Learning and Individual Differences 26 (2013) 20–29

Contents lists available at SciVerse ScienceDirect

Learning and Individual Differences

journal homepage: www.elsevier.com/locate/lindif

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.

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 (Taasoobshirazi & 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" (Taasoobshirazi & 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).

1041-6080/$ – see front matter © 2013 Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.lindif.2013.03.010

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 consituted 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 pretest–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 spilt-half reliabilities of .78–.85, indicting good internal consistency. Criterion validity was shown by a correlation of .61 between the PVST: 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; Taasoobshirazi & 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,

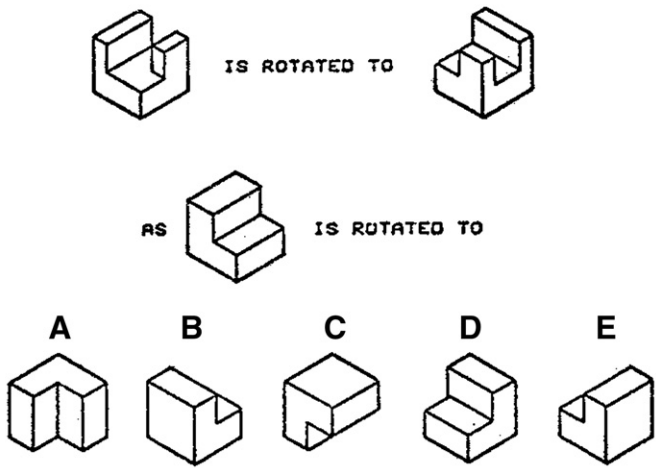

Fig. 1. Sample problem from the Purdue Spatial Visualization Test: Rotations (PSVT: R).

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, χ^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%), χ^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				
	n	Min	Max	*M*	*SD*	*n*	Min	Max	*M*	*SD*	*n*	Min	Max	*M*	*SD*
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. PVST-R pretest scores served as the assignment variable for intervention condition (cutoff = 18). PSVT-R posttest scores and calculus grades were dependent measures.

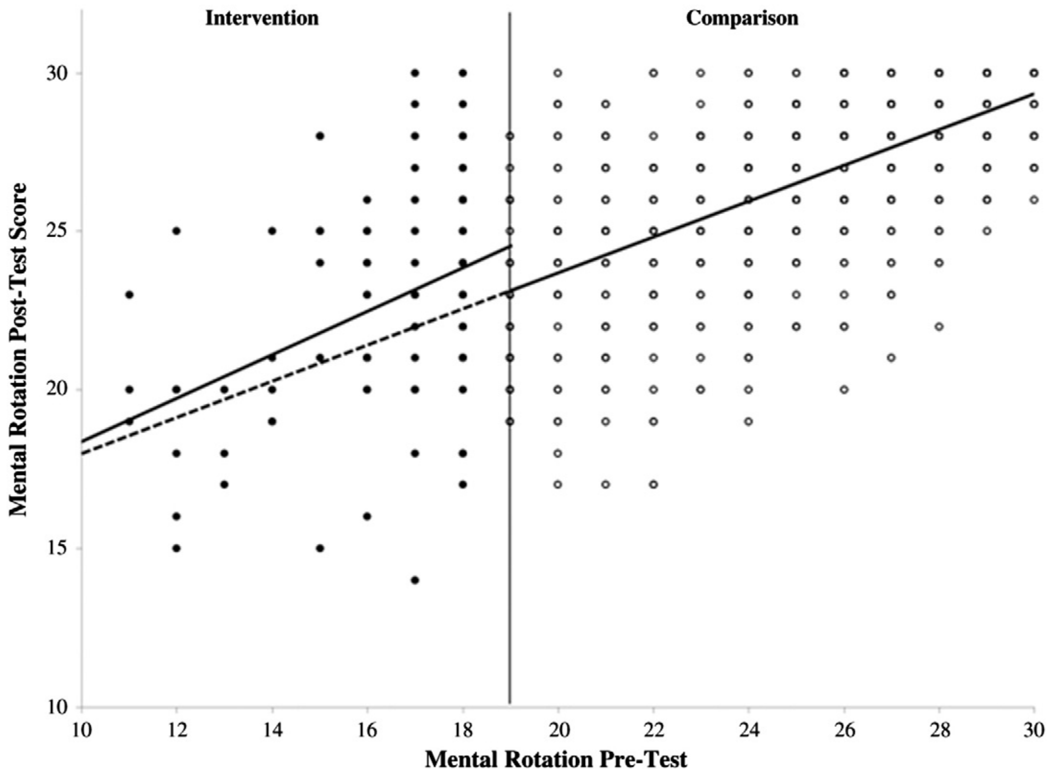

Fig. 2. Regression discontinuity of posttest mental rotation scores as a function of centered mental rotation pretest assignment variable and intervention group.

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	$\mathbf{Y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 Z_i + \hat{\beta}_2(X_{2i} - X_{2C}) + \mathbf{E}_{ij}$
1. Linear with interaction	$\mathbf{Y}_{ij} = \hat{\beta}_0 + \hat{\beta}_1 Z_i + \hat{\beta}_2(X_{2i} - X_{2C}) + \hat{\beta}_3 Z_i(X_{2i} - X_{2C}) + \mathbf{E}_{ij}$

Note. Each equation includes an estimate of the y-intercept, represented by $\hat{\beta}_0$; estimates of the regression coefficients of the predictors, represented by $\hat{\beta}_{1-3}$; 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 $\mathbf{E}_{ij}$.

[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, $p < 0.001$. 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

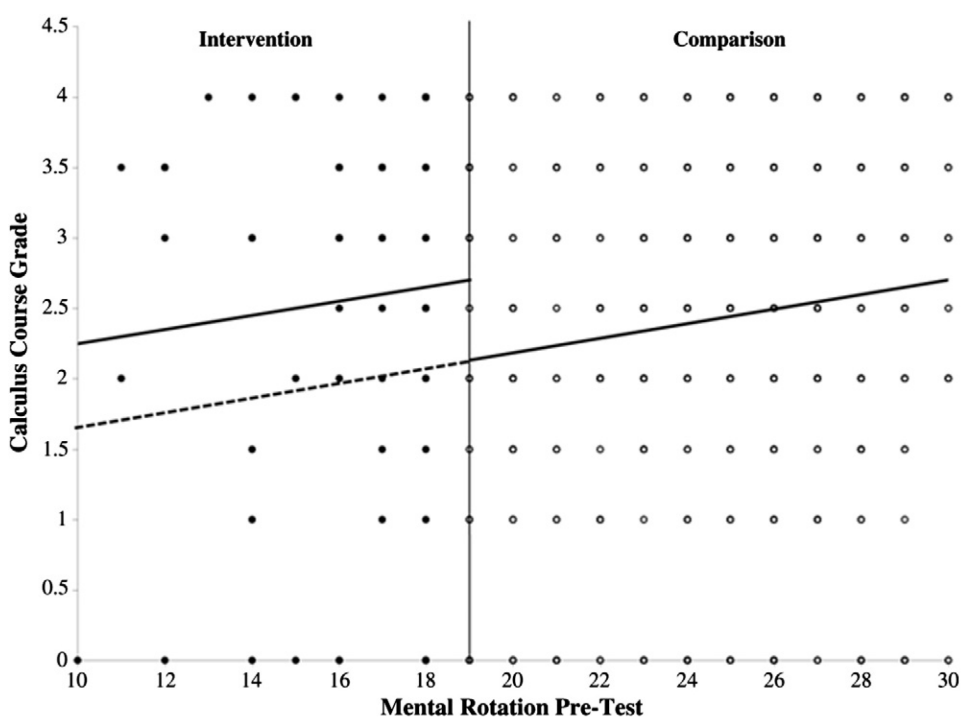

Fig. 3. Regression discontinuity of calculus grades as a function of centered mental rotation pretest assignment variable and intervention group.

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 (Taasoobshirazi & 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.htmhttp://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.

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., Saults, 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). Raeumliches vorstellungsvermoegenSpatial 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.

Taasoobshirazi, 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-I.

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.