concept of measurement.

There are several standardized tests available to measure a person's ability across the first two stages of spatial development. For example, the Purdue Spatial Visualization Test: Rotations (PSVT:R) was devised to test a person's ability at the second stage of spatial development.⁸ A sample problem from the PSVT:R is shown in figure 1. This testing instrument was used throughout this project to identify students who have weaknesses in spatial visualization skills and partially to assess the impact of the experimental course.

The Mental Rotation Test (MRT)⁹ is another test designed to assess a person's ability to visualize rotated solids. It consists of 20 items where students are shown a criterion figure on the left and asked to identify which two of four given choices represent the same object after rotation in space. There are 40 points possible on the MRT and a sample problem is shown in figure 2. The Mental Cutting Test (MCT)¹⁰ was first developed as part of a university entrance examination in the USA and consists of 25 items. For each test problem, students are shown a criterion figure that is to be cut with an assumed plane. They must choose the correct resulting cross-section from among five alternatives. A sample problem from the MCT is shown in figure 3.

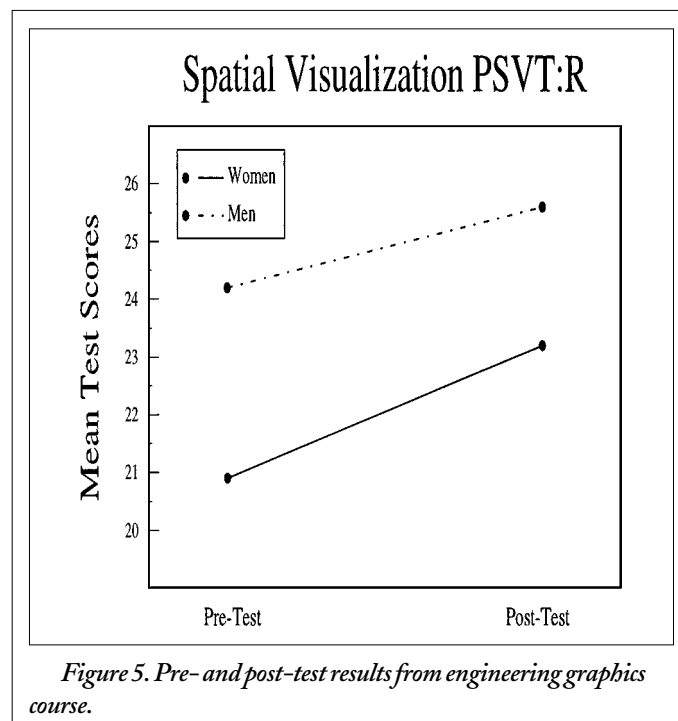
The Differential Aptitude Test: Space Relations (DAT:SR),¹¹ consists of 50 items. The task is to choose the correct 3-dimensional object from four alternatives that would result from folding the given 2-dimensional pattern. In one study,¹² it was found that a student's score on the DAT:SR was the most significant predictor of success in an engineering graphics course when compared to three other spatial visualization tests that were given (including the MCT). A sample problem from the DAT:SR is shown in figure 4.



B. Background Research at MTU

It is well-documented that the 3-D visualization skills of women lag behind those of their male counterparts.^{3,13-16} Studies conducted at MTU by Gimmestad (now Baartmans) support these findings. A course such as the one described in this paper can help women students address a deficiency in their background so that they are more likely to succeed in their engineering studies (and in particular their design graphics courses). In fact, Hsi et al.⁴ found that a Saturday tutorial session on spatial strategy instruction significantly improved the performance of men and women students in an engineering graphics course.

In 1985, Baartmans¹⁷ conducted a research study at MTU. The sample in the study included 365 entering freshman (65 women and



300 men) who had declared Mechanical Engineering as their major. During freshman orientation, the students were given the PSVT:R. A multiple regression analysis established that a student's score on the PSVT:R was the most significant predictor of success in the freshman graphics course (ME105) out of the eleven predictors studied. Two other factors were found significant in predicting student success in ME105: 1) math ACT subtest score, and 2) a combination of prior experience in shop, drafting and solid geometry. Mean scores for women lagged behind mean scores for men on two of the three significant variables—spatial visualization as tested by the PSVT:R and prior years of experience in drafting, shop, and solid geometry. The mean score for women on the spatial visualization test (20.9 out of 30) was significantly lower than that for men (24.2 out of 30). Furthermore, it was expected that students would improve their spatial visualization ability as a result of instruction and other activities in the freshman graphics course. In this study, both genders did improve their performance on the spatial visualization test, however, the mean post-test score for women (23.3) was still significantly lower than that for men (25.6). These results are shown in figure 5.

II. PROJECT DESCRIPTION

During the Spring and Summer of 1993, we wrote a textbook to be used in our introductory 3-D spatial visualization skills course (GN102, Introduction to Spatial Visualization). This course is viewed as a pre-graphics course at MTU. The course topics include hands on construction activities, paper and pencil activities, and computer activities. These topics and activities are sequenced in logical order for the development of 3-D spatial skills. The topical outline for the ten-week course follows:

Course Outline

Week 1 Course Introduction. Students are introduced to the need for visualization skills in fields such as engineering, medicine,

architecture, chemistry, and mathematics. The three stages of spatial visualization development are discussed.

Week 2 *Isometric and Orthographic Sketching*. Students are given a set of snap cubes so that they can construct a building according to coded plans. Then they learn how to make isometric and orthographic drawings of the building using grid paper. The use of the snap cubes enables the students to hold a concrete model in their hands as they are making the sketches.

Week 3 *Orthographic Drawings and Applications*. Objects which contain inclined surfaces are demonstrated and orthographic and isometric drawings are made of these objects. Students are also instructed how to set up an engineering drawing in a standard layout.

Week 4 *Pattern Development*. Flat patterns which can be folded into 3-D solids are studied. Students are also introduced to a sheet metal application.

Week 5 *Two- and Three-Coordinate Drawing*. Students are shown the principle involved in locating specific points in space. Then they use a table of coordinate data to draw wireframe geometry. A surveying application using traverse data is introduced.

Week 6 *Translation and Scaling*. Object transformations in 3-space are introduced. Students are required to draw objects after translation and scaling.

Week 7 *Rotation of Objects*. Students work with objects created from snap cubes and sketch isometric views of the objects as they are rotated about one or more axes. These objects are rotated first about one axis and then about two or more axes.

Week 8 *Reflection of Objects and Applications*. Students use MirasTM in class to construct reflected views of objects. The concept of a plane of symmetry for an object is also introduced. Applications from organic chemistry involving reflected molecules are investigated.

Week 9 *Cross-Sections of Solids*. Students are taught to graph planes in 3-space. Cross-sections for cubes, cones and cylinders for cutting-planes of different orientations are discussed.

Week 10 *Surfaces and Solids of Revolution and the Intersection of Solids*. Students are required to sketch the surface/solid which would be formed by revolution of a planar figure/region about an axis. Conversely, given the surface/solid of revolution, they sketch the shape of the planar figure/region which was revolved. The intersection of solids and its use in Computer Aided Design is discussed.

A. Computer Lab

As a part of this project, various computer exercises were initially developed which utilize I-DEAS software as a visualization tool. The exercises were written to adhere closely to those topics covered in the textbook. In January of 1998, we received additional funding from the NSF to develop stand-alone multimedia software and an accompanying workbook to supplement our original textbook. Preliminary assessment results indicate that the multimedia software is likely to be an effective replacement for the original computer exercises as well as for the other course materials originally developed.

III. ASSESSMENT

In 1993, incoming students who were enrolled in the fields of mechanical, civil, environmental, geological, and general engineering were administered the PSVT:R and a background questionnaire as part of their freshman orientation. A total of 535 students took the test, 418 males and 117 females. The average percent cor-

rect for male students taking the test was 79.6% compared to an average of 68.1% for the female students. Furthermore, of the 45 students who received perfect scores on the exam, only 3 were women. Thus, 10.0% of the male students received perfect scores compared to only 2.6% of the women students. Conversely, of the 96 students who received scores of 60% or lower, 50 were male and 46 were women. In other words, only 12.0% of the male students failed the exam; whereas, 39.3% of the females failed the exam.

Although women made up only 22% of the group being tested, they were almost 50% of the group failing the exam. Further statistical analyses of the data from the background questionnaire and the PSVT:R test scores revealed four significant predictors of success on the PSVT:R¹⁸ out of eleven factors studied. They were: 1) play as children with construction toys such as LegosTM, Lincoln LogsTM, Erector SetsTM, 2) gender, 3) math ACT scores, and 4) previous experience in design-related courses (like drafting, mechanical drawing, CAD, and art). Furthermore, male/female differences on predictors 1, 3 and 4 were tested for statistical significance. Play with construction toys and previous experience in design courses were found to be gender-biased (i.e., average scores for women on these variables were significantly lower than for men on these two predictors). This means that men were more likely than women to have participated in those activities that were found to be helpful in the development of spatial skills (i.e., play with construction toys and previous drafting course). Average scores on the ACT Math subtest did not differ significantly for men and women. The predictors which were not significant for a person's PSVT:R score were: 1) age, 2) right vs. left handedness, 3) previous experience in high school geometry courses—nearly all of the students had taken high school geometry, 4) participation in industrial arts courses in high school, 5) playing video games, 6) previous work experience involving spatial skills, and 7) participation in sports which involved placing an object in a specific location (e.g., basketball, hockey, etc.)

A. Gain Scores on Spatial Tests

Since its initial offering in the fall of 1993, the course has been taught a total of six times. Each time, the students have been administered the PSVT:R as a part of their final exam (post-test) and their gain scores analyzed. In addition, in the fall of 1996, 1997 and 1998, the students were pre- and post-tested using three other exams also designed to assess their spatial abilities. These exams include the Mental Rotations Test (MRT), the Mental Cutting Test (MCT), and the Differential Aptitude Test: Space Relations (DAT:SR). Gain scores on these exams for the students enrolled in the course have been analyzed, and these results will be presented in this paper.

The PSVT:R has been the primary instrument used both to recruit students for the course and to assess its effectiveness. Figure 6 shows the pre- and post-test results over the past six years for students enrolled in the course. Raw scores are indicated on the figure in parentheses. As illustrated by the data presented in figure 6, the average pre-test score across the six-year period has been approximately 50% compared to an average post-test score of approximately 80%. Dependent t-tests were used to analyze the average gain scores for the students in the class. The cumulative data over the past six years is shown in table 1. (Recall that the number of points possible on the exam is 30.)

Statistically significant gains on the average PSVT:R scores were made by students enrolled in GN102 for each of the six years.

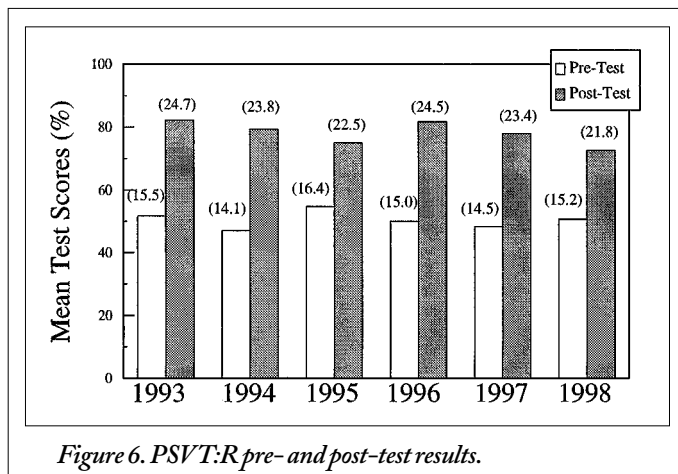


Figure 6. PSVT:R pre- and post-test results.

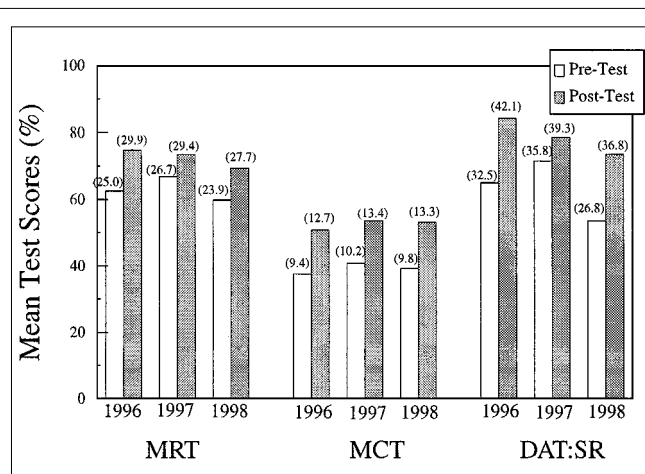


Figure 7. Pre- and post-test results.

Year	Gain (S. Dev.)	t-value	Level of Significance
1993 (n=24)	9.17 (3.58)	12.5	p<0.0001
1994 (n=16)	9.69 (3.62)	10.7	p<0.0005
1995 (n=47)	6.66 (3.74)	12.2	p<0.005
1996 (n=26)	9.42 (3.24)	14.8	p<0.0005
1997 (n=27)	8.89 (4.66)	9.91	p<0.0005
1998 (n=36)	6.53 (3.75)	10.46	p<0.0005

Table 1. PSVT:R average gain scores.

We believe that the post-test gains are due to the positive effects of the course rather than any practice benefit from having used the PSVT:R as both a pre- and as a post-test. Since these tests were administered almost 3 months apart, gains in score due to a practice effect should be negligible. Stanley et al.¹⁹ states that the average gain due to the practice effect for tests administered approximately 3 months apart is usually 0.2σ or less. However, the average gain scores for GN102 students were 2.56σ , 2.68σ , 1.78σ , 2.91σ , 1.91σ , and 1.74σ respectively. Thus, the average gain scores we determined are a factor of ten larger than would be expected from the practice effect alone.

Since the fall of 1996, the MRT, MCT and DAT:SR spatial visualization exams have also been administered to the students in GN102 as both a pre- and a post-test. The MRT has 40 possible points, the MCT has 25 possible points and the DAT:SR has 50 possible points. Figure 7 shows these test results over the past three years (raw scores are indicated in parentheses). Dependent t-tests were also used to analyze the average gain scores for the students in the class. The data for each test and year is presented in table 2. As

Test	Year	Gain (S. Dev.)	t-value	Level of Significance
MRT (40 pts)	1996 (n=26)	4.85 (5.60)	4.41	p<0.005
	1997 (n=26)	2.65 (9.23)	1.47	0.05<p<0.1
	1998 (n=36)	3.78 (6.12)	3.71	p<0.0005
MCT (25 pts)	1996 (n=26)	3.27 (3.05)	5.46	p<0.005
	1997 (n=26)	3.15 (3.64)	4.42	p<0.005
	1998 (n=36)	3.51 (2.93)	7.08	p<0.0005
DAT:SR (50 pts)	1996 (n=26)	9.54 (6.91)	7.04	p<0.005
	1997 (n=27)	3.52 (6.65)	2.75	0.005<p<0.01
	1998 (n=36)	9.94 (5.97)	10.00	p<0.0005

Table 2. Average gain scores.

it can be seen from the data, the GN102 students routinely made statistically significant gains on each test that was administered.

1) Spatial test reliabilities: Kuder Richardson-20 (KR-20) test reliabilities were calculated for the pre- and post-tests administered in this study. The KR-20 is a measure of the internal consistency of a test. A KR-20 greater than 0.8 is generally acceptable to most researchers. Table 3 contains the KR-20 data for the tests administered in the Fall of 1997.

As it can be seen from the data presented in this table, the tests that were administered can be deemed "reliable" except for the

Test	Pre-Test	Post-Test
PSVT:R	0.83	0.71
MRT	0.95	0.83
MCT	0.64	0.57
DAT:SR	0.89	0.95

Table 3. KR-20 test reliabilities Fall 1997.

MCT and the PSVT:R post-test. The 1997 PSVT:R post-test reliability is not of concern to us because over the past five years of administering the PSVT:R, the KR-20 has generally been greater than 0.8. As a partial explanation for the low reliabilities on the MCT, it should be noted that the mean scores on both the pre- and the post-test were relatively low (around 50% or lower). Therefore, guessing may have been prevalent due to the difficulty of the test for our students. While the low reliabilities of the MCT with this group of students is of concern, the “noise” introduced by low test reliabilities effectively makes it more difficult for statistically significant gain scores to be attained. Other researchers such as Magin and Churches²⁰ have reported KR-20 MCT test reliabilities in the 0.86–0.89 range.

B. Transcript Analysis

In the Fall of 1993, 96 of the 535 freshman engineering students tested failed the PSVT:R during new student orientation. Of these students, a random sample was selected for participation in GN102. Twenty-four students enrolled in GN102 (experimental group) and the remaining 72 students who initially failed the PSVT:R and did not enroll in GN102 constitute the control group for this study. Since the fall of 1993, virtually all of the students in both the experimental and control groups have graduated and/or left the university. Transcripts of the students in the experimental and control groups have been analyzed regarding their success in engineering graphics courses as well as their overall success in the various engineering curricula. Student transcripts were analyzed to compare performance of the experimental group (EG) and the control group (CG) in several ways: 1) performance in subsequent engineering graphics courses, 2) retention rates and final choice of major, 3) grade point averages, and 4) overall success. During the fall of 1993, the CG consisted of a total of 72 students (40 men and 32 women), and the EG consisted of a total of 24 students (13 men and 11 women). Gender differences were also examined when appropriate. Each of the factors under consideration were examined separately.

1) Performance in engineering graphics: Engineering students at MTU enroll in various graphics courses depending on their choice of major. Mechanical engineering students enroll in ME104-Engineering Spatial Analysis and ME105-Graphical Communication in Engineering Design. Civil and environmental engineering students have had changing graphics requirements over the past six years. Initially, civil and environmental engineering students were required to take ME104-Engineering Spatial Analysis and CET-103-Computer Applications. These requirements

changed over time resulting in a graphics requirement of GN201-Introduction to Computer Aided Drafting and Design.

ME104 is primarily a course in descriptive geometry; ME105 a course in engineering graphics and conventional practices; CET103 a course in the use of AutoCAD; and GN201 is a solids-based course in computer graphics with components of traditional engineering graphics. Each of the courses is three credits (quarter system) except for CET103 which is only two credits. In addition, there are courses in drafting and computer graphics offered by the School of Technology at MTU taken by some of the students in this study.

Due to changes in major selection or to attrition, not all of the students in this study went on to take a subsequent graphics course; however, 19 of the 24 in the experimental group and 44 of the 72 in the control group took one or more of the graphics courses in subsequent terms. Students sometimes repeated a course if they failed initially or received a D or a D+ in a given course. In these cases, the first grade that they received in a course was used in the analysis. Grade points were assigned to the graphics grades received by the students in the following manner: A = 4, B+ = 3.5, B = 3, C+ = 2.5, and so on. The conglomerate grade point averages for all graphics courses were calculated. The average GPA for the EG was 3.03 (n=29) versus and average of 2.70 (n=73) for the CG. Table 4 details the average GPAs for the two groups by specific course (not including repeats). It is interesting to note that of the students in the CG, no students earned an A in either ME104 or ME105. This was not true of the students in the EG.

On average, students in the experimental group outperformed those in the control group in their subsequent graphics courses. The exception to this was in CET103 which was primarily a training course in the use of AutoCAD—little or no spatial skills were required to perform the tasks in the course. Although there was only one person in the EG who went on to take GN201, the average grade in the course for all students who enroll is typically around a 3.50. Thus, the fact that the average for the CG was 2.77 means that they are performing well below average in that particular course.

Of the 29 first-time graphics grades received by students in the experimental group, only 2 (6.9%) were below a C (i.e., D+, D, or F) compared to 9 out of 73 (12.3%) for the students in the control group. Thus, students in the CG were 50% more likely to do poorly in their subsequent graphics course. Furthermore, for the students in the EG there was only one repeat of a graphics course (5.3%) compared to five repeats for students in the CG (11.4%). Of the ten

Course	Control	Experimental
ME104	2.23 (n=26)	2.75 (n=10)
ME105	2.63 (n=15)	2.60 (n=5)
CET103	3.60 (n=15)	3.33 (n=12)
GN201	2.77 (n=11)	3.50 (n=1)

Table 4. Average GPA by course.

total students who received a low grade in their first graphics courses, five eventually dropped out of MTU and three stayed at MTU but left the College of Engineering. Thus, 80% of the students who struggled with their engineering graphics courses were not retained in engineering at MTU.

2) Retention rates and choice of major: Retention rates for students in the control group and in the experimental group were analyzed at the end of the Winter 98–99 quarter. Since MTU offers degrees primarily in technical fields, students who drop out typically do so for lack of a non-technical choice of degree (they also drop out for other reasons, but this is true for a large number of those who leave the university). For the CG, twenty-one (29.2%) of the students left MTU compared to six (25.0%) of the students in the EG. Of the twenty-seven students who left MTU, eight (29.6%) left engineering *before* dropping out of the university. Interestingly, the course seems to be having a positive impact on the retention rate of women. Table 5 shows the retention rates by gender for the students in the EG and CG.

Of the students who were retained at the university, retention rates within the College of Engineering were examined. In other words, what percentage of those who persisted at MTU of each group have completed an engineering degree? Of the 51 students in the control group who persisted at MTU, only 42 (82.3%) have remained in engineering majors. For the experimental group, 16 of the 18 (88.9%) who remained are still in engineering. Figure 8 shows the retention rates by gender for the students in this study.

Retention within a curriculum with a strong graphics background was also examined for those students who persisted within the College of Engineering. In other words, what percentage of students ended up in a major where visualization skills are not con-

sidered to be critical to student success? (Specifically, these “non-graphically oriented” programs are considered to be chemical and electrical engineering.) None of the students in the experimental group ended up in a non-graphical major compared to eight (19.1%) of the students in the control group. It should be noted that the students in this study were those who had initially enrolled in majors other than chemical or electrical engineering. Thus, it seems that the students in the EG are not discouraged from participation in programs which require well-developed spatial skills.

Overall, 69.2% of the men and 63.6% of the women in the experimental group remained at the university in a graphically oriented engineering program. This compares to 47.5% of the men and 46.9% of the women in the control group. Thus, the students in the EG tend to have better retention rates in every aspect under consideration (overall, within engineering, etc.).

3) Grade point averages: Average GPAs were computed for those students who either graduated from the university or who are still pursuing their degrees. There are no differences in overall average GPA between the two groups. The average GPA for the control group is 3.01 and for the experimental group it is 3.00.

Differences in GPA were found within the group of students who dropped out of the university. For those who left the university, the average GPA for the control group was 2.25 compared to 2.47 for the experimental group. However, for the students in the CG, nine of the twenty-one who dropped out (42.9%) had a GPA below a 2.0 compared to only one of six (16.7%) for the EG. Thus, it seems that the students in the CG left for academic reasons whereas the students in the EG may have left for other reasons.

4) Student success: Transcripts of the students were examined after the Winter 98–99 quarter. Since the students in the experimental and control groups started in the Fall of 1993, this means that transcripts were examined after an average of approximately 17 quarters of enrollment. Each academic year consists of 3 quarters (most MTU students do not attend summer course offerings). In theory, an MTU degree requires 12 quarters for completion, although it usually takes longer than this in the College of Engineering. The average number of years to completion for engineering students is 4.5 years or 13.4 quarters.

The average number of terms to completion were analyzed for each group. For students in the control group, it took 14.8 terms on average to complete their MTU degree which is identical to the average of 14.8 terms for those in the experimental group. This is interesting in light of the fact that in subsequent years, students who have failed the PSVT:R but who choose not to enroll in GN102 have based their decision on a desire “not to get behind” in the pursuit of their engineering degrees.

IV. CONCLUSIONS

The overall goals of this project were to develop a course which would help “low visualizers” overcome their deficiencies in 3-D spatial visualization and which would help them become more successful in their engineering studies than they would have been without the course. On a variety of spatial tests (the PSVT:R, MRT, MCT, and DAT:SR), students who complete the GN102 course are year after year showing statistically significant gain scores on these tests. Analysis of the transcripts of the students who completed GN102 in 1993 versus transcripts of students in a comparable

Gender	Control	Experimental
Male	70.0%	69.2%
Female	71.9%	81.8%

Table 5. Retention rates by gender.

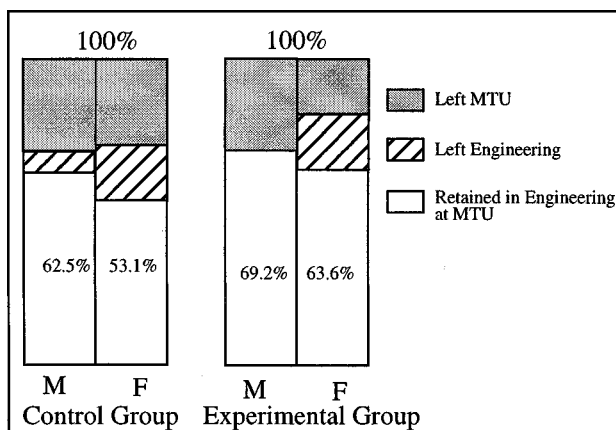


Figure 8. Retention statistics.

control group shows the GN102 students having higher graphics grades and higher retention rates in engineering.

ACKNOWLEDGMENTS

We would like to gratefully acknowledge the National Science Foundation for their support of this work through grants No. DUE-925407, DUE-9455143 and DUE-9752660.

REFERENCES

1. Miller, C.L., "Enhancing Spatial Visualization Abilities Through the Use of Real and Computer-Generated Models," *Proceedings, 1990 ASEE Annual Conference*, ASEE, 1990, pp. 131-134.
2. Pleck, M.H., "Visual Literacy-An Important Aspect of Engineering Design," *Proceedings, 1991 ASEE Annual Conference*, ASEE, 1991, pp. 1732-1734.
3. Deno, J., "The Relationship of Previous Experiences to Spatial Visualization Ability," *Engineering Design Graphics Journal*, vol. 59, no. 3, 1995, pp. 5-17.
4. Hsi, S., M.C. Linn, and J.E. Bell, "The Role of Spatial Reasoning in Engineering and the Design of Spatial Instruction," *Journal of Engineering Education*, vol. 86, no. 2, 1997, pp. 151-158.
5. Norman, K.L., "Spatial Visualization-a Gateway to Computer-Based Technology," *Journal of Special Educational Technology*, vol. XII, no. 3, 1994, pp. 195-206.
6. Pleck, M.H., et al., "Factors Affecting the Engineering Design Graphics Curriculum: Past, Present, Future," *Proceedings of the NSF Symposium on Modernization of the Engineering Design Graphics Curriculum*, Austin, Texas, August, 1990, pp. 43-52.
7. Bishop, J.E., "Developing Students' Spatial Ability," *Science Teacher*, vol. 45, 1978, pp. 20-23.
8. Guay, R.B., *Purdue Spatial Visualization Test: Rotations*, Purdue Research Foundation, West Lafayette, IN, 1977.
9. Vandenberg, S.G., and A.R. Kuse, "Mental Rotations, a Group Test of Three-dimensional Spatial Visualization," *Perceptual and Motor Skills*, no. 47, 1978, pp. 599-604.
10. *CEEB Special Aptitude Test in Spatial Relations*, developed by the College Entrance Examination Board, USA, 1939.
11. Bennett, G.K., H.G. Seashore, and A.G. Wesman, *Differential Aptitude Tests*, Forms S and T, New York, The Psychological Corporation, 1973.
12. Medina, A.C., H.B.P. Gerson, and S.A. Sorby, "Identifying Gender Differences in the 3-D Visualization Skills of Engineering Students in Brazil and in the United States," *Proceedings of the International Conference for Engineering Education*, Rio de Janeiro, August, 1998, CD-ROM.
13. Stafford, R.E., "Hereditary and Environmental Components of Quantitative Reasoning," *Review of Educational Research*, vol. 42, 1972, pp. 183-201.
14. Hier, D.B., and W.F. Crowley, Jr., "Spatial Ability in Androgen-Deficient Men," *New England Journal of Medicine*, vol. 306, no. 20, May 20, 1982, pp. 1202-1205.
15. Fennema, E., and J.A. Sherman, "Sexual Stereotyping and Mathematics Learning," *The Arithmetic Teacher*, vol. 24, no. 5, May, 1977, pp. 369-372.
16. Peters, M., P. Chisolm, and B. Laeng, "Spatial Ability, Student Gender, and Academic Performance," *Journal of Engineering Education*, vol. 84, no. 1, 1995, pp. 69-73.
17. Gimmestad, B.J., "Gender Differences in Spatial Visualization and Predictors of Success in an Engineering Design Course," *Proceedings of the National Conference on Women in Mathematics and Science*, Sandra and Philip Keith, eds., St. Cloud State University, St. Cloud, MN, Nov., 1989, pp. 133-136.
18. Parolini, L.L., "Gender Differences on Predictors of Success on the Purdue Spatial Visualization Test: Rotations," Master's Thesis, Michigan Technological University, May, 1994.
19. Stanley, J.C., and K.D. Hopkins, "Educational and Psychological Measurement," Prentice-Hall, Englewood Cliffs, NJ, 1972.
20. Magin, D.J., and Churches, A.E., "Reliability and Stability of Two Tests of Spatial Abilities," *Proceedings of the 6th ICECGDG*, Tokyo, Japan, 1994, pp. 801-805.



The role of spatial training in improving spatial and calculus performance in engineering students



Sheryl Sorby^{a,*}, Beth Casey^b, Norma Veurink^c, Alana Dulaney^b

^a Mechanical Engineering-Engineering Mechanics, Michigan Technological University, USA

^b Program in Applied Developmental and Educational Psychology, Boston College School of Education, 201 Campion Hall, Boston College, Chestnut Hill, MA 02467, USA

^c Engineering Fundamentals, Michigan Technological University, 1400 Townsend Drive, Houghton, MI 49931, USA

ARTICLE INFO

Article history:

Received 23 April 2012

Received in revised form 20 February 2013

Accepted 26 March 2013

Keywords:

Spatial training

Gender

Engineering

Spatial skills

Calculus performance

ABSTRACT

Freshman engineering students who took a mental rotation (MR) test as a pretest at freshman orientation and as a posttest at the end of the first semester (675 students; 542 males, 133 females) were divided into intervention and comparison groups based on a pass/fail MR pretest cutoff score. Those who failed the test were all assigned to a spatial intervention consisting of a 1-credit course meeting weekly over the semester; those who passed were assigned to the comparison group. The present study used a regression discontinuity (RD) analysis to determine the effectiveness of the intervention. A treatment effect was found for posttest MR performance as there was a discontinuity or jump in the regression intercepts at the cutoff score of the pretest variable, with the intervention group performing at higher levels than would be expected if there had been no intervention. Using the same RD analysis, the intervention also showed transfer effects, improving calculus performance of the students in the intervention condition.

© 2013 Elsevier Inc. All rights reserved.

1. Introduction

The ability to visualize objects and situations in one's mind and to manipulate those images is a cognitive skill vital to many career fields, especially those requiring work with graphical images (Smith, 1964). Evidence suggests that well-developed spatial skills of this type are critical to success in engineering, chemistry, computer science, mathematics, physics, medicine, dentistry, and many other fields (see Sorby, 2009 review). These types of spatial skills involve visualizing three-dimensional objects and perceiving what they will look like from different viewpoints or what they would look like if they were rotated or transformed in space. Mental rotation (MR) involves the ability to visualize what objects would look like if they were mentally rotated in space. Spatial visualization consists of the multi-step processing of spatial information, such as the ability to hold a shape in working memory and then search for the same shape hidden within a more complex figure, or to examine a group of shapes and then mentally combine them together to create a new design (Casey, 2013).

Spatial abilities have been widely studied and are known to be fundamental to higher-level thinking, reasoning, and creative processes. Spatial visualization and MR skills are particularly important to technical professions, such as engineering and computer science (Maier, 1994;

Norman, 1994). Spatial skills were shown to be the strongest predictor of success in using the computer interface in performing database manipulations (Norman, 1994), and are related to the ability to use computer-aided design software (Hamlin, Boersma, & Sorby, 2006).

Studies have shown quite clearly that students with high spatial ability scores perform better on organic chemistry questions requiring problem-solving skills (Pribyl & Bodner, 1987; Small & Morton, 1983). This was particularly true for questions involving the drawing or manipulation of molecular representations, and it was observed that students with higher spatial skills were more likely to draw correct structures and diagrams than those with lower spatial skills. These studies also noted that, as expected, spatial ability had little impact on those questions that required memorization or simple numerical procedures.

In physics, the ability to pictorially depict key variables and their relationships in physics problems distinguishes between expert and novice problem solvers (Taasobshirazi & Carr, 2008). "Before solving a problem, experts will represent the relationships in the problem by sketching a picture of the problem. Novices, in contrast, focus solely on representing the problem as a set of equations (e.g. Dhillon, 1998; Larkin, McDermott, Simon, & Simon, 1980). Pictorially representing problems before beginning to work on calculations is particularly important as problems become more complex and additional factors (e.g. angles, forces) begin to play a role in the problems" (Taasobshirazi & Carr, 2008, p. 155).

1.1. Review of the research on spatial skills and engineering

Engineers need strong spatial visualization skills to be able to represent and communicate their design ideas to others. The communication

* Corresponding author at: Engineering Education Innovation Center, 225 Hitchcock Hall, 2070 Neil Ave, Columbus, OH 43210, USA. Tel.: +1 906 231 7133; fax: +1 614 247 6255.

E-mail addresses: sorby.1@osu.edu (S. Sorby), caseyb@bc.edu (B. Casey), norma@mtu.edu (N. Veurink), dulaneya@bc.edu (A. Dulaney).

of a design is often done through the use of Computer Aided Design (CAD) models and 2-dimensional drawings. Engineers will either create the CAD models and 2-D drawings themselves or oversee their development, both of which require strong spatial visualization skills. Mechanical and biomedical engineers need to be able to visualize how parts will fit and move together. For example, when designing an artificial joint, biomedical engineers need to be able to visualize how the bone will fit inside the joint and the different ways the bone should be able to rotate inside the joint. They need to visualize where stresses in the joint will develop and how they will vary across the joint. Electrical and computer engineers need to visualize how to most efficiently assemble several small components when designing electronic equipment. They need to be able to graphically present data and visually interpret data. Civil engineers need to read maps and visualize 3-dimensional topography when designing transportation systems in order to design roadways that minimize cut and fill operations and have curves that are easily navigated.

The Johnson O'Connor Research Foundation (2005) tested nearly 32,000 individuals across the country (approximately half women and half men) on a number of cognitive variables, including spatial visualization factors. They found that the spatial visualization skills of engineers, both as students and as practitioners, are highly developed compared to other professions. It is unclear at this time what the causal nature of this relationship is (i.e., if people with well-developed spatial skills are attracted to engineering, or if engineering education and practice help to develop these critical skills).

Spatial skills have been found to be particularly important to success in engineering graphics. Students' scores on the MR component of the Purdue Spatial Visualization Test (PSVT: R) were shown to be the most significant predictor of success based on final scores in an engineering graphics course out of eleven variables tested (Gimmestad, 1989). Interestingly, component scores on the ACT (math, verbal, science, and composite) were not significantly correlated with success in engineering graphics.

1.2. Gender differences in spatial skills

Of all cognitive processes that have been investigated, spatial cognition shows some of the most robust gender differences favoring males, especially in the ability to mentally rotate 3-dimensional objects (see reviews by Voyer, Voyer, & Bryden, 1995; Halpern et al., 2007). This gender difference has been documented as early as age four, and thus, is present prior to formal schooling (Levine, Huttenlocher, Taylor, & Langrock, 1999). Furthermore, for girls in particular, these types of spatial reasoning skills are linked to math performance (Casey, Nuttall, Pezaris, & Benbow, 1995). Lippa, Collaer, and Peters (2010) assessed MR performance in more than 90,000 women and 111,000 men from 53 nations. In all nations, men's mean performance exceeded women's. For both men and women, greater gender equality and higher economic development were associated with better spatial skills. This may have something to do with greater dependence on spatial skills in the workforces of increasingly technological countries. What is surprising, however, is that *greater gender equality* (as assessed by the United Nations indices) and *higher economic development* (as assessed by per capita income and life expectancy) were positively associated across nations with *larger sex differences*. It seems that increasingly developed and technological societies produce wider gender gaps in spatial skills.

1.3. Women in engineering

Given the importance of spatial skills for solving different types of engineering problems, it is not surprising that engineering remains one of the most male-dominated of all STEM fields. In 2005, 20% of engineering BS degrees were awarded to women, a modest gain from 17% in 1985. Likewise, in 2003, 11% of engineers in the workforce were women, up from 9% in 1993. Some areas of engineering

(environmental and biomedical) attract nearly equal numbers of men and women; however, these are among the smallest of engineering disciplines. The two largest engineering disciplines, mechanical and electrical, routinely attract about 10% women to the field, resulting in about 20% of the national average of women enrolled in engineering across all disciplines. Since spatial skills in general, and MR skills in particular, are important to success in engineering, it is likely that improving women's spatial skills could help them succeed and persist in engineering. Further, spatial skills interventions would help men with low spatial skills who are majoring in engineering.

1.4. Training effects on spatial skills

Thus, an important question is whether or not spatial skills can be improved through training, particularly among engineering students. Multiple sources of evidence suggest that spatial abilities can, in fact, be developed through spatial activities and training. For example, meta-analyses of the literature on spatial experience (Baenninger & Newcombe, 1989; Uttal, 2009) indicate that: 1) participation in spatial activities such as sports, crafts, and other hobbies, is positively related to scores on spatial ability measures; and 2) that performance on spatial ability measures can be improved through training. In addition, investigations within specific domains of science and mathematics education have provided evidence that spatial skills can be improved through practice and instruction (Brinkmann, 1966; Lord, 1985). Uttal (2009) conducted a meta-analysis of the effects of training on gender differences in spatial skills. They found that, although gender differences in spatial skills were not eliminated as a result of training, both males and females showed substantial benefits from spatial training.

1.5. Spatial skills training with engineering students

1.5.1. The program of research at Michigan Tech

In 1993, a spatial skills training course was developed and implemented at Michigan Tech to help first-year engineering students develop their 3-D spatial skills. The training course included topics such as isometric and orthographic sketching, flat pattern development, and rotation of objects. During freshman orientation in 1993, students who had declared majors of mechanical, civil, environmental, metallurgical, or general engineering were administered the Purdue Spatial Visualization Test: Rotations (PSVT: R) (Guay, 1977). In this initial year, a total of 96 out of 535 students failed the PSVT: R with a score of less than 60%. In a pilot study, a random sample of 24 students who failed the PSVT: R were selected for participation in the experimental course and the remaining 72 students became the comparison group. In examining the results from the PSVT: R for this initial group of students, the following observations were made. Although women made up only about 17% of the group taking the PSVT: R, they constituted about 43% of the group failing the test, making women nearly three times as likely to fail the PSVT: R than their male counterparts. From this study, it also appeared that the spatial intervention had a positive impact on student success; however, the sample size was small.

Consequently, from 1994 to 2008, engineering students were given the PSVT: R during orientation and those who failed the exam with a 60% or lower were encouraged to enroll in the spatial skills course. The school administration did not allow for random assignment of failing students to intervention and comparison conditions. Instead, the failing students who chose to enroll in the training course were compared to the students who opted out of the course. At the end of the semester, test scores indicated that the students in the training course had experienced significant gains in spatial skills. Also, the retention rates at Michigan Tech were higher for those students who participated in the course than they were for those who initially failed the PSVT: R and who did not enroll—76.8% versus 70.0% for male students and 87.4% versus 71.7% for female students

(Sorby, 2005b). Finally, the overall grade point average of students who opted to take the spatial training course was significantly higher than those who chose not to take the course (Sorby, 2005b).

1.5.2. Spatial training studies

There are very few other published spatial training studies specifically using engineering students beyond that of the program of research conducted at Michigan Tech (Uttal, 2009). Following the design of Sorby (2005a), one group of researchers examined the effects of spatial training on students who had initial low spatial scores (Onyancha, Derov, & Kinsey, 2009). They studied the effects of targeted training in mechanical engineering students through a computer assisted design (CAD) course (Onyancha et al., 2009). They identified students who failed the MR test at the beginning of the course, and offered spatial training to these students. The researchers found that the targeted training produced a significant improvement in the spatial ability test scores of the failing students who chose to take the training compared to the students who selected out of the training. Using a briefer spatial intervention, Hsi, Linn, and Bell (1997) identified students at risk for poor spatial skills in an introductory computer aided design course and provided targeted spatial training. They found that pre-course gender differences were eliminated and that the overall course grade was improved as a result of the targeted spatial training.

1.5.3. Spatially based engineering courses

Some studies have not used targeted spatial interventions, but instead have examined the pre/post changes in spatial performance as a function of taking engineering courses that include a spatial component. For example, one group of researchers (Nemeth, 2007) assessed changes in a mental cutting task prior to and at the end of a year of an engineering program that included a course in descriptive geometry with perspective drawing. They found that learning technical drawing improves a person's ability for spatial visualization. They also found that the males showed greater improvement than the females by the end of the academic year. Prieto and Velasco (2010) administered spatial visualization and inductive reasoning tests at the beginning and end of a course in technical drawing in samples of first year engineering students. In both studies it was observed that a moderate percentage of students improved their visualization test performance; there was no improvement on the inductive reasoning test. The improvement on the spatial test was similar in men and women. These researchers concluded that their results support the conclusion that spatial visualization ability can be improved with training.

1.5.4. Purpose of the present research

A major problem with the prior research investigating the benefits of spatial training on engineering students' spatial skills is the possibility of alternative explanations. Previous studies have not included randomly assigned comparison groups to determine whether it was the specific components of the spatial intervention/engineering courses that resulted in improved performance, or whether the improvement occurred as a result of general experiences, a range of activities within the first academic year, or participant characteristics correlated with enrollment in spatial training. For example, in several previous research studies addressing this question in engineering students (Onyancha et al., 2009; Sorby, 2001a; Sorby, 2001b; Sorby, 2005b; Sorby & Baartmans, 2000), the intervention group consisted of failing students on the pretest who chose to take the intervention course, while the comparison groups consisted of failing students who chose not to take the course. Because of the lack of random assignment of participants to intervention and control conditions, there may have been a tendency for the more motivated students to choose to take the course. The use of a non-experimental design is problematic because it presents the potential for selection bias.

The present study was designed to examine the benefits of an intervention targeted to the freshmen engineering students who failed the initial spatial assessment during orientation at Michigan Tech.

However, the design of the present study required all students who failed the mental rotation test during orientation to enroll in the spatial intervention course. This enabled us to address the question of whether the spatial intervention was successful in raising students' spatial skills through an alternative non-experimental design. This type of design, called regression discontinuity (RD), allows for elimination of selection bias when implemented properly. In fact, in cases where a randomized design is not possible, an RD design is the recommended alternative to quasi-experimental and associational designs because it allows for an unbiased detection of treatment effects (Cook, 2008; Institute of Educational Sciences (IES), Technical Methods Report, 2008; Shadish, Galindo, Wong, Steiner, & Cook, 2011). The RD design is based on a pre-test–posttest treatment-comparison group design, in which individuals are assigned to a treatment condition based on a cutoff score from a pre-intervention measure. Participants scoring on one side of the cutoff receive the intervention while participants scoring on the other side of the cutoff do not receive the intervention. As long as assignment to the intervention and comparison conditions strictly follows the cutoff criterion, any selection effects correlated with the impact of the intervention are also perfectly correlated with the pre-intervention measure, which, when held constant in the statistical analysis, allows for an unbiased estimate of the intervention impact on a post-intervention measure (Shadish, Cook, & Campbell, 2001; Thistlethwaite & Campbell, 1960; Trochim & Cappelleri, 1992). Consequently, researchers are starting to use the RD design to obtain unbiased impact estimates of education-related interventions when random assignment is not possible, as recommended by recent IES requests for proposals (U.S. Department of Education & Institute of Education Sciences, 2008). Like the experimental design, the logic underlying the RD design supports statistically valid conclusions, as evidenced by statistical proofs (Cappelleri, 1991; Rubin, 1977).

The RD design in the present study used a MR test pretest as the assignment variable, with students scoring 18 or lower receiving the intervention, and those scoring 19 or higher not receiving the intervention. Therefore, only the students who needed the intervention actually received it, based on the pretest cutoff and in line with the RD design assignment protocol. The MR measure was also administered to all participants at posttest, following completion of the intervention. To estimate the intervention impact, posttest scores can be regressed on pretest scores and on a dichotomous intervention group membership variable. The pretest assignment variable must be centered at the cutoff score in the model, which allows for an estimate of the intervention impact at the cutoff score by detecting a discontinuity in the regression line between the intervention and comparison groups. If there were no effect of the intervention, then the regression line for the group scoring at or below the pretest cutoff (the intervention group) would be the same as the regression line for the group above the cutoff (the comparison group). A treatment effect is documented when there is a discontinuity or jump in the regression intercepts at the cutoff score of the pretest variable.

The present study utilized this RD analysis to determine whether the spatial intervention course would be effective in improving the spatial skills of freshmen engineering students who initially failed the spatial test administered at freshman orientation. We also used the RD approach to determine whether there was an interaction effect between gender and the intervention. (A detailed discussion of the RD design and analysis is described in the Research design and analytic approach section (Section 2.4) of this article).

A second series of questions in the present study relates to gender differences in spatial performance at pretest and posttest: First, do the female engineering students perform more poorly than the males on the pretest? Second, do the low spatial females benefit more from the intervention than the low spatial males?

A further purpose of the present study was to examine whether the benefits of the spatial intervention course for the low spatial skills students had any wider transfer of learning effects beyond its specific

impact on their spatial performance. As is typical of most engineering programs, all students are required to take calculus in their undergraduate program at Michigan Tech. Many components of calculus depend on reasoning with visual representations. Bremigan (2005) has pointed out that “understanding fundamental calculus concepts (e.g., limits, derivatives, and integrals) requires the use of visual representations” (p. 249), while Zimmerman (1991) identified prerequisites for visual thinking in calculus, which include the ability to extract specific information from diagrams, an understanding of algebra and geometry as alternative languages for the expression of mathematical ideas, and knowledge of the rules and conventions associated with mathematical graphics. Thus, another question of interest in the present study is whether the students who received the spatial intervention would receive better calculus grades than would be expected by their initial low level of functioning on the spatial pretest. We conducted a second RD analysis using the same cutoff on the spatial pretest, but using calculus grades, rather than posttest spatial performance, as the dependent measure.

2. Material and methods

2.1. Participants

At Michigan Tech, orientation for first year students occurs during the week prior to the start of the academic year. Students who had declared a major in engineering were administered the PSVT: R during freshman orientation and at the end of the first semester. As part of informed consent, all students who chose to participate in the study had agreed to let these tests be included as group data in a research study. A total of 675 students (542 males; 133 females; 20% females) completed the test and signed the consent form during orientation.

Students who scored 60% or lower on the spatial pretest were required to enroll in a 1-credit course designed to help them improve their 3-D spatial skills. The course is described in a subsequent section of this paper. The spatial skills course has been in existence at Michigan Tech since 1993; however, prior to 2009, students who scored 60% or lower *elected* to take the course rather than being *required* to take the course. Thus, it is important to note that the students below the RD cut-off did not consider the spatial course to be part of a research study. Instead, they correctly assumed that this was part of the Engineering Program requirements for those who did not pass the test. In the fall 2009 semester, there were 87 students who enrolled in the spatial skills course and completed the pretest and post-test (46 males; 38 females).

2.2. Design

The study consisted of an intervention group and a comparison group. The intervention group included the students who scored at 18 or lower (60% or lower on the PSVT: R during orientation; in contrast to previous years at Michigan Tech, these students were required to enroll in the spatial skills intervention course. Participants in the comparison group scored 19 or higher on the PSVT: R during orientation.

2.2.1. Description of spatial intervention course

The 1-credit spatial skills intervention course met for one 1.5-hour session each week. The course materials consisted of a workbook and multimedia software developed by Sorby and Wysocki (2003). The session began with a short mini-lecture on the topic for the day. Following the mini-lecture, students worked through the software module either individually or in groups of two to three. Students finished the session by completing several assigned workbook pages. The software and workbook consisted of nine separate modules. Each module contained a background section as well as several exercises. The exercises in the software modules are primarily fill in the blank or multiple-choice. The exercises in the workbook require students to complete sketches

of simple objects in addition to fill-in-the-blank or multiple-choice exercises similar to those in the software. A tenth module is included in the course instruction with materials provided by the faculty teaching the course. The following list comprises the modules for the course:

- *Surfaces and Solids of Revolution*. In this module, 2-D shapes are revolved about an axis to form a 3-D surface or a 3-D solid.
- *Combining Objects*. Creating new objects by cutting, joining, or intersecting two objects is covered in this module.
- *Isometric Sketching*. Students learn about how coded plans are used to define objects constructed of cubes and blocks and are then required to sketch them from several corner views.
- *Orthographic Projection with Normal Surfaces*. In this module, students learn how to construct the top, front, and right side views of objects. Hidden features of objects are also discussed.
- *Orthographic Projection with Inclined and Curved Surfaces*. In this module, students learn how to construct the top, front, and right side views of objects that include inclined or curved surfaces.
- *Flat Pattern Developments*. This module covers the transformation from 2-D to 3-D for flat patterns that are folded up to create 3-D objects.
- *Rotation of Objects about One Axis*. In this module, students learn about object rotation and sketch objects as they are rotated about a single axis.
- *Rotation of Objects about Two or More Axes*. Continuing from the lessons learned in the previous module, students now rotate objects about two or more axes and sketch the result from this transformation.
- *Object Reflection and Symmetry*. This module covers the concept of reflecting an object across a plane. Also covered is the concept of object symmetry.
- *Cross-sections*. In this module, students imagine a cutting plane passing through an object to determine what the cross-section would look-like.

2.3. Measures

2.3.1. The pretest and posttest measures of mental rotation ability

The PSVT: R was the test used to measure spatial ability (30 items). The pretest was initially administered to first-year engineering students during freshman orientation in August to measure mental rotation (MR) ability (Guay, 1977). The posttest was administered near the end of the first semester in the 13th or 14th week. It is a time-limited paper-and-pencil test that requires the 3-D mental rotation of objects in space. In this task, the students are asked to: (1) study how the object in the top line of the question is rotated, (2) picture in your mind what the object shown in the middle line of the question looks like when rotated in exactly the same manner, and (3) select from among the five drawings (A, B, C, D, or E) given in the bottom line of the question, the one that looks like the object rotated in the correct position (Bodner & Guay, 1997). A sample problem from the PSVT: R is shown in Fig. 1.

Within Computer Science, Engineering, and Technology education, the PSVT: R is considered the “gold standard” of spatial assessments (Waller & Lourenco, 2010). The test has good reliability and validity with split-half reliabilities of .78–.85, indicating good internal consistency. Criterion validity was shown by a correlation of .61 between the PSVT: R and the Vandenberg Mental Rotation Test, another widely used measure of mental rotation ability. It has also been shown to predict problem-solving skills in chemistry, physics, and engineering (Gimmestad, 1989; Small & Morton, 1983; Taasobshirazi & Carr, 2008).

2.3.2. The ACT math test

The American College Testing (ACT) math test is a standardized test used in college admissions (ACT, 2007). The math test is a 60-minute, 60-question math test covering algebra, plane and coordinate geometry,

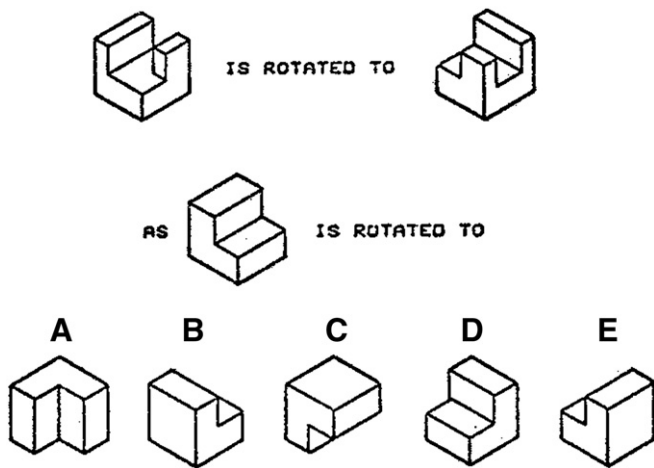


Fig. 1. Sample problem from the Purdue Spatial Visualization Test: Rotations (PSVT: R).

and elementary trigonometry. Students took the exam as part of their admissions into the Engineering program at Michigan Tech. The ACT Technical Manual (ACT, 2007), reports internal reliability coefficients between .87 and .92. Validity was assessed using multiple regression analysis, and the model for the ACT mathematics subtest score showed the greatest prediction accuracy ($R^2 = .65$), compared to models for ACT scores in other content areas.

2.3.3. Calculus grades

The Calculus with Technology I courses are offered in the Fall and Spring semesters of the freshmen year for the Michigan Tech Engineering students. The content is an introduction to single-variable calculus, which includes a computer laboratory. Topics include trigonometric, exponential, and logarithmic functions, differentiation and its uses, and basic integration; it integrates symbolic tools, graphical concepts, data and numerical calculations. The grade range include: A (4), AB (3.5), B (3), BC (2.5), C (2), CD (1.5), D (1), F (0). For all students in the study, the percent of students who received each grade consisted of: A = 19.8%, AB = 14.9%, B = 13.4%, BC = 11%, C = 12.2%, CD = 6.8%, D = 6.4%, F = 15.5%.

2.4. Research design and analytic approach

It was not possible to randomly assign participants to intervention and control conditions in the present study. However, as discussed previously, a standard quasi-experimental design is problematic because it presents selection bias. To address this issue, the present study used an RD design. Accordingly, group assignment was strictly determined by the pretest cutoff, without any participant crossover between groups. The PSVT: R served as the pretest assignment variable, with 18 as the cutoff score, and students were unaware of the assignment criterion for the intervention course. As a result, it was possible to completely account for differences between groups by statistically controlling for the pretest variable in the analysis.

This cutoff score was chosen based on the university's pass-fail criteria of spatial skills for incoming freshman. Often, the cutoff score for an RD design is the mean of the pretest measure because the high distribution of scores at the mean maximizes statistical power, but this is not a requirement. The pretest cutoff score in the present study is not far from the sample mean ($M = 23.97$, range = 29–11) and holds the potential for more meaningful university policy implications. Students who fail the MR pretest demonstrate a weakness in skills crucial for their success in the engineering program, and, therefore, require remediation of these skills. A significant intervention impact estimated at the MR test cutoff provides an empirically-based solution for this targeted population of students that aligns with university policy.

Approaches to obtaining counterfactual evidence of an intervention impact and to ruling out selection bias differ between the RD design and a typical (cross-sectional) quasi-experimental design. Accordingly, the present study calls for a different analytic approach than that typically used in a cross-sectional design. While the latter could detect an intervention effect by comparing the means of the intervention and comparison groups on the posttest (e.g., with a t-test or ANOVA), this analysis would not be effective for an RD design. This is because the mean posttest score of the intervention group is not expected to be higher than the mean of the comparison group; after all, participants were assigned to the intervention group based on their low pretest scores in relation to the cutoff. Instead, linear regression is used to determine if, at the cutoff score, the predicted outcome of the intervention group is expected to be higher than the predicted outcome of the comparison group.

The cross-sectional quasi-experimental design attempts to maximize the equivalence of the intervention and comparison groups on the pretest measure—as is done in randomized experiments—in order to rule out alternative explanations of between-group differences on the posttest. In other words, group equivalence on the pretest provides counterfactual evidence. However, in an RD design, intervention and comparison groups are not intended to be equivalent on the pretest—in fact, group differences on the pretest measure are maximized. This is because the counterfactual evidence in the RD design is found in the predicted outcome for the comparison group at the cutoff, where, without the intervention, the predicted posttest scores would be linearly continuous with those of participants in the intervention group. In addition, when a quasi-experimental design is used, selection bias must be controlled by measuring any participant characteristics known to bias the detection of the intervention impact, and statistically holding these variables constant in the analysis. However, in the RD design, as previously discussed, potential bias of the intervention impact estimate is captured solely by the pretest variable used to assign participants to groups. Therefore, this variable must be held constant when estimating the intervention impact, which makes OLS regression and ANCOVA better-suited analyses for the RD design than a t-test or ANOVA.

In addition to eliminating selection bias, the RD design also avoids problems of regression to the mean presented by a cross-sectional analysis, because regression to the mean is expected across the range of the pretest scores, and is described by the regression line itself. While regression to the mean influences the slope of the pretest variable, it does not influence the estimate of the regression intercepts at the cutoff score (Shadish et al., 2001). Therefore, without selection effects and regression to the mean, the RD design avoids threats to internal validity that are inevitably posed by the cross-sectional design, making it the best approach for the present study (for a complete discussion of the RD design and internal validity, see Shadish et al., 2001).

To detect an unbiased treatment effect under this design, scores of a post-intervention measure are regressed both on the pretest assignment variable—with observed scores centered at the cutoff score—and a dummy-coded intervention assignment variable. Covariates of the assignment variable can also be included in the model but are not required for reducing selection bias, unless they also co-vary with intervention group membership; in this study, there are no covariates of intervention group membership other than the pre-test assignment variable. By centering the assignment variable at the cutoff score, and including this variable as a predictor in the regression equation, the regression coefficient for the intervention group-membership predictor is estimated at the cutoff score, rather than at $X = 0$. Consequently, the regression coefficient for the intervention variable becomes statistically synonymous with measuring a discontinuity in the regression line at the cutoff score. Thus, a treatment effect is documented when there is a statistically significant regression coefficient for the intervention group-membership variable (Thistlethwaite & Campbell, 1960).

In the present study, a regression model that approaches the cutoff ($X = 18$) from the left corresponds to the intervention group, while the regression model that approaches the cutoff from the right

corresponds to the comparison group. If no intervention had been administered, or if the intervention has no impact, predicted outcomes would be linearly continuous for values above and below $X = 18$; conversely, an effective intervention results in significantly different predicted outcomes at $X = 18$ between the intervention and comparison groups.

3. Results

3.1. Preliminary analyses

Descriptive statistics for participants' scores on the pretest and posttest measures of MR, as well as calculus grades, split by intervention group membership and gender, are provided in Table 1. The effect sizes relating to gender differences, as measured by Cohen's d -statistic, was $d = .78$, favoring males, for the pretest mental rotation score, and $d = .40$, favoring males, for the post-test mental rotation scores. Thus, the effect size was large for the pretest and moderate for the post-test. For the calculus grades, there were no meaningful gender differences, with an effect size of $d = .12$, favoring females. It should be noted that 28% of the freshmen engineering females failed the spatial pretest, while only 7% of the females failed the posttest (MR score below 19); 9% of males failed the pre-test, and 2% of males failed the posttest.

3.2. Meeting evidence standards for a regression discontinuity analysis

In order for the RD analysis to yield unbiased intervention impact estimates, certain evidence standards must be met (Schochet et al., 2010; Shadish et al., 2001). The analyses reported in this section are a prerequisite to conducting the main analysis, reported in Section 3.3. The first standard is the institutional and statistical integrity of the assignment variable, which prohibits any systematic manipulation of the assignment variable. This standard parallels pretest equivalence of assignment groups in a randomized controlled trial. Second, the study should not have high overall or differential levels of attrition. Third, the functional form of the assignment-outcome variable relationship must be properly specified in the statistical model used to test the impact of the intervention, and the bandwidth of the assignment variable must be appropriate for sample selection. These standards are addressed in sequence below.

The integrity of the assignment variable was determined using a chi-square goodness-of-fit analysis, which tested for a statistically significant difference in the frequency of scores just above and below the pre-test cutoff score. A non-significant difference provides evidence that participant pre-test scores were not manipulated in order to allow or deny any individuals assignment to a particular condition; a statistically higher density of scores either just above or below the cutoff would suggest that individual scores had been manipulated. Here, a

null finding indicates that there are a statistically equal number of participants scoring just below the cutoff (at score 18) and just above the cutoff (at score 19). The results of this analysis were, indeed, non-significant, $\chi^2(1, 77) = 1.57, p > .05$, thereby showing that the data fit this criterion.

Overall and differential attrition of the sample was assessed by creating a dummy variable for participant dropout (0 = completed pretest and posttest and 1 = dropped out), and conducting a chi-square test of association. The overall attrition rate for the sample of 742 participants who completed the MR pre-test was low at 9.4%. This analysis found a non-significant difference in dropout frequency between the intervention (10%) and comparison groups (5%), $\chi^2(1, 742) = 2.72, p > .05$. This finding demonstrates that the evidence standard for attrition has been satisfied in the present study.

Third, the functional form and bandwidth standard was assessed using a systematic RD model over-specification procedure. This process began with over-specifying the regression model used for the main analysis of the intervention impact, followed by paring down of non-significant terms in subsequent models. This process aids in determining the functional form of the data and is a prudent approach, as it ensures limited bias in the final parameter estimates. Over-specification is typically used to test for effects of higher order polynomials and their interaction with the pretest MR scores (Shadish et al., 2001). This can help in determining whether the final model is properly specified.

The over-specification procedure started with a visual inspection of a scatterplot with the pretest and outcome variables, with regression lines fitted separately for each intervention group. If there is a visual indication that the functional form of the data may be cubic or quadratic, these terms should be included in the initial models along with their corresponding terms for an interaction with the intervention group variable. If the visual inspection of the data is not suggestive of a non-linear relation, then the prudent approach for determining the proper model specification is to include a term for the interaction between the intervention variable and the centered pretest variable.

The visual inspection of the bivariate scatterplots (Figs. 1 and 2) indicated that a linear relation was most appropriate. Therefore, the proper specification of the regression models was determined by initially including the basic interaction terms. Thus, two regression analyses were run for the over-specification procedure: one with posttest MR scores as the outcome and once with Calculus grades as the outcome. Table 2 displays equations for these models. When MR posttest scores were the dependent variable, the term for the interaction between the pretest variable and the intervention group variable was non-significant, $\beta = 0.05, t(671) = 1.17, p > .05$. This was also the case for the model with calculus grades as the outcome variable, $\beta = 0.02, t(484) = 0.19, p > .05$. Thus, the over-specification procedure indicated that the functional form of the assignment-outcome variable

Table 1

Descriptive statistics for pretest assignment variable and posttest measures by gender and intervention group.

Measure	Comparison					Intervention					Total				
	<i>n</i>	Min	Max	<i>M</i>	<i>SD</i>	<i>n</i>	Min	Max	<i>M</i>	<i>SD</i>	<i>n</i>	Min	Max	<i>M</i>	<i>SD</i>
PSVT-R pretest															
Females	96	19	29	23.67	2.85	38	11	18	15.39	2.27	134	11	29	21.32	4.61
Males	496	19	30	25.37	3.07	46	11	18	16.57	1.82	542	11	30	24.62	3.86
Total	592	19	30	25.09	3.10	84	11	18	16.04	2.11	676	11	30	23.97	4.23
PSVT-R posttest															
Females	96	18	30	25.99	2.66	38	14	30	22.42	4.23	134	14	30	24.98	3.56
Males	496	17	30	26.68	2.89	46	16	30	22.59	3.19	542	16	30	26.33	3.13
Total	592	17	30	26.56	2.86	84	14	30	22.51	3.67	676	14	30	26.06	3.26
Calculus grade															
Females	60	0	4	2.67	1.36	27	0	4	2.41	1.31	87	0	4	2.59	1.38
Males	364	0	4	2.41	1.43	37	0	4	2.66	1.36	401	0	4	2.43	1.31
Total	424	0	4	2.52	1.29	64	0	4	2.62	1.35	490	0	4	2.46	1.33

Note. PSVT-R = Purdue Spatial Visualization Test: Rotations. PSVT-R pretest scores served as the assignment variable for intervention condition (cutoff = 18). PSVT-R posttest scores and calculus grades were dependent measures.

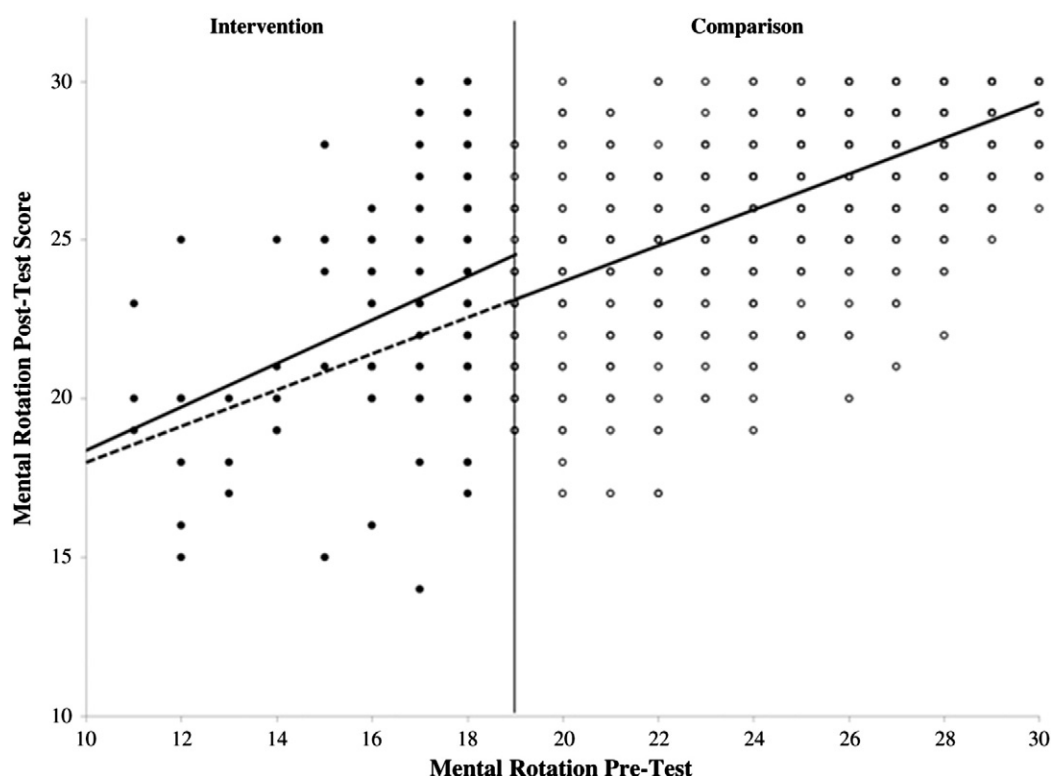


Fig. 2. Regression discontinuity of posttest mental rotation scores as a function of centered mental rotation pretest assignment variable and intervention group.

relationship is linear and that the impact of the intervention does not vary with different levels of the pretest variable (i.e., the intervention impact is uniform for all levels of X). Further, Fig. 2 reveals that this functional form applied to the complete bandwidth of the assignment variable.

In addition, we examined evidence that if an intervention had not been implemented, there would be a linearly continuous relationship between the outcome and assignment variables through the cutoff score. This preliminary analysis was conducted to confirm that there would have been no regression discontinuity if the intervention had not occurred; this is good practice for meeting evidence standards, but is not the main analysis for determining the intervention impact. This analysis provided evidence that there would not have been a discontinuity in the outcome variable without the intervention.¹

3.3. Main analyses

Given that the evidence standards for unbiased intervention impact estimates were met, two RD analyses were conducted to test the impact of the spatial intervention course on posttest MR scores and Calculus grades. This analysis was conducted with the total sample of freshmen

engineering students who gave permission to participate in the study, and who were available to take the posttest at the end of the Fall semester.

3.3.1. Posttest MR scores as the dependent variable

First, pretest MR scores were centered at the cutoff score ($X = 18$) by subtracting observed scores from the cutoff. Next, using multiple OLS regression, a series of regression models were estimated with a dummy coded intervention condition variable (0 = no intervention, 1 = intervention) as the primary predictor variable, pre-test MR scores and a dummy coded gender variable (0 = females, 1 = males) as covariates, and posttest MR scores as the outcome variable. In addition, a dummy variable for gender (0 = female, 1 = male) and a term for the interaction between gender and intervention condition were included as predictors, in order to determine if the intervention impact on posttest mental rotation differed for males and females. However, the interaction term did not significantly predict students' MR posttest scores, $\beta = -0.02$, $t(671) = -.39$, $p > .05$. This result indicates that influence of the intervention on posttest mental rotation scores did not differ between the males and females in the sample.

Table 2

Regression model equations estimating impact of spatial skills intervention course on posttest MR scores and calculus course grades.

Model	Equation
2. Linear	$Y_{ij} = \beta_0 + \beta_1 Z_i + \beta_2 (X_{2i} - X_{2C}) + E_{ij}$
1. Linear with interaction	$Y_{ij} = \beta_0 + \beta_1 Z_i + \beta_2 (X_{2i} - X_{2C}) + \beta_3 Z_i (X_{2i} - X_{2C}) + E_{ij}$

Note. Each equation includes an estimate of the y-intercept, represented by β_0 ; estimates of the regression coefficients of the predictors, represented by β_{1-3} ; a dummy variable representing individual assignment to treatment and comparison conditions, represented by Z_i ; a calculation centering individual pre-test scores around the pre-test cutoff score (so that the y-intercepts for both groups are at the cutoff score, facilitating the detection of a regression discontinuity), represented by $(X_{2i} - X_{2C})$; and an estimate of random error, represented by E_{ij} .

¹ Since it is impossible to eliminate the intervention to evaluate this standard, a recommendation is to see if an alternative baseline measure (related to the dependent measures) is continuous (i.e., does not show a discontinuity) at the cutoff (Schochet et al., 2010). In the present study, the students' Math Subtest ACT scores, assessed prior to the intervention, were selected as the alternative baseline measure. Math ACT scores were significantly correlated with posttest MR scores, $r = .33$, $p < .001$ and with calculus grades, $r = .24$, $p < .001$. Therefore, this served as a reasonable alternative baseline measure. Next, Math ACT scores were regressed on the dummy coded intervention variable and the centered pre-test variable. Holding pre-test MR scores constant, the intervention condition variable did not significantly predict Math ACT scores, $b = -0.29$, $t(674) = -0.29$, $p > .05$, indicating no impact of the intervention on this baseline covariate. In other words, the relation between Math ACT scores and the posttest MR scores was continuous at the cutoff (i.e., there was no discontinuity).

The final model testing the intervention impact on posttest MR scores included intervention condition and centered MR pretest scores as predictors. Intervention course significantly predicted students' posttest MR scores when holding constant centered pretest MR scores, $\beta = 0.12$, $t(673) = 2.85$, $p < .01$. The effect size for this predictor was $r = .11$, which is equal to Cohen's $d = .22$, a small but meaningful effect size. In other words, after partitioning variance for the assignment variable, and thus completely accounting for selection effects, a significant impact of the spatial skills intervention course on posttest MR scores was observed. This was indicated by a regression discontinuity at the MR pretest cutoff score. Thus, as predicted, the intervention was effective in raising posttest mental rotation scores, significantly above scores that would have been predicted if no intervention had occurred. As would be expected, the centered pretest variable (although not the predictor of interest) also significantly predicted posttest MR scores, $\beta = 0.75$, $t(673) = 18.55$, $p < .001$. As can be seen in Fig. 2, which displays a bivariate scatterplot of centered pretest variable and MR posttest scores, the regression line is higher for the intervention group than the comparison group at the cutoff.

3.3.2. Calculus scores as the dependent variable

Next, a parallel analysis was conducted with students' calculus course grades as the dependent variable. The centered pretest MR scores were also used as a covariate in this analysis. Since doing calculus depends, in part, on spatial reasoning, participants were assigned to intervention groups based on this skill. This allowed us to examine whether the spatial skills intervention for low spatial students might help to improve these students' calculus grades. Further, by including pretest MR scores in the model, the impact of the intervention on calculus grades was calculated at the cutoff score, allowing for selection effects to be controlled. As with the analysis of MR posttest scores, the interaction term for gender and intervention course did not significantly predict calculus grades, $\beta = 0.12$, $t(483) = 1.43$, $p > .05$, indicating that the impact of the intervention did not differ for males and females. Consequently, gender and the interaction term were dropped from the model.

Subsequently, the final analysis of calculus grades included the intervention course variable and centered pretest variable as predictors. As expected, the intervention course variable was statistically significant, $\beta = .14$, $t(487) = 2.24$, $p < .05$. The effect size was small but meaningful at $r = .10$, which is equivalent to Cohen's $d = .20$. When holding constant the centered pretest MR scores. In other words, there was a regression discontinuity of calculus grades at the MR pretest cutoff score. Fig. 3 displays the regression discontinuity of calculus grades as a function of intervention condition and centered MR pretest scores. Thus, the analysis of calculus grades yielded findings parallel to those of the first analysis. The intervention had a positive impact on both posttest MR scores and calculus grades. Similar to the first analysis, the centered pretest variable was also a significant predictor of calculus grades, $\beta = .16$, $t(487) = 2.51$, $p < .05$.

4. Discussion

There is substantial evidence that spatial skills are a major component of the underlying abilities contributing to the development of expertise in science, technology, engineering, and mathematics (STEM). A recent longitudinal study following 400,000 high school students 11+ years later, investigated both their choice of college major and career, and found that adolescent spatial reasoning skills were predictive of choice of STEM majors and careers, above and beyond the effects of verbal and math abilities (Wai, Lubinski, & Benbow, 2009). Based on their findings and prior research, they concluded that for decades, spatial ability assessed during adolescence has surfaced as a salient psychological attribute among those adolescents who subsequently go on to achieve advanced educational credentials and occupations in STEM. Unfortunately, the importance of this skill as a major factor in predicting success in STEM fields has been ignored in favor of focusing primarily on math and verbal skills. Based on this longitudinal research, the importance of spatial reasoning as an alternate route to success in STEM is finally becoming clearer to policy makers at the national level, along with the recognition that level of spatial ability is critical in structuring students' educational and occupational choices and -outcomes—either

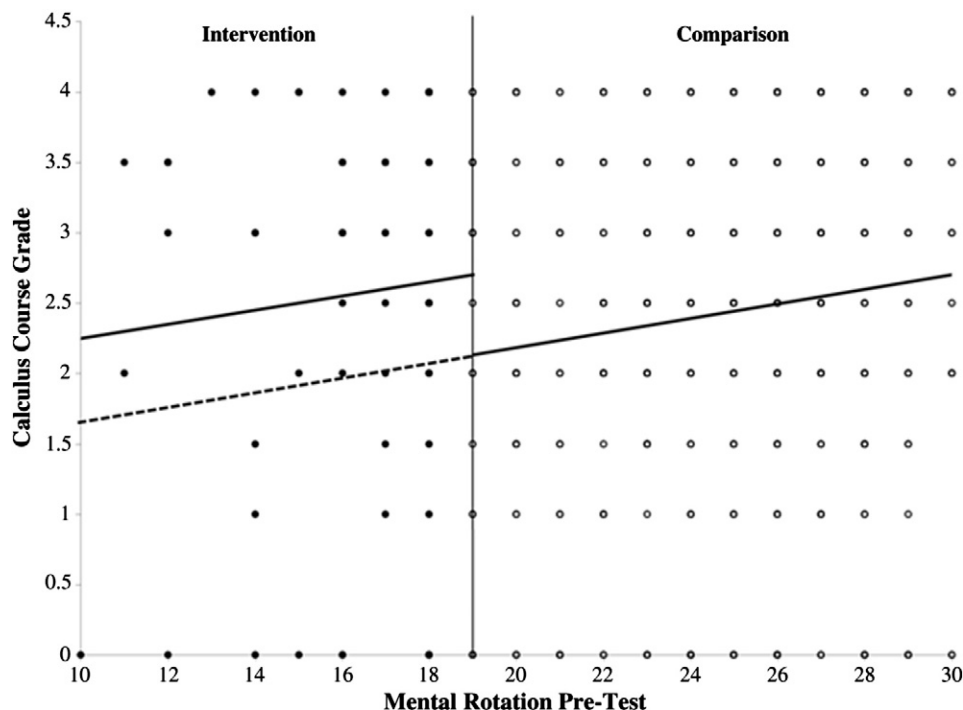


Fig. 3. Regression discontinuity of calculus grades as a function of centered mental rotation pretest assignment variable and intervention group.

towards or away from STEM fields (National Science Foundation & National Science Board, 2010).

The major contribution of the present results is to confirm and validate prior findings on spatial skills previously conducted at the same Engineering Program at Michigan Tech—by applying a more statistically sophisticated approach. One problem interpreting the prior results from this Program is that there might have been a selection bias, since students who failed the initial spatial test could decide to either take the intervention or serve as the comparison group. Thus, the students choosing to take the intervention may have had higher motivation levels than the students not taking it.

In the Fall of 2009, we obtained permission from the school to require all failing students to take the intervention course. We compared their progress to students who just passed the cutoff on the spatial task. By using this higher performing comparison group, we set the bar higher than just randomly assigning failing students to the two conditions. Nevertheless, we confirmed the findings of the previous research from the Michigan Tech Engineering Program—spatial training is effective in improving engineering students' spatial skills. Future research will need to specify which aspects of the intervention provide the most benefit. For example, the requirement to complete sketches of objects visualized in different orientations may have been more effective than the fill-in-the-blank or multiple-choice exercises. Even having to take this additional course, and receiving the extra attention that being in the class entails, could lead students to pay more attention, and could have contributed to students' improvement.

Another key contribution of the present findings was that the spatial training of the low spatial students was not limited to improving their spatial skills alone; transfer of learning effects were also shown through the significant impact of spatial training on students' success in a calculus math course. These transfer results are particularly encouraging when compared to a prior study, which did not find any benefit of a spatial training program on either spatial visualization or calculus performance (Ferrini-Mundy, 1987). This prior study provided a much less intensive spatial training program and was provided to college students across a range of majors. The present training was based on a spatial intervention program developed over 15 years, and designed explicitly to develop spatial reasoning skills in engineering students.

There are a number of possible mechanisms by which spatial training could lead to improvement in calculus performance. A lot of calculus problems involve visualizing slopes and areas under curves. Students must also understand inflection points and how they relate to maximum and minimum slopes. Frequently, students will be shown a 2-D graph on the board and asked to visualize it as a 3-D function. In physics, it's been found that experts make sketches to first visualize the problem, while novices, in contrast, focus solely on representing the problem as a set of equations (Taasobshirazi & Carr, 2008). Just having extra practice visualizing and sketching objects in 3-D space may have increased the intervention students' capacity to use visualization when solving calculus problems.

4.1. Implications for remediating spatial skills

The present findings suggest that students with initially poor spatial skills should be identified at the outset of their engineering programs and provided with remediation and spatial content should be included in first year engineering courses. It is likely that a subset of engineering students may need more time to develop and practice spatial skills, especially as these types of skills have not been identified or taught previously in the large majority of elementary and high school educational systems within the US.

These findings also have implications for other STEM fields as well. They suggest that researchers in these disciplines should explore the potential advantages of early identification of students with poorer spatial skills, and to follow up with spatial interventions to improve these skills. This would be advantageous for elementary and middle school

students as well; it has been found that having more advanced spatial skills is linked to better mathematics achievement (Casey, Nuttall, & Pezaris, 1997; Casey et al., 1995; Geary, Saults, Liu, & Hoard, 2000), and in a recent longitudinal study, it was found that spatial skills in kindergarten were stronger predictors of ninth grade math school achievement than fourth grade math achievement scores (Krajewski & Ennemoser, 2009).

4.2. Gender differences in spatial skills

Individual differences in spatial skills have implications for gender equity as well. Research has shown large gender differences favoring males on key types of spatial skills found in children as young as four years of age (Levine et al., 1999; Voyer et al., 1995). The effect sizes for gender differences in mental rotation skills in particular has been found to be close to 1 standard deviation, with males performing higher than females on average. This pattern of findings was confirmed in the present study, with large gender effect sizes favoring males found on the pretest and moderate effect sizes occurring on the post-test. It is clear that even engineering females come to engineering programs at the outset of their training with lower spatial abilities compared to males in engineering.

Nevertheless, both the low spatial performing males and females on the pretest showed significant improvement from the spatial intervention course at post-test. Thus, the intervention did not have differential effects as a function of gender, but instead, both low spatial ability gender groups benefited from the intensive spatial intervention. The fact that the gender differences were not eliminated through the intervention is consistent with prior training research. Although it has been shown that spatial skills are highly malleable for both males and females, a recent meta-analysis found that gender differences are not eliminated as a result of training (Uttal, 2009).

Most of the prior research on gender differences has involved relatively short mental rotation training sessions (Uttal, 2009). Thus, the present study contributes to the relatively small group of studies that has examined the effects of extended spatial training on male and female performance (Terlecki, Newcombe, & Little, 2008). In the present study, although 28% of the female engineering students failed the initial spatial pretest compared to 9% of the males, by the end of the year only 7% of the females and 2% of the males were still failing the spatial test.

One key point to be made is that gender differences in spatial skills do not need to be eliminated to achieve gender equity. In order for female engineering students to succeed at spatial problem solving in engineering, they simply need to be proficient at the types of spatial reasoning skills that are necessary for completing a range of engineering problems. It is also important to recognize that there is wide variability within genders as well as between genders, and there are substantial numbers of females with high spatial skills, especially those entering STEM fields.

5. Conclusions

Spatial skills are critical to success in most STEM fields, including engineering. However, there is little guarantee that students who graduate from high school in the US have developed these critical skills. The present study shows that spatial skills among engineering students are malleable, and that spatial interventions for students majoring in STEM fields can substantially improve the performance of students who initially showed poor spatial skills. Specifically, we found that spatial interventions are effective in raising spatial skills for the engineering students who failed the spatial test at the outset of their freshman year. It also was shown that improvements in spatial skills resulted in improved grades in an introductory calculus course. These findings could have far-reaching implications as we strive to increase the number of people who earn STEM degrees.

References

- ACT (2007). The ACT technical manual. Retrieved from <http://www.act.org/research/researchers/techmanuals.html>
- Baenninger, M., & Newcombe, N. (1989). The role of experience in spatial test performance: A meta-analysis. *Sex Roles*, 20(5–6), 327–344. <http://dx.doi.org/10.1007/BF00287729>.
- Bodner, G. M., & Guay, R. B. (1997). The Purdue visualization of rotations test. *The Chemical Educator*, 2(4). Available from: <http://chemeducator.org/bibs/0002004/00020138.htm>
- Bremigan, E. G. (2005). An analysis of diagram modification and construction in students' solutions to applied calculus problems. *Journal for Research in Mathematics Education*, 36(3), 248–277.
- Brinkmann, E. H. (1966). Programed instruction as a technique for improving spatial visualization. *Journal of Applied Psychology*, 50(2), 179–184. <http://dx.doi.org/10.1037/h0023068>.
- Cappelleri, J. C. (1991). *Cutoff-based designs in comparison and combination with randomized clinical trials*. Ph.D. Dissertation, Cornell University, Ithaca, NY.
- Casey, B. M. (2013). Spatial abilities and individual differences. In D.A. Waller, & L. Nadel (Eds.), *Handbook of spatial cognition* (pp. 117–134). Washington, DC: American Psychological Association.
- Casey, M. B., Nuttall, R. L., & Pezaris, E. (1997). Mediators of gender differences in mathematics college entrance test scores: A comparison of spatial skills with internalized beliefs and anxieties. *Developmental Psychology*, 33(4), 669–680. <http://dx.doi.org/10.1037/0012-1649.33.4.669>.
- Casey, M. B., Nuttall, R., Pezaris, E., & Benbow, C. P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology*, 31(4), 697–705. <http://dx.doi.org/10.1037/0012-1649.31.4.697>.
- Cook, T. (2008). Waiting for life to arrive: A history of the regression-discontinuity design in psychology, statistics, and economics. *Journal of Econometrics*, 142(2), 636–654. <http://dx.doi.org/10.1016/j.jeconom.2007.05.002>.
- Dhillon, A. S. (1998). Individual differences within problem-solving strategies used in physics. *Science Education*, 82(3), 379–405. [http://dx.doi.org/10.1002/\(SICI\)1098-237X\(199806\)82:3<379::AID-SCE5>3.0.CO;2-9](http://dx.doi.org/10.1002/(SICI)1098-237X(199806)82:3<379::AID-SCE5>3.0.CO;2-9).
- Ferrini-Mundy, J. (1987). Spatial training for calculus students: Sex differences in achievement and visualization ability. *Journal for Research in Mathematics Education*, 18, 126–140.
- Geary, D. C., Sauls, S. J., Liu, F., & Hoard, M. K. (2000). Sex differences in spatial cognition, computational fluency, and arithmetical reasoning. *Journal of Experimental Child Psychology. Special Issue: Sex and Gender Development*, 77(4), 337–353. <http://dx.doi.org/10.1006/jecp.2000.2594>.
- Gimmestad, B. J. (1989). Gender differences in spatial visualization and predictors of success in an engineering design course. *Proceedings of the national conference on women in mathematics and the sciences*, St. Cloud, MN (pp. 133–136).
- Guay, R. B. (1977). Purdue spatial visualization test: Rotations. West Lafayette, IN: Purdue Research Foundation.
- Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbache, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, 8(1), 1–51. <http://dx.doi.org/10.1111/j.1529-1006.2007.00032.x>.
- Hamlin, A. J., Boersma, N., & Sorby, S. A. (2006). Do spatial abilities impact the learning of 3-D solid modeling software? *Proceedings of the annual conference of ASEE*. Chicago, IL [CD-ROM].
- Hsi, S., Linn, M. C., & Bell, J. E. (1997). The role of spatial reasoning in engineering and the design of spatial instruction. *Journal of Engineering Education*, 86(2), 151–158.
- Johnson O'Connor Research Foundation (2005). Statistical bulletin 2005-4: Major field plots for the foundation's standard test battery. USA: Chris Condon and David Schroeder.
- Krajewski, K., & Ennemoser, M. (2009). The impact of domain-specific precursors and visuospatial skills in kindergarten on mathematical school achievement in Grade 9. *Paper presented at the biennial meeting of the Society for Research in Child Development*, Denver, CO.
- Larkin, J., McDermott, J., Simon, D. P., & Simon, H. A. (1980). Expert and novice performance in solving physics problems. *Science*, 208(4450), 1335–1342. <http://dx.doi.org/10.1126/science.208.4450.1335>.
- Levine, S. C., Huttenlocher, J., Taylor, A., & Langrock, A. (1999). Early sex differences in spatial skill. *Developmental Psychology*, 35(4), 940–949. <http://dx.doi.org/10.1037/0012-1649.35.4.940>.
- Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior*, 39(4), 990–997. <http://dx.doi.org/10.1007/s10508-008-9460-8>.
- Lord, T. R. (1985). Enhancing the visuo-spatial aptitude of students. *Journal of Research in Science Teaching*, 22(5), 395–405. <http://dx.doi.org/10.1002/tea.3660220503>.
- Maier, P. H. (1994). *Räumliches vorstellungsvermögen* Spatial imagination. New York: Lang.
- National Science Foundation, & National Science Board (2010). Preparing the next generation of STEM innovators: Identifying and developing our nation's human capital (NSB-10-33). Retrieved from <http://www.nsf.gov/nsb/stem/innovators.jsp>
- Nemeth, B. (2007). Measurement of the development of spatial ability by mental cutting test. *Annales Mathematicae et Informaticae*, 34, 123–128.
- Norman, K. L. (1994). Spatial visualization—A gateway to computer-based technology. *Journal of Special Education Technology*, 12(3), 195–206.
- Onyancha, R. M., Derov, M., & Kinsey, B. L. (2009). Improvements in spatial ability as a result of targeted training and computer-aided design software use: Analyses of object geometries and rotation types. *Journal of Engineering Education*, April, 2009, 157–167.
- Pribyl, J. R., & Bodner, G. M. (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching*, 24, 229–240.
- Prieto, G., & Velasco, A. D. (2010). Does spatial visualization ability improve after studying technical drawing? *Quality and Quantity: International Journal of Methodology*, 44(1015), 1024.
- Rubin, D. B. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*, 2(1), 1. <http://dx.doi.org/10.2307/1164933>.
- Schochet, P., Cook, T., Deke, J., Imbens, G., Lockwood, J. R., Porter, J., & Smith, J. (2010). Standards for regression discontinuity designs. Retrieved from What Works Clearinghouse website. http://ies.ed.gov/ncee/wwc/pdf/wwc_rd.pdf
- Shadish, W. R., Cook, T. D., & Campbell, D. T. (2001). Experimental and quasi-experimental designs for generalized causal inference. Boston: Houghton Mifflin.
- Shadish, W. R., Galindo, R., Wong, V. C., Steiner, P. M., & Cook, T. D. (2011). A randomized experiment comparing random and cutoff-based assignment. *Psychological Methods*, 16(2), 179–191. <http://dx.doi.org/10.1037/a0023345>.
- Small, M. Y., & Morton, M. E. (1983). Spatial visualization training improves performance in organic chemistry. *Journal of College Science Teaching*, 13(1983), 41–43.
- Smith, I. M. (1964). Spatial ability—Its educational and social significance. London: University of London.
- Sorby, S. (2001a). Improving the spatial skills of engineering students: Impact on graphics performance and retention. *Engineering Design Graphics Journal*, 65, 30–35.
- Sorby, S. (2001b). A course in spatial visualization and its impact on the retention of female engineering students. *Journal of Women and Minorities in Science and Engineering*, 7, 153–172.
- Sorby, S. (2005a). Assessment of a "new and improved" course for the development of 3-d spatial skills. *Engineering Design Graphics Journal*, 69, 6–13.
- Sorby, S. (2005b). Impact of changes in course methodologies on improving spatial skills. *International Journal for Geometry and Graphics*, 9, 99–105.
- Sorby, S. (2009). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education*, 31, 459–480.
- 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*, 89, 301–307.
- Sorby, S., & Wysocki, A. (2003). Introduction to 3d spatial visualization: An active approach. Clifton Park, NJ: Thomson/Delmar Learning.
- Taasobshirazi, G., & Carr, M. (2008). Gender differences in science: An expertise perspective. *Educational Psychology Review*, 20(2), 149–169. <http://dx.doi.org/10.1007/s10648-007-9067-y>.
- 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, 996–1013.
- Thistlethwaite, D. L., & Campbell, D. T. (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, 51(6), 309–317. <http://dx.doi.org/10.1037/h0044319>.
- Trochim, W. M. K., & Cappelleri, J. C. (1992). Cutoff assignment strategies for enhancing randomized clinical trials. *Controlled Clinical Trials*, 13(3), 190–212. [http://dx.doi.org/10.1016/0197-2456\(92\)90003-1](http://dx.doi.org/10.1016/0197-2456(92)90003-1).
- U.S. Department of Education, & Institute of Education Sciences (2008). Technical methods report: Statistical power for regression discontinuity designs in education evaluations (NCEE 2008-4026). Retrieved from <http://ies.ed.gov/ncee/pdf/20084026.pdf>
- Uttal, D. H. (2009). Training spatial skills: What works, for whom, and for how long? *Paper presented at the Society for Research in Child Development*, Denver, CO.
- 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(2), 250–270. <http://dx.doi.org/10.1037/0033-2909.117.2.250>.
- Wai, J., Lubinski, D., & Benbow, C. P. (2009). Spatial ability for STEM domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of Educational Psychology*, 101(4), 817–835. <http://dx.doi.org/10.1037/a0016127>.
- Waller, A. A., & Lourenco, S. (2010). Proceedings from 40th ASEE/IEEE frontiers in education conference: Special session—Frontiers of spatial cognition: Moving beyond rotating cubes. Washington, DC: ASEE/IEEE (American Society of Engineering Education/Institute of Electrical and Electronics Engineers).
- Zimmerman, W. (1991). Visual thinking in calculus. In W. Zimmerman, & S. Cunningham (Eds.), *Visualization in teaching and learning mathematics* (pp. 127–137). Washington, DC: Mathematical Association of America.