

2016-01-03

Measuring the Effectiveness of Software-Based Training to Improve the Spatial Visualization Skills of Students in STEM Disciplines in Higher Education Institutions

Peter Cole
Technological University Dublin

Follow this and additional works at: <https://arrow.tudublin.ie/scschcomdis>



Part of the [Computer Engineering Commons](#), and the [Higher Education Commons](#)

Recommended Citation

Cole, P. (2016) *Measuring the Effectiveness of Software Based Training to Improve the Spatial Visualization Skills of Students in STEM Disciplines in Higher Education Institution*. Submitted in partial fulfilment of the requirements of Technological University Dublin for the degree of MSc. in Computing (Advanced Software Development) January 2016.

This Theses, Masters is brought to you for free and open access by the School of Computing at ARROW@TU Dublin. It has been accepted for inclusion in Dissertations by an authorized administrator of ARROW@TU Dublin. For more information, please contact yvonne.desmond@tudublin.ie, arrow.admin@tudublin.ie, brian.widdis@tudublin.ie.



This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 3.0 License](#)

DUBLIN INSTITUTE OF TECHNOLOGY



MSc COMPUTER SCIENCE

Advanced Software Development

DT228A

Measuring the Effectiveness of Software-Based Training to Improve the Spatial
Visualization Skills of Students in STEM Disciplines in Higher Education Institutions

Submitted By:

Peter Cole

Supervisor:

Damian Gordon

A dissertation submitted in partial fulfilment of the requirements of Dublin Institute of Technology for the degree of MSc. in Computing (Advanced Software Development)

January 2016

I certify that this dissertation which I now submit for examination for the award of MSc in Computing (Information Technology), is entirely my own work and has not been taken from the work of others save and to the extent that such work has been cited and acknowledged within the text of my work.

This dissertation was prepared according to the regulations for postgraduate study of the Dublin Institute of Technology and has not been submitted in whole or part for an award in any other Institute or University.

The work reported on in this dissertation conforms to the principles and requirements of the Institute's guidelines for ethics in research.

Signed: _____

Peter Cole

Date: 03 January 2016

Measuring the Effectiveness of Software-Based Training to Improve the Spatial Visualization Skills of Students in STEM Disciplines in Higher Education Institutions

Abstract

This research investigates how software can be used to teach spatial skills leading to greater success in Science, Technology, Engineering, and Mathematical (STEM) fields. Existing research indicates that spatial skills can be taught and that good spatial skills are common to people who succeed in STEM fields. In this work, a software-only testing system with a direct targeted, training intervention module was implemented to measure and teach spatial skills using mental rotations, which are believed to be one of the most significant indicators of success in STEM fields. Spatial skills were tested using a standardized and validated test that measures spatial visualization skills for rotation. Tests were administered at the beginning and end of the autumn semester 2015 using the online system under a Pretest, Posttest, and Comparison (PPC) model. Experimental and comparison groups were used to administer a spatial skills software-based intervention. The standardized test, the Revised Purdue Spatial Visualization Test for Rotations (PSVT:R) was used to measure students' spatial skills. Results showed higher gains for students who used the intervention software that were statistically significant.

Keywords: Spatial, Visualization, Mental Rotation, STEM, Software, Education, Training, Intervention, Pretest, Posttest

To my wife, my partner, my inspiration, my love.

ACKNOWLEDGEMENTS

Thank you to all the students at Rowan University who participated in this study.

Thank you to the professors and instructors at Rowan University who helped me complete my case study and this study, especially Mariano Savelski, Joe Stanzione, Michael Ciocco, Kevin Dahm, Jess Everett, Jenahvive Morgan, Kaitlin Mallouk, and Bruce Oestreich. You all helped in some way whether you knew it or not.

Thank you to all my instructors at DIT, Brenden Tierney, Deirdre Lawless, Luca Longo, Pierpaolo Dondio, John Gilligan, Aiden Meade, Ciaran Cawley, Sarah Jane Delany, John Burns, Robert Ross, Andrew Hines, and Brian Keegan. I learned from each of you. Thank you to the staff at DIT and all those who kept the wheels turning, especially Andrea Curley.

Thank you to my classmates and friends, for all the good times in the study halls, the pubs and the pool halls. Especially thanks to my drinking buddy Luke, and to Robert and Eamon, to Daniel, Oran, and Venkat, to Peter, Jack, and Liam, to Bastian and Abhi, to Fahad, Ahmed, and Hamad, and, of course, to France.

Thank you to Patricia, Brian, Doreen, Joshua, Laura and Ben, you made our time in Ireland a sunny affair.

Thank you Damian Gordan for your encouragement and your confidence in me, for your opinions and insights. A true force of positive thinking and a good guy!

Thank you Phoebe Q, I love you more than a poke in the eye with a sharp stick, and more than anything else for that matter. And thank you Suggy for everything, words fail me for how much I owe you; you are the best person I know.

And thank you *Boojum* ... for the 10 kilos.

TABLE OF CONTENTS

Abstract.....	iv
Chapter 1 Introduction.....	1
1.1 Background.....	1
1.2 Project Description.....	2
1.3 Research Aims and Objectives	2
1.4 Research Method	3
1.5 Scope and Limitations.....	3
1.6 Organization of Dissertation	4
Chapter 2 Literature Review.....	5
2.1 Introduction	5
2.2 Spatial Skills and Success in STEM	5
2.3 Types of Spatial Skills	5
2.4 Measuring Spatial Skills	7
2.4.1 Types of Spatial Skills Tests	7
2.4.2 Computer-Based Testing.....	10
2.5 Spatial Reasoning Strategies.....	10
2.6 Gender Differences in Spatial Skills.....	11
2.6.3 Spatial Skills Development & Prior Experiences.....	13
2.7 Spatial Skills Training.....	13
2.7.1 An Assessment of Software for Improving Spatial Visualization Skills	14
2.8 Conclusion	15
CHAPTER 3 DESIGN	18
3.1 Introduction.....	18
3.2 Pretest, Posttest, Comparison (PPC) Model	19
3.3 Psychological Testing	19
3.3.2 Psychometrics of Spatial Skills	19

3.3.3 Test Construction	26
3.3.4 Test Reliability	26
3.3.5 Test Validity	27
3.3.6 Test Question Characteristics	27
3.3.7 Intervention Learning Module.....	28
3.4 Human Subject Research	29
3.5 Methodology	30
3.5.1 System Implementation	30
3.5.2 Program Design	31
3.6 Conclusion	38
Chapter 4 Experiment	40
4.1 Introduction	40
4.2 Intervention Time Tracking	41
4.2.3 Filtering Algorithm.....	42
4.3 Amount of Intervention.....	44
4.4 Student Motivation.....	44
4.5 Conclusion	45
Chapter 5 Evaluation	46
5.1 Introduction.....	46
5.2 Roadmap	46
5.3 Data Analysis	48
5.3.4 Total Population	48
5.3.5 Filtered by Student Intervention Effort	50
5.3.6 Intervention Effort Segmented by Gender	52
5.4 Conclusion	61
Chapter 6 Conclusions and Future Work.....	63
6.1 Introduction.....	63

6.2 Findings.....	63
6.3 Research Objects.....	64
6.4 Limitations	65
6.5 Future Work	66
Bibliography	69
Chapter 7 Appendices.....	76
7.1 Institutional Review Board (IRB) – Spatial Skills Study Rowan University	76
7.2 Institutional Review Board (IRB) – Certification.....	78
7.3 Invitation to Participate in Intervention	80
7.4 Study Participation Agreement	81
7.5 Questionnaire Questions	81
7.5.1 Pretest Questionnaire.....	81
7.5.2 Posttest Questionnaire	84
7.6 Intervention Full Descriptive Statistics.....	85
7.7 Cronbach α Results Summary.....	86

TABLE OF FIGURES

Figure 1 From (Wai et al., 2009) Spatial Ability against Degree Groups	6
Figure 2 Sample question from the PSVT:R Rotations Test (From Bodner & Guay, 1997)	8
Figure 3 MRT Sample Items (From Vandenberg & Kuse, 1978)	9
Figure 4 A Question as Presented in the Paper Version of the Revised PSVT:R	21
Figure 5 A Question as Presented in the Online Version of the Revised PSVT:R	22
Figure 6 High Level System Architecture	31
Figure 7 Screen image of the Match Analogy Section of the Intervention Software	36
Figure 8 Screen image of the Analogy Challenge Section of the Intervention Software	37
Figure 9 High-Level Entity Relationship Diagram	39
Figure 10 Pseudo Code for Intervention Effort Filtering	43
Figure 11 Invitation to All Students to Participate in the Intervention	80
Figure 12 Study Participation Agreement Screen Capture from Website	81

TABLE OF TABLES

Table 1 Testing Presentation Latency	25
Table 2 Pretest & Posttest Administration Dates	41
Table 3 Effect of Filtering Intervention Timing Data	42
Table 4 Descriptive Intervention Statistics for the Different Population Subgroups	44
Table 5 Population Sizes for all Populations Analyzed	48
Table 6 Total Population T-Tests (EG) Experimental Group; (CG) Comparison Group	48
Table 7 (EG) Experimental Group & (CG) Comparison Group Filtered by Intervention Effort	52
Table 8 EG Filtered by Intervention Effort EG vs CG; Segmented by Gender	53
Table 9 EG & CG Filtered by Failure on Pretest	56
Table 10 EG vs CG; Males & Females	59
Table 11 Descriptive Intervention Statistics for the Different Population Subgroups	85
Table 12 Descriptive Intervention Statistics for Population Subgroups for Challenges by Level	86

TABLE OF PLOTS

Plot 1 Total Population EG Gain vs CG Gain	49
Plot 2 Cohen's d Effect Size for Total Population	50
Plot 3 Filtered EG Population vs CG Gain Pretest to Posttest	51
Plot 4 Female EG vs Female CG Pre & Post Scores	54
Plot 5 Male EG vs Male CG Pre & Post Scores	55
Plot 6 Failed Pretest EG Gain vs CG Gain	57
Plot 7 Intervention Effect Size for Student's that Failed the Pretest (≤ 18)	57
Plot 8 Effect Size; Females who Failed Pretest	61

Chapter 1 Introduction

1.1 Background

Spatial skills are a group of cognitive abilities that are recognized to be different from other kinds of intelligence such as verbal and mathematical skills (Gardner, 1984). They are mental abilities performed through an understanding of the dimensional structure of objects and visualizing relationships between them. They include mental tasks such as rotating three-dimensional objects, understanding the mirrored representation of objects, rotating one object to the same degree as another, as well as recognizing one's own physical relationship to other objects (Linn & Petersen, 1985).

Spatial skills' importance can be documented through decades of research showing correlations between spatial skills and professional achievement in Science, Technology, Engineering, and Mathematical (STEM) fields such as Computer Science (Wai, Lubinski, & Benbow, 2009). Wai *et al.*, (2009) state that, "...spatial ability has emerged as a salient psychological characteristic among young adolescents who go on to develop expertise in STEM domains." Furthermore, studies have shown correlations between spatial skills and academic entrance, academic retention and higher graduation rates in STEM fields such as computer science and lead to success in future endeavors in those fields (Newcombe, 2010).

Fortunately, studies have also shown that spatial skills are learnable and that students who are taught spatial skills have improved academic success in STEM fields (S. Sorby, Casey, Veurink, & Dulaney, 2013). This is true even for students who initially have low spatial skills (Uttal *et al.*, 2013). Uttal *et al.* explain that students with low spatial skills could ultimately do well in STEM subjects, "if these students could just get through the early phases of learning that appear to be particularly dependent on decontextualized spatial skills, then their lack of spatial skills might become less important as semantic knowledge increased." Sorby (2009) has demonstrated from a course she designed on three-dimensional spatial visualization skill, that "long-term assessment results ... have shown that engineering students who initially failed the Purdue Spatial Visualization Test for Rotations (PSVT:R) and who subsequently enrolled in the spatial skills course went on to perform better in their engineering graphics courses by a significant margin." The PSVT:R is a psychometric test that is frequently used to measure mental rotation ability.

Furthermore, scientific fields have been shown to have a significant importance in the economic success and security of developed nations. Unfortunately, there is a workforce shortage of professionals for these fields. Since women and minorities have long been known to be underrepresented in STEM disciplines, it is of significant interest to understand what factors limit or prevent them from succeeding in STEM fields (Burke & Mattis, 2007). Female's spatial skills have been recognized as being below that of males. However it has been demonstrated that females can improve their spatial skills with training (Connor, Schackman, & Serbin, 1978). If the discrepancy between males and females can be addressed through teaching spatial skills, it could help increase the number of professionals in STEM fields.

1.2 Project Description

Many research projects have been conducted that attempt to teach spatial skills to students. Some of the techniques that are commonly used to teach spatial skills include the use of workbooks, drawing and sketching exercises, manipulative objects, such as Snap Cubes®, and computer learning software. Despite its potential, software-based training has not always been shown to add significant value when used as an intervention in teaching spatial skills, especially compared to other interventions such as classroom lessons, workbooks, and practicing drawing three-dimensional objects (S. A. Sorby, 2006). This raises several questions. Is software intervention for teaching spatial skills effective? Has existing software used for spatial skills training simply not had the proper features? Before addressing these questions, it must be determined what teaching strategies have been used to improve spatial skills at the higher education level and which are most effective. Additionally, what features have been used in computer software to teach these strategies and how can they be more effective?

1.3 Research Aims and Objectives

The aim of this project was to evaluate the effectiveness of software only-based training to improve the spatial skills of students in STEM disciplines in higher education institutions. This was done by creating a software-based spatial skills training intervention, and administering pre- and post-intervention standardized tests that measured outcomes from the change in spatial skills scores. This project contributes to the body of knowledge about whether software-based interventions can be used to help teach spatial skills to students by demonstrating an increase in student spatial skills through the applied model. Specifically,

what is the effect of a short, targeted, online software intervention on mental rotation ability of first year STEM students? One objective for this study was to segment the collected data and analyze it in several ways, including by total population, by gender and by those students who started with low spatial skills, in an attempt to draw insight from those segments and how they may differ from each other. Other objectives for this project included designing an experiment to test students' spatial skills using the PSVT:R standardized test, creating an online spatial rotations testing system, developing a software-based intervention teaching module, and creating a software-based infrastructure to collect, administer, and evaluate data related to student testing, performance and improvement. Methods and approaches for teaching spatial skills were considered in order to understand what methods have been applied, especially using computer software, and measures were taken for improvement. An experiment was designed that used a pretesting and posttesting model with experimental and comparison groups. Students in the experimental group used the software-based intervention to improve their spatial skills. This involved creating an educational software-based lesson that taught mental rotations through practice and challenge modules.

1.4 Research Method

A pretest-posttest experimental research method with comparison groups was used on a large cohort of students to evaluate the effect of using a software-only teaching module as an intervention for increasing the spatial abilities of first year engineering students. Statistical methods employed to analyze the results included Welch two sample and paired t-tests, two-way Anova, and Cohen's d effect size estimates.

1.5 Scope and Limitations

The experiment was limited to first year engineering students at a single University. All first year engineers over 18 years of age were eligible for inclusion in the study. All first year engineering students in the engineering college were required to participate in the pretest; however, participation in the intervention and posttest was at the discretion of individual instructors since the intervention and posttest impinged on class time. There were 16 sections of a mandatory freshman clinic course that were invited to participate in the study with individual instructors determining at their discretion whether their section would opt out entirely, join the experiment as a comparison group whereby they would just complete the posttest without any intervention, or join the experiment for the intervention as well as the posttest. The amount of intervention given to students were not uniform. Students were

encouraged to use the intervention software as much as they desired. Some instructors assigned a certain amount of intervention work as required work external to the class (i.e. homework / continuous assessment). Additionally, some instructors used a portion of class time (typically one class session) for students to use the intervention software. The amount of time and work spent on the intervention was tracked for each student. Due to time constraints a pilot study for the experiment could not be performed.

1.6 Organization of Dissertation

Presented in this dissertation is a literature review related to spatial skills testing, a design and methodology description of the experiment performed, an explanation of the experiment, an analysis of the experimental results, and the conclusions reached as well as recommendations for directions of future research.

In the literature review seven areas of research are targeted that reflect the key areas of concern for this study. Topics such as the recognized types of spatial skills, measuring spatial skills through standardized testing and using computer-based testing, solution strategies for solving spatial problems, spatial skills development, gender differences in spatial skills, spatial skills and success in STEM, and spatial skills training are examined.

Following the literature review, the design discusses the background, approach and methodology and covers topics such as the pretest / posttest / comparison (PPC) model, psychological testing, test construction, reliability, and validity, the characteristics of test questions, the intervention learning module and human subject research. Further, the methodology, system implementation and program design are discussed.

Details of how the experiment was conducted including metrics, details of problems encountered, and motivations are discussed. Finally a statistical analysis is performed and conclusions and future direction are discussed.

Chapter 2 Literature Review

2.1 Introduction

In this chapter seven areas of research are targeted that reflect the key areas of concern for this study. Spatial skills as they relate to success in STEM fields is discussed and lays the groundwork for the justification of this work. Literature on the types of spatial skills has been examined as is discussed to provide a framework of where mental rotations fit in the scheme of spatial skills. Several psychometric instruments for testing spatial skills are discussed, focusing on the PSVT:R; their limitations and advantages are considered, followed by a discussion on using computer-based testing for spatial skills. Solution strategies for solving spatial problems is discussed. Spatial skills training is discussed including a detailed look at an assessment of an extant software system for improving spatial visualization skills. Finally a brief discussion on gender differences in spatial skills examines work that has tried to understand the gender differences that have been found in spatial abilities.

2.2 Spatial Skills and Success in STEM

A longitudinal study which examined an accumulation of over fifty years of psychological knowledge concludes that spatial ability in pursuit of education especially in STEM fields solidifies “the generalization that spatial ability plays a critical role in developing expertise in STEM [fields] (Wai et al., 2009).” They examined results from project TALENT which was a longitudinal study of over 400,000 high school student in the US conducted since 1960. Furthermore, they concluded that there is a pool of individuals with high spatial abilities, but lower math and verbal skills, who could succeed in STEM fields, but that pool remains untapped. (Figure 1)

Humphreys, Lubinski, & Yao, (1993) found that because traditional measures of intelligence such as the SAT, ACT and GRE [U.S. academic entrance exams] do not test for spatial abilities, disciplines such as engineering and the physical sciences may be losing talented candidates who would otherwise succeed in those fields. Lubinski (2010) states that spatial ability is “totally neglected” as a criteria for higher education selection despite that it is one of the mostly important abilities for excellence in STEM fields.

2.3 Types of Spatial Skills

Broadly speaking, spatial abilities are one’s aptitude for capturing images and mentally manipulating them through transformations or rotations. In 1979 McGee identified Spatial Visualization and Spatial Orientation as the two main types of spatial skills (as cited in Tartre,

SPATIAL ABILITY FOR STEM

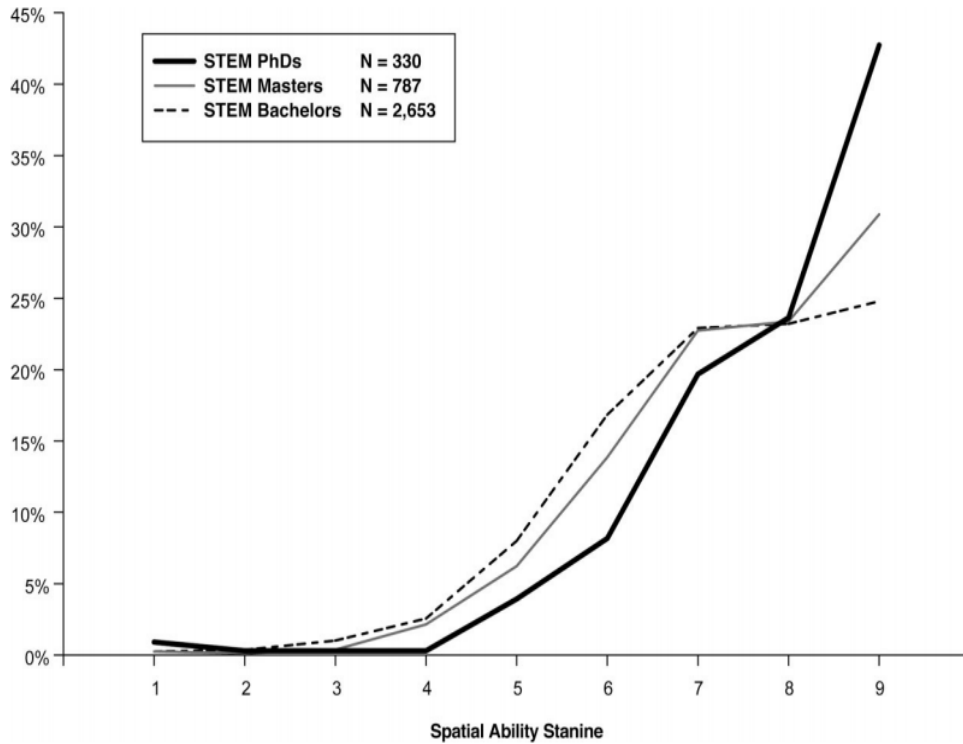


Figure 1 From (Wai et al., 2009) Spatial Ability against Degree Groups

1990). Tartre (1990) further breaks spatial visualization into the categories of mental rotation and mental transformation. Mental rotation is concerned with rotating an object in its entirety without transforming the size or shape of the object. Mental transformation deals with manipulations that transform a part of an object.

Some research makes a distinction between mental rotations performed on two-dimensional objects versus those performed on three-dimensional objects while others do not (Ho, Eastman, & Catrambone, 2006). Ho *et al.* (2006) found a correlation between two-dimensional and three-dimensional tasks, that is hierarchical, meaning those who can solve three-dimensional problems have the spatial skill required to solve two-dimensional rotation problems, but the converse is not true. They noted that the study was performed on static images on paper or computer screen and recommended that real-time, three-dimensional representations of objects be used for further study. In the current research, the intervention uses real-time three-dimensional object representations.

Linn & Petersen (1985) categorize the main spatial skills into three groups: Spatial Perception, Mental Rotation, and Spatial Visualization. Spatial Perception tasks challenge a person to recognize and understand spatial relationships in relation to their own body position even when distractions are present. However, they do not require one to mentally move an object. Mental Rotation tasks challenge a person to quickly and accurately rotate mental images. Spatial Visualization tasks challenge a person to perform more complex operations, often with multiple steps, that may have multiple solution strategies, which may include mental rotation or spatial perception as a component step and require an analytical strategy to solve. (Linn & Petersen, 1985) demonstrated that mental rotation tasks are separate from visualization tasks.

Many courses designed to improve spatial skills for STEM students focus specifically on the development of mental rotation ability because it is most strongly correlated with academic success in STEM fields (Sorby, 2009b). For that reason, this research will also focus on the spatial skill of mental rotation.

2.4 Measuring Spatial Skills

2.4.1 Types of Spatial Skills Tests

There are many psychometric instruments to measure various types of spatial skills. Some of the more common are the Purdue Spatial Visualization Test: Rotations (PSVT:R) (Bodner & Guay, 1997), the Mental Cutting Test (MCT) first developed in 1939 by CEEB, the College Entrance Examination Board and the Mental Rotation Test (MRT) by Vandenberg & Kuse (1978). Kozhevnikov (2012) has identified over 50 tests used to assess spatial skills. These spatial tests can be used as a source to design interventions for teaching spatial skills.

In the Revised PSVT:R which consists of 13 symmetrical and 17 nonsymmetrical 3-D objects students are shown a drawing of a three-dimensional reference object and an accompanying version that has been rotated (Maeda, Yoon, Kim-Kang, & Imbrie, 2013) (Yoon, 2011). Another three-dimensional object is shown and the student must identify how that object would look if it were rotated in the same way as the original reference object. A typical question on the PSVT:R test is a multiple choice question that depicts one or more three-dimensional objects and asks the student to somehow perform a mental rotation of the picture, for example by rotating it a specific way, and then select the correct matching result from an array of choices. All the objects are shown as two-dimensional isometric drawings (Figure 2). This test requires a higher level of spatial visualization ability than other spatial tests by

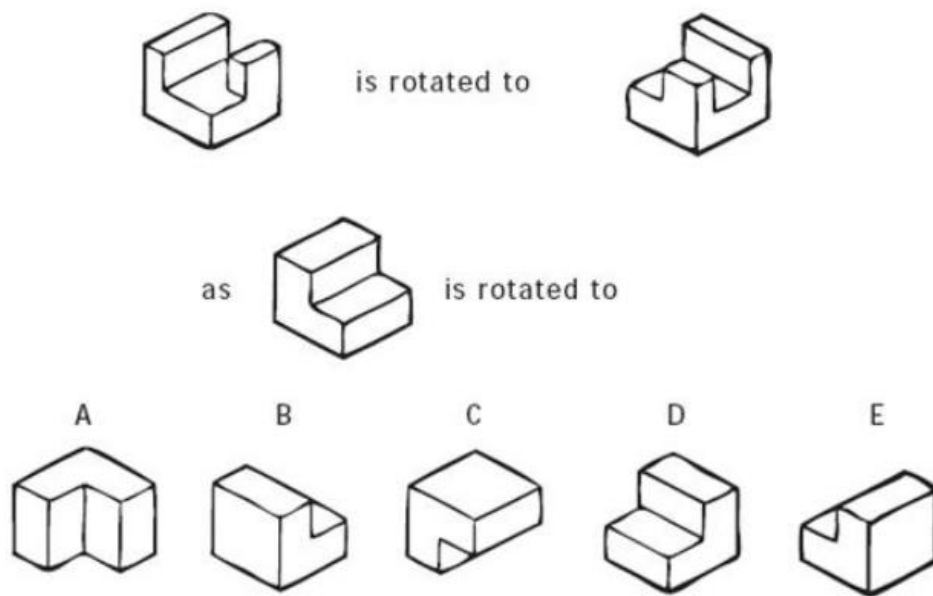


Figure 2 Sample question from the PSVT:R Rotations Test (From Bodner & Guay, 1997)

maximizing Gestalt-type processing and minimizing analytical processing. Gestalt processing is a holistic approach to solving spatial tasks where objects are considered in their entirety rather than broken down into their individual components. (Bodner & Guay, 1997).

According to Maeda & Yoon (2013) the PSVT:R is frequently used as the preferred instrument to measure mental rotation abilities for several reasons that they synthesized from various resources. They report that according to Bodner & Guay (1997) its questions are less likely to be solved by analytical thinking and more likely to use holistic or gestalt thinking. It has been demonstrated that there is a positive relationship between solving tasks in STEM disciplines and solving rotations in the PSVT:R (Yue, 2004). Strong testing reliability and test validity evidence has been reported that support the use of the PSVT:R (Maeda & Yoon, 2013). The item difficulty of the questions in the PSVT:R are high enough that they allow researchers to distinguish STEM student's mental rotation ability (Yue, 2006) (Black, 2005).

The PSVT:R uses inclined, oblique, and curved surfaces which require a greater spatial ability to mentally rotate than other simpler shapes, such as cubes, used in other rotation tests (Yue, 2004).

Several limitations have been noted in the PSVT:R such as the fact that how the test is administered can affect the results, for example, whether by paper or by computer, the types

of images, isometric or trimetric, used in the test, the length of time given to complete the test and whether the test has been truncated (Maeda & Yoon, 2013).

Yue (2006) (2009) found a number of figural errors on questions from the original PSVT:R from Bodner and Guay (1997). There were 10 problems such as missing lines found on 7 questions. A revised test was created by Yoon (2011); using Classic Test Theory and Item Response Theory she subsequently validated the test as a psychometrically sound instrument. In the current research the Revised PSVT:R is used for pretesting and posttesting.

Yue (2009) conducted a study in which the traditional orthogonal, line-based isometric images used in the PSVT:R test were replaced with “*more realistic*” pictorial representations created on CAD software that allowed for shading and better surface representation. Each version of the test was given to students and the scores were compared to determine if the performance was improved when more realistic images were used. While the mean scores showed improvements of up to 15% the differences were not found to be statistically significant.

The Mental Cutting Test (MCT) presents a three-dimensional image to the student who must mentally slice the image with a plane (S. A. Sorby, 2006). Five choices of possible resultant cross-sections are presented to the student who must identify the correct match.

The Mental Rotation Test (MRT) presents a three-dimensional image to the student with four additional images that may be rotated versions of the original (Vandenberg & Kuse, 1978). Students must indicate which two of the four additional images are matches for the original image. (Figure 3)

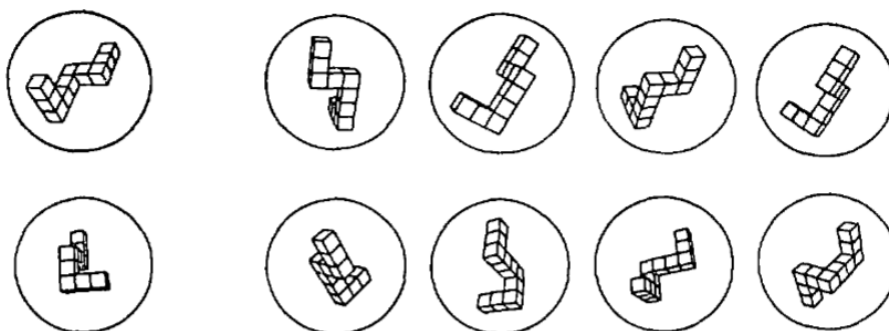


FIG. 1. Sample items from the Mental Rotations Test. The first and fourth alternatives in the first item are correct; the distractors are mirror images of the criterion figure. In the second item, the second and third alternatives are correct; the distractors are rotated images of other criterion figures.

Figure 3 MRT Sample Items (From Vandenberg & Kuse, 1978)

2.4.2 Computer-Based Testing

Veurink & Hamlin (2015) explored differences between testing spatial skills in engineering students using online methods compared to paper-based tests. They found that students who took the spatial test using an online Learning Management System on average earned lower scores. It was true for the total population and for males; however, females had no statistically significant difference between the paper and online versions of the test.

Pedrosa, Barbero, & Miguel (2014) created a web-based “interactive learning manager for graphic engineering to develop spatial vision,” called ILMAGE_SV. The aim of the study was to see if the tool was able to teach spatial visualization. It compared a traditional classroom-based teaching course against the ILMAGE_SV system which covered the same material. The MRT test was used as a pretest and posttest. A Differential Aptitude Tests-Space Relations (DAT-SR) test was used to assign students to spatial visualization aptitude levels for training. After training a visualization test made up of questions selected from the DAT-SR, the MRT, PSVT:R & the Lappan test was administered online. Pedrosa *et al.* report that results from students in the experimental group were similar to those taught in a traditional classroom. The ILMAGE_SV system was reported to be more efficient for students challenged with low spatial skills and for student without technical drawing experience. The system is written in Spanish and only works using Microsoft’s Internet Explorer and requires a separate third-party, ActiveX control, called eDrawing from SolidWorks, Inc., to be installed on the user’s computer.¹ Only a small portion of the website could be evaluated due to technical difficulties getting Internet Explorer to recognize the eDrawing software.

2.5 Spatial Reasoning Strategies

There are two categories of spatial reasoning that are given attention in the literature on mental rotations: Gestalt-based and analytical. Gestalt-based reasoning is holistic in its approach to rotating an object, meaning that the act of mentally rotating an object rotates the entire object all at the same time. Analytical approaches do not take a holistic approach, instead there are many approaches that can be taken but none involve rotating the object as a whole. Analytic rotation can be pattern-based, can involve deconstructing the object mentally, can use other analysis methods, such as the color of the object, or some other identifying characteristic, or

¹ <http://www.edrawingsviewer.com/>

they can even be verbal based (Harris, Hirsh-Pasek, & Newcombe, 2013). Individual strategies used to perform mental rotations may vary from student to student and even may vary from problem to problem.

Bethell-Fox and Shepard (1988) showed that the amount of time it took to perform mental rotations on unfamiliar patterns (two-dimensional 3 x 3 grids with patterns of filled squares) increases as the patterns become more complex. Additionally, they showed that when students are given appropriate practice, rotation time was independent of the pattern's complexity.

(Harris et al., 2013) examined mental rotation skills compared to mental folding skills and found that, "robust sex-related differences," existed for mental rotation skills but not for mental folding skills. Mental rotation skills skill have been shown to have a gender-based gap in several studies. One reason frequently suspected for the discrepancy between genders is that females are analytical rotators and males are holistic or Gestalt-like rotators; this conclusion is often made on evidence that females are solving mental rotation problems more slowly than males (Geiser, Lehmann, & Eid, 2006). However, Geiser *et al.*, (2006) could not be definitive on the point; they felt that possibly females were using holistic rotation in a slower way, that they were using other analytic strategies, or perhaps some combination of the two. Linn & Petersen, (1985) have made similar conclusions.

Rather than just looking at Gestalt vs. Analytical strategies as it applies to the object being rotated, Just & Carpenter, (1985) examined the use of alternative coordinate systems that are applied to the demands of a given task. Instead of using a single, invariant coordinate system, for example the Cartesian system, they demonstrate that a different coordinate system can be used as necessary to meet the requirements of different problems. They describe each required rotation around an axis in the selected coordinate system as a trajectory. This type of rotation has a shorter trajectory, with a rotation around a single axis. A step-by-step rotation around several axis is said to have a longer trajectory. Which trajectory method a person chooses will have implications for what must be achieved mentally to perform the required rotation. Further they found that subjects with low spatial skills never chose shorter trajectories and that those with higher spatial skill usually did use them. (Just & Carpenter, 1985)

2.6 Gender Differences in Spatial Skills

Halpern (1986) indicates, "*There is considerable evidence that spatial ability can be improved with training, and that under normal circumstances, boys receive more spatial training.*" Halpern expresses that the reason males outperform females is due to social pressures for

females to avoid professions that are considered to be meant more for men and that society offers males more opportunities for spatial skills training. However, not all research agrees on the degree of spatial capabilities based on sex. Caplan *et al.* (1985) have expressed that it is premature to conclude that there exists a superiority in males over females in spatial abilities. They cite problems with the way the individual studies that support differences based on sex are conducted indicating that some tests, such as the rod-and-frame test, that have been used in other studies are not true measures of spatial abilities as students' scores may be affected by other factors such as assertiveness, fear, or uneasiness. Additionally, they indicate that results to studies supporting the claim may be overgeneralized; when told that the test used was a test of empathy, females performed better. They felt that ultimately it is, "*premature and indeed inappropriate to ask about sex differences in spatial ability.*" Alexander and Evardone (2008) demonstrated that females performed better on MCT tests when the test stimuli, i.e. block shapes, were replaced with human figures. They feel the reason males perform better on the MCT test may be that using some shapes may give an advantage to males in a way that is unrelated to their spatial abilities.

Parsons (2004) reproduced the difference in spatial abilities in males and females when testing was performed using MCT tests; however, in a Virtual Reality Spatial Rotation (VRSR) system they developed, predicted sex differences were not found. He surmises that the difference could be due to a change in the cognitive domain used when using the VRSR system.

Kail, Carter, & Pellegrino (1979) tested males and females in a mental rotation experiment that used the Spatial Relations Test from the Primary Mental Abilities (PMA) battery. They found that males and females performed similarly in most regards including: using mental rotation to judge pairs of objects, accuracy in judging pairs, precision (when compared with latency), and the amount of time used when comparing objects both familiar and unfamiliar. Males only outperformed females in the speed with which they rotated objects, suggesting that possibly females and males use different strategies when performing mental rotations, i.e. females use a slower, but equally effective analytical strategy, while males use a faster holistic strategy.

Voyer (2010) performed a meta-analysis of studies that measured mental rotation abilities, using tests such as the MRT and the PSVT:R. They found that whether or not the tests were administered with time limits made a significant difference in the scores between genders.

Males were favored over females when time constraints existed. They found that this effect was present as long as there was a time limit, even when they were “long.” Additionally, in a meta-analysis study, Voyer, Voyer, & Bryden, (1995) found that the MRT mental rotation test produced the more significant discrepancies between genders than any other spatial ability test. They were emphatic that there should be no further discussion as to whether or not there exists any discrepancy between genders in spatial abilities.

Linn & Petersen (1985) established that in timed, mental rotation tests, females performed slower and more poorly than males. They conclude that the reason for the differences between males and females in spatial abilities has not been determined and offer several possible explanations including that females may simply perform mental rotations more slowly than males, that females apply less efficient strategies when solving mental rotation problems, they may use analytical rather than holistic approaches, or they may be more inclined to use testing time to go back and double check their work, or perform rotations more than once, perhaps using analytical methods.

2.6.3 Spatial Skills Development & Prior Experiences

Deno (1995) demonstrated that for men non-academic activities had a stronger relationship to visual spatial skills than formal academic subjects. However, there was no statistically significant indication of the same for women. This was true across a wide variety of activities for men, further Deno states, “*the sole significant positively correlated high school activity for women was playing computer video games.*”

2.7 Spatial Skills Training

In Uttal *et al.* (2013) a meta-analysis of 217 research studies that investigated training for spatial skills was performed. They determined that three different types of mutually exclusive spatial training categories could be used to categorize the approaches taken. They included training via video games, training via a semester-long course of a relevant subject matter covering spatial abilities, e.g. a graphics drawing course, and training through practice, strategic instruction, or software lessons. These were further classified as either indirect or direct training. Video game training and related course training were classified as indirect training. Practice, strategic instruction and software-based lessons were classified as direct training. Sixty-seven percent of the studies analyzed used a direct training approach which is also the approach proposed here. The research addresses studies that were conducted from 1984 through the 2000s that used some type of intervention to improve spatial performance.

Connor, *et al.* (1978) demonstrate that with training the performance of girls on spatial aptitude tests “*improved significantly*” from pretest to posttest in which the Children's Embedded Figures Test (CEFT) was used. Further in cross cultural testing, the author found, “*in general, it appears that sex differences on tasks with large spatial components are more likely to be found in cultures having rigidly defined sex roles.*” (Connor *et al.*, 1978)

Hsi, Linn, & Bell (1997) trained over 500 students in spatial skills over a four year period and demonstrated that spatial ability was a predictor of success in an introductory engineering course. They taught various approaches for solving spatial skills that seemed to increase students' ability and confidence, especially for females. They recommend teaching spatial strategies as part of the curriculum for introductory engineering courses.

2.7.1 An Assessment of Software for Improving Spatial Visualization Skills

Gerson *et al.* (2001) have evaluated and assessed the development of a multimedia software system for improving three-dimensional spatial skills. Their study was based on work done by Sorby, Baartmans, and Wysocki who received NSF funding to develop the multimedia software and a separate workbook for a course that was previously developed to improve spatial skills. The earlier course developed by Sorby *et al.* used a computer-aided design software package, called I-DEAS, that was originally developed for the automotive industry and was not designed specifically for spatial skills learning. The level of hardware required to use this software made it impractical for widespread use. Sorby *et al.* performed a quasi-experimental quantitative study using newly developed multimedia software across two years. Four modules from the replacement software and workbook were used in a new course in 1998. The course was comprised of 35 students who were strongly encouraged to enroll after failing a spatial rotation pretest with a score below 60%. The students evaluated the new multimedia software as being more user-friendly; however, they also felt they learned more from the I-DEAS software even though it was complicated. The I-DEAS software package contained only exercises while the multimedia software had both background information and exercises.

In the second year, fall 1999, students who were required to take spatial training were given a choice between two versions of the course. Both versions used the multimedia software that had since had another five modules added to it, for a total of nine. In the first version, students also had a weekly two-hour lecture. In the second version, the lecture was replaced with a workbook in a two-hour supervised lab with no formal instruction. Eleven students took version 1, with lectures, and nine took the software / workbook version. Pretests and posttests

were administered to each module to measure gains. For most spatial skills tested, the pretest and posttest scores were similar between the two modules; however, for mental cutting tests (MCTs) and Purdue Spatial Visualization Test: Rotations (PSVT:R) there was a statistically significant gain for both groups. Both modules showed similar, significant gains in spatial skills by the measures taken.

There were two types of quizzes administered to the modules, each tested different types of spatial abilities. In the first quiz, which tested making isometric sketches and orthographic projections both module versions scored similarly. In the second quiz, which tested pattern folding and rotation abilities, the non-lecture module outperformed the module with lectures. The authors attribute this difference to differences in the number and type of exercises that students completed during the sessions. The non-lecture module completed workbook pages with more difficult problems and completed more of them, thereby receiving more practice in what they were learning. Additionally, students in the lecture class had some additional topics in their syllabus so their time spent on spatial training was shortened.

Some qualitative evaluation data was collected from students during the study as a final evaluation of the software. Student responses that were most common were of the form, “the software was easy to use and to understand,” “a combination of software and written materials is best for learning the course material,” and “the exercises on the multimedia software were generally too easy.”

The authors concluded that multimedia software is a useful method for improving spatial skills; however, each module had other factors effecting outcomes. In version one, there were a significant number of additional lectures given in conjunction with the software. In version two of the module, the software was used in conjunction with a workbook that provided additional samples and examples. The authors also strongly felt that multimedia software could not fully replace interaction between students and faculty. Significant gains were made across both modules indicating their value to spatial learning. The study did not isolate the software portion of each module, which could have conclusively demonstrated that the multimedia software was responsible for the measured gains. It was not stated whether the qualitative evaluations were significant in the stated conclusions.

2.8 Conclusion

Spatial skills have been studied for over a century. A broad search of published research articles relevant to spatial skill and its measurement were examined for this study. Spatial skills are

believed to be a unique type of mental processing skill different from other skills such as mathematical skills or language-based skills. Some research has shown that spatial skills significantly improve for females when given spatial skills training. Several studies that use pretest, posttest research models, as well as, pretest, posttest, and comparison models were examined. There is a good body of work that is similar to this study in the use of the PSVT:R instrument as a measure of spatial skill. One study, Pedrosa *et al.*, (2014) was discussed that performs a similar software-based only intervention mechanism as used in this study. Additionally, no studies were uncovered that used a custom developed testing and training system using the Revised PSVT:R to administer tests and train students specifically in mental rotation spatial skills. The PSVT:R rotation test requires a higher level of spatial visualization ability than other spatial tests by maximizing Gestalt-type processing and minimizing analytical processing. Many of the courses designed to improve spatial skills focus specifically on the development of mental rotation ability because it is most strongly correlated with academic success in STEM. It has been shown that the amount of time it takes for a student to perform mental rotations can be improved through practice irrespective of the complexity of the object. In one longitudinal study, software was used as a part of a spatial skills intervention, but was not tested as an isolated variable. In one study students felt the questions given on multimedia software were generally too easy and researchers strongly felt that multimedia software could not fully replace interaction between students and faculty.

The above discussion, covers several topics that are significant to the research conducted in this study. They all relate directly to the research objective to evaluate the effectiveness of software only-based training to improve the spatial skills of students in STEM disciplines in higher education institutions. That spatial skills matter for success in STEM fields is a fundamental premise for the value of this research. Understanding the types of spatial skills and the psychometric instruments used to assess them provide necessary context as does an understanding of prior work concerning computer-based testing and training for spatial skills. Understanding spatial solution strategies offers insight into the approaches taken when designing and implementing the intervention software as is used in this study. It is common knowledge among educators that females are underrepresented in many STEM fields today. Gender as it relates to spatial skills is relevant to this study because there is a well-documented but not well-understood significant difference in the performance between males and females on many spatial skills tests. Females are at a disproportionate disadvantage in STEM fields compared to males. How the genders perform using the intervention in this study may

contribute to the understanding of how females and males differ in response to using intervention software.

CHAPTER 3 DESIGN

3.1 Introduction

In this project a spatial skills testing and teaching module will be designed and implemented. It was designed to be used without additional pedagogical implements, such as classroom lessons or separate workbooks or other human interventions. The effectiveness of the software was measured by giving a standardized spatial skills rotation test before and after making the teaching module intervention available to students. The experiment was designed to space out the administration of the pretest and posttest results to minimize the “learning factor,” the effect of the pretest experience on the posttest scores. An attempt was made to minimize the delays in between the intervention and the posttesting; the amount of time between training and retesting was tracked in a database. In Branoff (1998) a “significant learning factor” was reported when students were given a modified PSVT:R test within 1 to 5 minutes of the unmodified version of the test. Additionally, the amount of time that students spent using the teaching software and the number of tasks completed in the software were tracked.

This project was implemented through the creation of a complete software-based system. First year engineering STEM students were tested using Yoon’s (2011) Revised PSVT:R test. The PSVT:R has been used extensively for spatial skills testing for many decades. The literature review has not revealed the existence of any available extant fully online management and testing system for administering the PSVT:R test. All of the software developed for this research was web-based and designed specifically to teach spatial skills. This is a difference from other spatial skills interventions software, such as that by Yue (2009) and Yue & Chen (2001), that were not designed specifically to teach mental rotations but were used to that end, for example several spatial skills computer-based training efforts used available CAD programs in their studies. In other research, such as by Sorby (2009) traditional desktop (as opposed to web-enabled) modules have been created and were used in conjunction with other intervention methods such as classroom-based lessons using paper workbooks. All of the software developed for this research was designed with a focus on holistic, Gestalt-type processing. Often in the spatial skills domain Gestalt processing is used to describe the holistic approach that can be taken when understanding an object. Bodner & Guay (1997) use the example of how we understand and recognize another person’s face, we do so holistically, not by knowing the specifics such as how far apart the eyes are from each other, the size of the lips, etc. Another difference to the design approach of this research is that it has also focused on having students practice and understand rotation in a more holistic manner. That is by

having students directly rotating objects around a single axis (using the mouse) rather than by using the more analytic, step-by-step approach, of breaking the rotation into pieces around the more familiar Cartesian axes. For example, when students complete a challenge question in the intervention they are showed an animation of how to correctly rotate the object in the problem, however this is done by showing the student how to accomplish the rotation using a single axis of rotation rather than through a series of steps around several (Cartesian) axes.

3.2 Pretest, Posttest, Comparison (PPC) Model

This research followed a Pretest/Posttest/Comparison (PPC) design. In PPC designed studies subjects are either in a comparison group or a group receiving an intervention. Participants are measured before and after the intervention which helps limit the effect of preexisting differences between participants (Morris, 2008).

When studying an experimental intervention using a PPC design, a distinction is made between a comparison and a control group. A control group is a group that receives no intervention at all whereas a comparison group may receive an alternative intervention. Creating a pure control group in an educational research setting, using human subjects, can be problematic. For example, students may discuss the relevant intervention topics between themselves or receive other training that could have an impact on the prescribed training or lack of intervention training. These factors would likely occur outside the researcher's control. For this reason, the term, "comparison group," is preferred over "control group."

Experiments that measure spatial skills have been conducted by S. A. Sorby & Baartmans (2000) and many others (Branoff, 1998) (Yue & Chen, 2001) (Yue, 2002) (Yue, 2009) have used similar pretest and posttest frameworks to evaluate improvement in students' spatial skill scores.

3.3 Psychological Testing

3.3.2 *Psychometrics of Spatial Skills*

One area of the field of psychometrics, which is a field that focuses on psychological measurements, is objectively measuring skills through an understanding of how testing instruments should be constructed and what procedures should be established (Kline, 1986). The revised paper PSVT:R instrument, which has been psychometrically validated by Yoon (2011), was used as a model for the testing instrument used in this research. The online version of the Revised PSVT:R used in this study has not yet been psychometrically validated. There

is precedent for variations of the PSVT:R to be used in research before they have been psychometrically validated (Maeda & Yoon, 2013).

There are several differences that may be of significance between the online version used in this study and the paper version psychometrically validated by Yoon. In the online version there is a timer presented on the user's screen as the test is taken. The amount of time that the student has remaining to complete the test is shown in a countdown fashion in minutes and seconds. Obviously this is not a feature of the paper version of the test. (Although it is presumed that students are somehow aware of the time remaining to complete the paper test.) Having a live timer on the screen raises several questions that should be answered when the online version is psychometrically validated. Does having a proximate timer change the way students manage their time when taking the test? From one point of view, students may find it easier to manage their time as they take the test by being made more aware of the remaining time, yet from another point of view the countdown timer may increase a student's feelings of pressure by having a continuous reminder of the impending time limit. Another difference related to time management has to do with how students move from question to question if they want to review their work. With a paper test a student can move between questions by estimating the number of pages to go back and jumping directly to (or close to) the correct page, however, on the online version the student must click back through each question sequentially. Will this induce students to spend less time checking their work and possibly receive lower scores using the online version of the test? Another difference between the two versions is that the online version will warn a student when they indicate, through a button click, that they are finished with the test if there are unanswered questions. Also, if using the online version, if the user runs out of time, the test ends immediately. Students who accidentally leave a question unanswered on the paper test may not realize their error and submit the test without even guessing at the correct answer.

In the instructions for the paper test students are told not to write directly on the paper test. Frequently answers have been collected through the use of "Scantron answer sheets" that can be processed electronically. These paper answer sheets may be easy for a student to visually scan and determine if questions have been left unanswered. As mentioned above, students are not supposed to write on the paper version of the test, although some still do so (N. Veurink & Hamlin, 2015). Students may, for example shade in parts of a shape to aid them in determining the rotation. In none of the literature reviewed for this study was it addressed whether or how policing of this policy occurs. Likewise, there is no mechanism in the online version for

students to “write” on the shape as an aid to determining the rotation. In either version of the test students are free to use scrap paper at their discretion.

What may be the most significant difference between the two versions is that the online version presents the questions to students in a random order. In the revised paper version questions are asked progressively from least to most difficult depending on the rotation of the shape and the number of axes about which it rotates Yoon (2011). The computer labs that were used for testing were comprised of long tables in ordered rows with several desktop computers with upright flat screen monitors. Consequently, the layout made it quite easy for one student to see the monitor of a fellow test taker. As a preventative measure to deter copying the questions in the inline version of the test were presented in a random order. Yoon (2011) does not explain the rationale for presenting the questions ordered to be progressively more difficult. A final difference to note is that the online version gives the student instant feedback on their scoring when the test is submitted raising a question as to whether there is a psychological impact to the student that might affect their motivation, either positively or negatively, to engage in the intervention. Students who score high may not see the value in spending time on the intervention or be motivated to improve. Those who score low may be motivated to try to

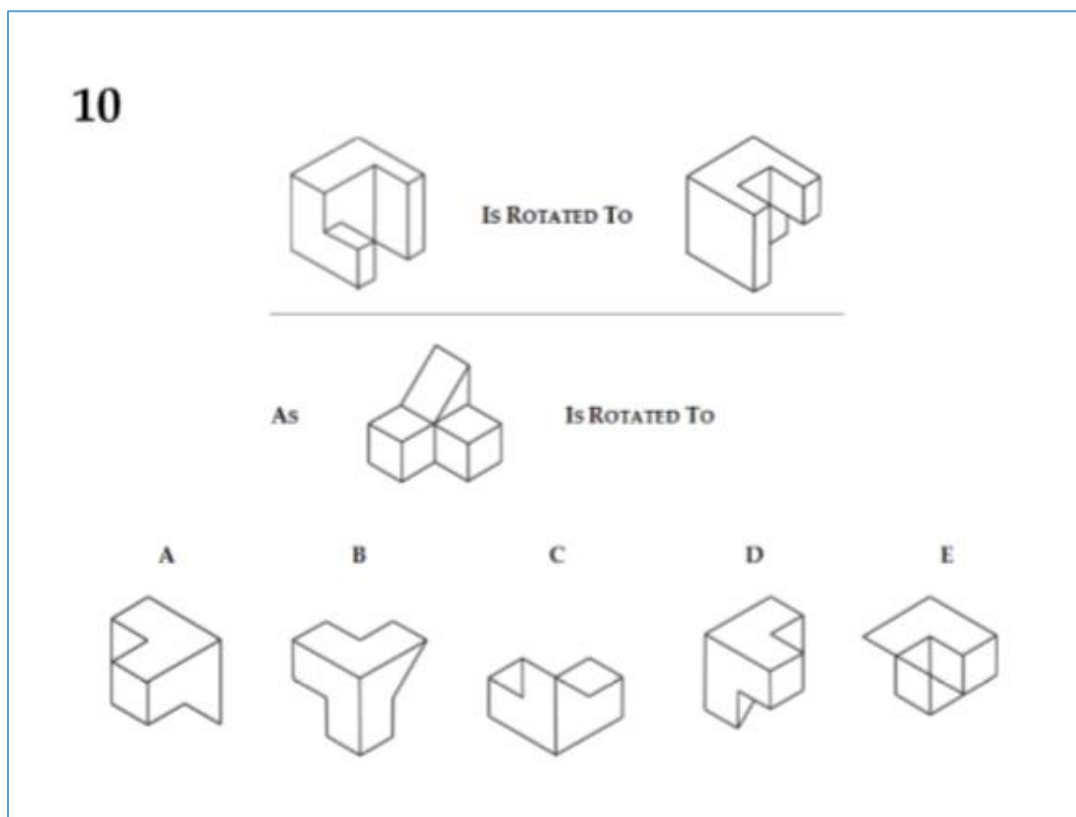


Figure 4 A Question as Presented in the Paper Version of the Revised PSVT:R

improve or may be discouraged and become disinterested in the intervention. These are all issues that can be addressed further when the online test is validated.

There are many similarities between the online version and the paper version of the Revised PSVT:R. (Figure 4)(Figure 5) The two introductory example questions are the same in both versions. The 30 questions presented to the students are exactly the same. The images used in the online version are digital replicates of the paper version. All images were captured directly from an electronic pdf version of the revised paper PSVT:R and have the same exact isometric properties. In both versions of the test students are presented with one question at a time. On the paper test there is only one question per page. Veurink & Hamlin (2015) have created an online version of the PSVT:R using a Learning Management System (LMS) which used an uploaded version of the paper version of the Revised PSVT:R. (Yoon, 2015) The student receives no feedback from the online version during the test. (As mentioned above the score is reported to the user on completion of the test.) The online version was administered with a twenty minute time limit as prescribed by Yoon (2011).

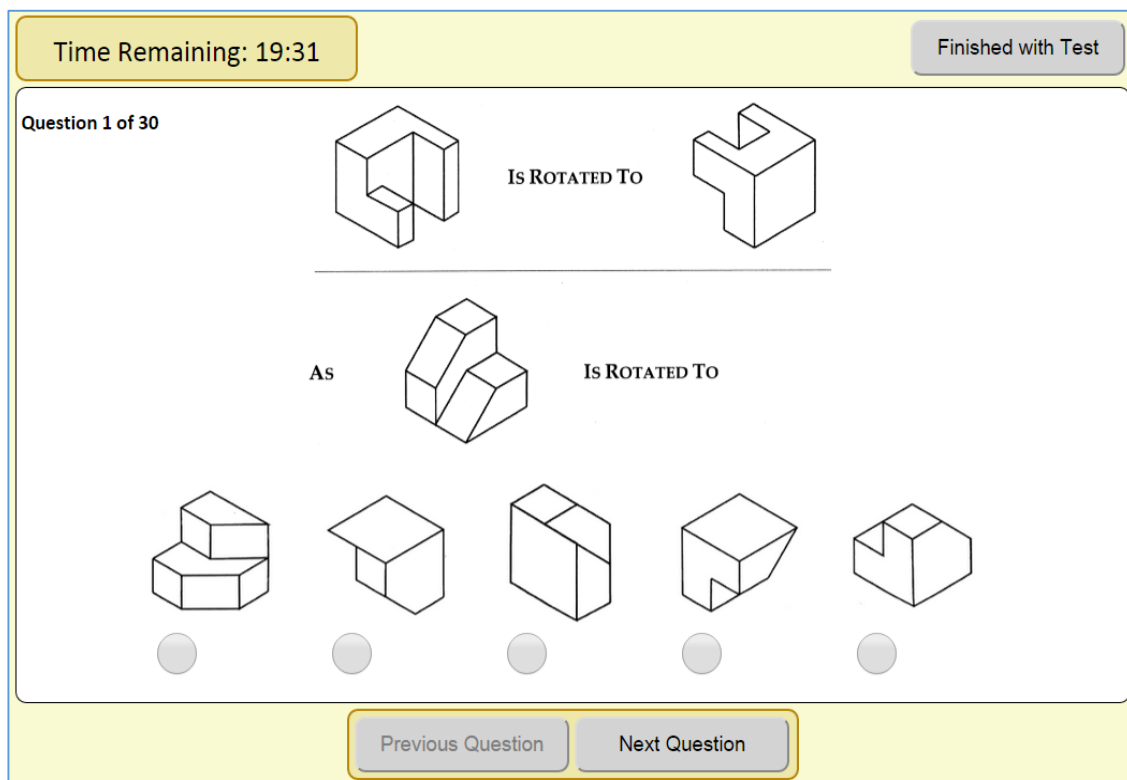


Figure 5 A Question as Presented in the Online Version of the Revised PSVT:R

In an effort to stay as close to the paper version as possible, care was taken to minimize the differences when implementing the online version of the PSVT:R, however, the introduction of differences were inevitable simply by changing the mode of the test. Changing the mode from paper-and-pencil based to computer-based created many subtle differences that should be addressed in further research when the test is validated.

Mode effect is a widely studied phenomenon wherein a survey or test yields different data depending on the administration mode of the instrument used. For example, a survey taken over the telephone may yield different results than the same survey administered face-to-face for reasons such as social or psychological pressures (Bridgeman, Lennon, & Jackenthal, 2003). Generally mode effect differences can be categorized as either participant issues or technological issues (Leeson, 2006). Participant-based issues are those that rise from differences in how the test taker participates in taking the test under the new mode. For example, how they react to using a computer, some people, especially the uninitiated, experience computer anxiety. The comfort level with the new technology, or any psychological or cognitive effects that working in the new mode may cause are other examples of participant-based issues leading to the mode effect (Leeson, 2006). Technology-based issues have to do with how the difference in the technology of the new mode changes how the user takes the test. For example, when taking a test on a computer the user needs to be comfortable using the devices, such as a mouse or keyboard, and they must understand how to use the specific software that presents the test, i.e. clicking on controls, windows, etc.

There are several areas of concern regarding mode effect that were considered when moving the paper version of the Revised PSVT:R to an online version. Bridgeman *et al.* (2003) found that when moving testing instruments to an online mode a low resolution presentation could cause the instrument to require scrolling by the user. They reported that test participants found scrolling to be an interference with taking the test. In tests that required little scrolling, there were no significant differences recorded in scores. The online version of the Revised PSVT:R created for this study required no scrolling.

The equipment used for giving the pretest and posttest at Rowan University in this study was identical for all users. No student was put at a disadvantage compared to another based on the configuration of the equipment in the computer lab used for testing. (Students were not assigned to a specific computer in the lab, instead they selected their computer through their

own choice on a first-come, first-served basis.) All computers were set to have the same screen resolution. Because the test was presented in a standard HTML 5 browser, users had the option to zoom in or out to a level that they were comfortable viewing the questions and to resize the browser window to their own preferred size. Even when used on a screen with resolution as low as 1024 x 768 pixels no scrolling was required.

It is important to recognize that the tested cohort, incoming first-year engineers, was already computer savvy. In the questionnaire given after the posttest, for the question, “How often do you use a computer browser in your daily life?” 251 students answered “frequently”, 18 answered “occasionally” and only 1 answered “seldom.” For the question, “How would you describe your comfort level at using a browser (Chrome, Firefox, Safari, etc.) to complete tasks on a computer?” 230 answered “strong”, 37 answered “moderate”, and 5 answered “weak.”

Another area of concern when putting a test online concerns presentation latency. Presentation latency is the delay that might occur when a user moves between questions on a test (Bridgeman et al., 2003). For example, when a user moves from question number one to number two, the presentation latency is the amount of time that the user must wait after completing question number one but before being able to work on question two. This latency time can be considerable for web-based systems in which requests for the next question must be sent to a remote server. One concern of a high presentation latency is that it might deter users from going back to check their earlier work and thereby not scoring as well as they would on the paper version of a test. In this study the online test was implemented as a web-based test; however, presentation latency was addressed in the implementation. Presentation latency was nearly eliminated by preloading the entire test into browser memory before the test commences, thereby eliminating the potential for time consuming messaging to a remote server during the test. The amount of latency was minimal as reported below. (Table 1) By tracking the amount of time spent on each question as well as the amount of time spent on the entire test it was possible to calculate the amount of presentation latency. In this study, there were 636 online PSVT:R tests administered on 24 different days for a total of 19,080 questions. The total amount of latency across all tests was 251.6 seconds with an average latency of 13.19 milliseconds per question.

Date Taken	Test Type Description	Number of Total Time on		Total Time on Questions (seconds)	Delta (seconds)	Average Latency per Question (milliseconds)
		Tests Taken	Tests (seconds)			
2015-09-01	PRE	142	130,801.89	130,599.25	202.648	47.57
2015-09-02	PRE	127	117,634.76	117,628.90	5.856	1.54
2015-09-04	PRE	62	50,912.53	50,909.37	3.165	1.70
2015-11-04	PRE	2	1,894.08	1,893.96	0.12	2.00
2015-11-06	PRE	1	442.25	442.13	0.12	4.00
2015-11-10	PRE	8	4,857.19	4,856.74	0.454	1.89
2015-11-17	PRE	1	945.44	945.39	0.053	1.77
2015-11-18	PRE	1	905.34	905.27	0.07	2.33
2015-11-25	POST	1	275.40	275.28	0.113	3.77
2015-11-26	POST	1	1,036.01	1,035.85	0.158	5.27
2015-11-27	POST	2	1,417.68	1,416.58	1.097	18.28
2015-11-28	POST	4	2,820.97	2,820.20	0.769	6.41
2015-11-29	POST	3	3,578.10	3,577.61	0.497	5.52
2015-11-30	POST	33	25,025.15	25,012.21	12.946	13.08
2015-12-01	POST	24	16,713.68	16,712.21	1.469	2.04
2015-12-02	POST	68	53,373.74	53,368.57	5.168	2.53
2015-12-03	POST	23	16,527.55	16,524.80	2.757	4.00
2015-12-04	POST	21	14,887.98	14,884.99	2.989	4.74
2015-12-04	PRE	1	814.20	814.04	0.157	5.23
2015-12-05	POST	1	300.79	300.72	0.066	2.20
2015-12-06	POST	2	1,713.59	1,713.40	0.193	3.22
2015-12-07	POST	3	2,236.66	2,236.31	0.342	3.80
2015-12-08	POST	102	72,885.01	72,874.91	10.108	3.30
2015-12-11	POST	3	2,487.65	2,487.37	0.281	3.12
Totals		636	524,487.64	524,236.05	251.60	

Table 1 Testing Presentation Latency

The online version of the test does not add significant new usability features or require new user understanding of how to answer questions that are significantly different from the paper test. The most significant difference in how to answer a question is that in the online version a user selects their answer by clicking on a radio button. In the paper version a user must write their answer down on paper, possibly by filling in a bubble on a Scantron sheet.

Results in this study were not compared to results that were collected from paper-and-pencil-based tests. Ultimately the goal of this study is not to validate the online version of the Revised PSVT:R as a testing instrument but to measure the effect of the intervention on spatial skills.

3.3.3 Test Construction

When constructing a valid measurable test, the goal is to create an instrument that can reliably represent the population on a scale that appropriately reflects the population's ability on the test. A certain amount of error is assumed to exist in the raw score achieved by a test taker. The raw score is said to consist of a true score plus some amount of error. Classical Test Theory is applied to construct a test that minimizes the amount of error contained in the raw scores of test takers, thereby giving a raw score that more correctly reflects the true score (Kline, 1986).

When Yoon (2011) revised and validated the PSVT:R she performed an item analysis in which she determined each item's difficulty that is defined as the proportion of the population that answered the question correctly. Item difficulty is important in test construction in order to create a scale with real meaning. If the degree of difficulty is low, or too easy, everyone in the population does well, if too hard, no one can answer correctly, in either case the resulting scale is unable to meet the desire to measure the difference between people. Ideally a sample population will score over a normal range (Kline, 1986).

The PSVT:R has been implemented for many spatial skills studies in ways that deviate from the specifics that were developed by Guay and implemented by Yoon. Examples of differences include offering the students different time lengths (Maeda & Yoon, 2013), supplying additional visual information on the shapes, such as Cartesian axes (Branoff, 1998), modeling the shapes with more realistic shading (Yue, 2009), or as trimetric images rather than isometric images (Maeda & Yoon, 2013), or as computer-based rather than paper-based tests.

3.3.4 Test Reliability

The reliability of a test is defined as how consistent the test instrument is in measuring that which it is designed to measure. Testing errors may come from tests that are not well designed or from test that are not well administered. For example, if the test instructions are not followed or if the test is administered without adhering to the allotted time limit.

A Cronbach α calculation was performed on the pretest results for the online version of the Revised PSVT:R indicating the online version of the test has good internal consistency (Cronbach's $\alpha = 0.86$, $N = 285$) (Appendix 7.7). The test was performed using the psych library, "Procedures for Psychological, Psychometric, and Personality Research", version 1.5.8, for the R language. When Yoon (2011) validated the Revised PSVT:R, she reported a Cronbach α of 0.862 with $N=1022$.

3.3.5 Test Validity

The validity of a test is concerned with how well a test measures that for which it was designed to measure. Often a test may require other skills from a test taker than that for which the test was designed. For example, a math test consisting of word problems requires a certain reading comprehension level as well as mathematical ability. If the test is intended strictly to measure mathematical ability, a requirement for reading comprehension may skew the results from the intention of the test. By implementing the PSVT:R test with more realistic views Yue (2009) attempted to demonstrate that the test measures not just the test takers ability to mentally rotate the shapes but also to understand two-dimensional representations of three-dimensional objects.

The purpose of this study is not to validate the psychometrics of the online PSVT:R test, rather to use the test as a repeatable instrument to measure the difference between students' pre-intervention and post-intervention spatial skills.

3.3.6 Test Question Characteristics

Yoon (2011) characterized questions on the PSVT:R as each having one of four different types of rotation where the rotation type loosely relates to the question or item difficulty. Item difficulty is computed as the number of examinees who answered correctly divided by the number of examinees taking the test. It assumes the items are scored dichotomously, without partial credit for an answer, meaning a correct response can be scored as a 1 and an incorrect response with a 0. Yoon defines the types of rotation for a problem based on the complexity of the rotation, including the number of Cartesian axes that the object is rotated around, the number of degrees (either 90° or 180°) rotated, and the direction of the rotation, either positive or negative. The actual configuration of the shape itself is not considered in this characterization, however Yoon suggests that the complexity of the rotation is not the only factor and that together with the complexity of the object's shape, may more closely define the degree of difficulty.

Further, Yoon (2011) looked at item discrimination when validating the Revised PSVT:R. Item discrimination is the ability of a test question to discriminate between high and low levels of ability for a test taker. It correlates the scores received on a question against the total score received on the test. When the item correlation is high it is expected that high ability test takers tend to answer the question correctly while low ability test takers tend to give an incorrect answer. After a full item analysis Yoon concluded that item difficulty was close to optimal for

the test, that the range of difficulty for questions in the Revised PSVT:R were appropriate to differentiate the abilities of students at an institute of higher learning. Yoon also determined that the validation results of the Revised PSVT:R implied there is no bias against gender in the test.

It is interesting to note that both Yoon (2011) and Guay (1980) categorized the problems in the PSVT:R in terms of rotation about multiple Cartesian axes, X, Y, and Z. It was not necessary for them to take that approach. The uniform rotation of an object can always be expressed as being rotated around a single axis. When given two images of the same shape in different positions it is always possible to express the rotation from the first image to the second image using a single axis, albeit not necessarily an axis that is parallel to one of the Cartesian axes. In this study an attempt was made to design the intervention in such a way that students were not required to take a step-by-step, or axis-by-axis, approach when rotating an object to the required position, instead a more “holistic rotation” approach was the aim. In this work the term holistic rotation is used to express the strategy of solving a rotation problem by completing the desired rotation using a single rotation around a single axis rather than in multiple rotations on an axis-by-axis fashion using the Cartesian X, Y, and Z axes. This approach considers the applied coordinate space as a whole, rather than broken down in its components. Just & Carpenter, (1985) described a Gestalt-like approach to performing mental rotation, where the path or trajectory of rotation was treated in a holistic manner – a holistic rotation. Holistic rotation has a shorter trajectory, with a rotation around a single axis.

3.3.7 Intervention Learning Module

A spatial skills teaching module was created that provides practice and challenges students to solve three-dimensional spatial rotation problems and offers students additional learning features that are more easily implemented on a computer, such as the ability to rotate images to manipulate a shape’s orientation, which could not typically be shown in a classroom setting with static graphical representations of three-dimensional objects presented in two dimensions. The images were presented so students were able, through user input, to manipulate the images presented on the screen to give them a greater insight and understanding about three-dimensional space than might be achieved through traditional methods of representing three-dimensional images in a static two-dimensional space. Students could manipulate the shapes using multiple forms of controls, including using sliding UI controls, buttons, the keyboard or the mouse. In an attempt to minimize the analytical part of

mental rotation and maximize a holistic approach, students were directed to prefer using the mouse as a method for rotating shapes. However, since mapping the X, Y two-dimensional nature of a mouse pointing device to an X, Y, and Z three-dimensional space is imperfect, controlling the mouse can be problematic for some students. Therefore, multiple methods of user input are presented in the practice part of the intervention software to enable them to focus on the rotation and not be negatively impacted by learning to control a two-dimensional device in a three-dimensional space.

Students were able to take the standardized test and the teaching module on an HTML 5 capable browser, such as Google Chrome or Firefox, which ensured a familiar interface.

A design decision was made to use traditional isometric views when creating images for the standardized tests rather than create new three-dimensional solid modelling versions from software, for example a CAD program. This allowed for the use of exact digital replications of the images used in the paper version of the Revised PSVT:R. Solid modelling shapes may provide images that are more understandable by students and contain fewer errors thereby giving a more accurate representation of a student's spatial abilities (Yue, 2009). While evidence from Yue may be correct, the decision was made to stay as close to the Revised PSVT:R test when putting the instrument online. Such a modification may make any future comparisons to PSVT:R results from other research less effective, or possibly less valid.

3.4 Human Subject Research

In order to allow students to take the pretest, the intervention, and the posttest over a time period of several months, it was necessary to track students' scores and the amount of time they spent using the intervention software. This required the system to know how and when students used the system. To facilitate this, students were required to create accounts with the system, supplying personally identifying information that included name, student identification number, and email address. Additionally, students could optionally report their date of birth, gender, and the country where they completed their high school education as a way to further classify test takers if necessary and with an eye toward further research. All user accounts were protected by passwords supplied by the students at registration. Additionally all personally identifying information was stored in an encrypted format, using AES encryption with a 128-bit key length. Email accounts were validated by sending an email with a unique code to the email address supplied by students which they could use upon logging into the system to validate the account. Collecting the personally identifying

information described classifies this research as human subject research therefore requiring that it receive approval from the Rowan Institutional Review Board (IRB). To receive IRB approval an application must be submitted by an applicant who is certified in Human Subject Research and is affiliated with the University. Dr. Stephanie Farrell, a professor in the Rowan Chemical Engineering Department, assisted in getting IRB approval for deployment of the experiment. Certification in Human Subject Research and inclusion on the IRB application was required for any person working on the project who would have access to the personally identifying student information (Appendix 7.2).

The study was approved by the Rowan University Institutional Review Board. (Appendix 7.1) Students were given the opportunity to opt out of having their results included in the study. (Appendix 7.4) Students under the age of 18 years were not permitted, based on the conditions of the IRB approval received, to be included in the study. All students, including those in the comparison group, were given the opportunity to use the intervention software after the posttest was administered. (Appendix 7.3)

3.5 Methodology

3.5.1 System Implementation

An Apache webserver (Version 2.4.12 Amazon) hosted the spatial skills testing and intervention system.² The system was comprised of two distinct services, one used for hosting data management and one for serving user webpages. The database server handled all data and metadata managed and collected by the system and interfaced to an SQL relational database system. The page server was used to present the pretest and posttest, the intervention, as well as all supporting management functionality necessary for students to use the system, such as registering, logging in and out, and reporting. The servers were written in PHP and were implemented using the MVC pattern. Programs that executed in the browser were written using JavaScript with JQuery.

The platform used to host the system was a cloud-based virtual Amazon EC2 server running Amazon Linux AMI as the operating system. (Figure 6)

Several third party libraries, platforms, and frameworks were used to support implementation of the system including SLIM, JQuery, Three.js and SendGrid. SLIM is a PHP micro

² <http://spatialresearch.us>

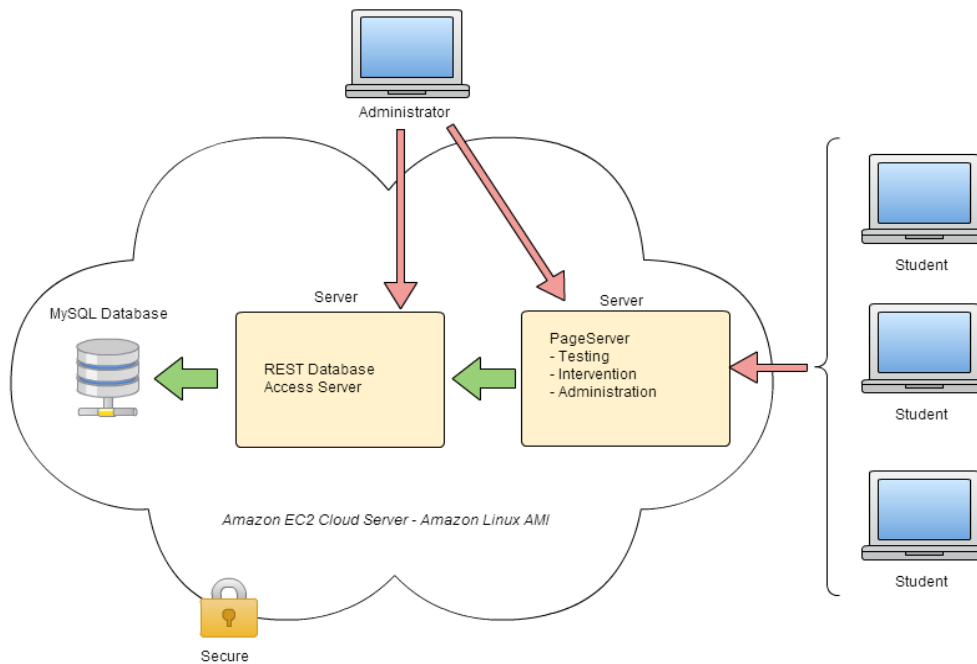


Figure 6 High Level System Architecture

framework used to implement server APIs.³ JQuery is a JavaScript library that simplifies HTML document access and AJAX transactions.⁴ Three.js is a library that simplifies use of the WebGL feature of HTML 5 capable browsers which enables animation.⁵ SendGrid is a third party email server used to communicate with registered users in order to validate email addresses and reset passwords. The relational database used within the system was MySQL (Version 5.5.42 Linux).

3.5.2 Program Design

3.5.2.1 Implementation of Testing

Because the Revised PSVT:R test is a time-sensitive instrument, it was necessary to implement the test so that all necessary information was preloaded into memory, including all questions and images. This eliminated network latency issues that might have occurred during testing that would affected the actual time a student was able to spend on the test. The system tracked both the amount of time a student spent on a test, as well as the amount of time spent on each question. The difference between the amount of time spent on the entire test and the sum of time spent on each question is the amount of latency experienced by students. Of the 331 Revised PSVT:R pretests completed by Rowan students, the average network latency

³ <http://www.slimframework.com/>

⁴ <https://jquery.com/>

⁵ <http://threejs.org/>

experienced after the test commenced was only 639 milliseconds, less than one second per student.

Student results on the test were stored in the system database via AJAX requests. A failsafe system that allowed users to reports their scoring and timing information for the test was implemented that allowed them to email in their test results in the event network issues occurred (i.e. connection problems or timeouts). Several students reported their scores through this mechanism. However, while the system had timed out in responding to these students -- for presumably network related issues -- all of their scores and timing information had been successfully registered with the system; therefore, no student data needed to be entered manually.

3.5.2.2 Questionnaires

Upon completion of both the pretest and posttests, students were presented with a questionnaire. The pretest questionnaire consisted of 49 questions and the posttest questionnaire consisted of just two questions. Broadly, the questions were concerned with gathering three types of information. The types of activities that students had engaged in that are thought to be relevant to developing spatial skills, the comfort and familiarity students have with using computers and browsers, and students' preferences for performing tasks in a traditional classroom setting or using computer software as an alternative. Earlier research into students with well-developed spatial skills has attempted to determine which activities have been common among them (Deno, 1995) (S. A. Sorby, 2009b) (Leopold, Gorska, & Sorby, 2001). Activities that develop eye-hand coordination, such as using construction toys for building structures (e.g. Lego or Erector Sets), taking middle school courses in shop (e.g. woodworking), drafting, or mechanics, playing three-dimensional oriented computer games, playing sports, especially ball-based sports, and having good mathematical skills. (S. A. Sorby, 2009b) A number of questions were asked that concern the level of engagement students had in specific activities, such as playing 3-Dimensional action video games, building models, carpentry, drawing or painting, dressmaking, repairing machines, playing contact sports, playing ball-and-bat sports, playing racket sports, etc. (Appendix 7.5)

Additionally, some demographical information was asked to determine students' home environment such as whether it was rural, suburban or urban. Students were also asked to supply the highest degree achieved by their parents in STEM fields. All questions were presented in a multiple choice format. Students were not required to take the questionnaire for

inclusion in either the experimental or comparison group of the intervention study. The software allowed students to answer as many or as few questions as they desired.

Of 285 students included in the study, 284 students answered some or all of the pretest questionnaire. 273 students answered all 49 questions, 7 answered 48 questions, and 3 students answered just 1 question. Of those included in the study 272 answered some or all of the posttest questionnaire. Of the 272 who answered the posttest questionnaire 270 answered both questions and 2 answered just 1 question. On average, 98.6% and 99.6% of the questions were answered for the pretest and posttest respectively.

As mentioned above, the questionnaire questions concerned with how frequently students used a computer and their comfort level with using a browser, indicated that on the whole students were quite familiar with using a computer to complete their daily tasks and quite comfortable using a browser.

3.5.2.3 Implementation of the Intervention Software

The training module consisted of five sections, comprised of an introduction, a rotation matching section, a rotation challenge sections, a rotation matching analogy section, and a rotation analogy challenge section. The intervention was designed to use a practice-challenge format, wherein a student can practice performing rotations or solving analogies in one section before moving on to the next section that challenges them to answer questions at various difficulty levels using the skills they practiced. As mentioned above, the intention in implementing the intervention software-based training module was to maximize students' ability to learn to use Gestalt-type processing over analytical processing and perform rotations in a holistic manner around a single axis.

In the introduction, spatial skills are defined for the user and the purpose of the module is explained to them, i.e. to improve one's mental ability to work with three-dimensional objects. Students are given a brief explanation of rotation of objects about an axis and they are shown how a three-dimensional object can be rotated around different axes. The introduction includes an overview of how to use the intervention software's features. In all of the sections students are presented with three-dimensional objects which they can control through user input. Using the mouse or keyboard, students can rotate, pan, and zoom the shapes to satisfy the requirements of the section. The "right-hand rule" for rotation about an axis is explained to students. While this is a useful piece of information for students to know, it is not necessary to

use the software. However, it was included in order to provide students the necessary language to talk about three-dimensional rotation in other contexts. In addition to the introduction, there is a detailed help page to which students can refer when using the intervention module.

There are two types of rotation a student can perform: *synchronized* and *relative*. The type of rotation permitted on a three-dimensional space is indicated by an icon on the lower left. With synchronized rotation the user can rotate an entire three-dimensional space; when doing so, all other related three-dimensional spaces being displayed will be rotated to the same exact degree thus allowing students to examine all shapes from multiple perspectives in order to gain a fuller understanding of how the shape occupies the space and how it is constructed. The aim is two-fold, to make it possible for the student to more easily understand the shape being presented, and to encourage the student to think of the shape more holistically. Using relative rotation the user can rotate a shape within its three-dimensional space, changing the shape's position in the space. This is how a user would change the shape's rotation to match that of another shape.

In the *Match Rotation* section of the intervention software, students are presented with two three-dimensional spaces each containing an identical shape that are randomly rotated differently within their respective spaces. Students are asked to rotate the shape on the right side to match the rotation of the shape on the left side. For a match to be considered successful students must come to within twenty degrees of the rotation of the left shape. The purpose is not get students to match the shapes to the exact degree of rotation, but to be able to rotate the shape in such a manner so that it appears to be the same. In a three-dimensional space, a rotation difference of twenty degrees, or about 95% the same, is not very perceptible to the eye. When students rotate to within that degree they get practical rotation experience and demonstrate to themselves that they are understanding the shape and rotating it in a similar manner. As an aid to find the correct rotation students can see, in real time, how many degrees their rotation differs from the solution. There are four main methods that can be used to manipulate a shape into position: the mouse can be dragged directly on the shape to manipulate its orientation. Because a typical computer mouse is designed to map coordinates to a two-dimensional space, i.e. the computer screen, it can prove challenging for some users to control a shape's rotation when the mouse is mapped to a three-dimensional space. For this reason, several other input methods for rotating the shape are provided. There are three sliding controls each mapped to the X, Y, and Z axis in the space. Each slider can be used to rotate the shape independently around each respective axis. There are a set of six rotation buttons that increment or decrement the rotation

of the shape about the three axes. The buttons are coupled with a set of four radio buttons with which the user can set the number of degrees to rotate the shape when a rotation button is clicked. Finally the user can control the rotation of the shape using the keyboard by pressing the desired axis key (**X**, **Y** or **Z**). They can select the desired number of degrees using the **1**, **4**, **5** or **9** keys where the **1** key is for 1° , the **5** key is for 5° , the **4** key for 45° , and the **9** key for 90° . The **P** key is used for positive rotation and the **M** key for negative rotation. Because the slider method, the rotation button method, and the keyboard method of shape control all rotate the shape on a single, isolated axis, these methods are considered more analytical than when the shape is rotated using the mouse. For example, using a more analytical method, such as the slider provide, a user might first rotate the shape 90° around the X axis and then 180° around the Z axis breaking the ultimate rotation down into different parts. A more holistic method would be to use the mouse to rotate the shape to the final position using a single rotation axis that does not align with the normal Cartesian X, Y, and Z axes.

In the *Rotation Challenge* section the user must first select a difficulty level, either easy, medium, or hard. The difficulty level affects three properties of the questions presented: the color, the degree of rotation, and whether or not the X, Y and Z axes are displayed. In the easy level the user rotates a colorized shape. In the medium and hard levels the student does not get to see any color on the rotated shape, instead it is a uniform white or gray. The shape on the left is always colorized. The shape rotation for easy problems is in 90° increments, for medium it is 45° increments and for hard it is 5° increments. For easy problems the three-dimensional space for both shapes have Cartesian axes displayed, for medium just the left shape has axes, and there are no axes displayed on the hard level.

After selecting the difficulty level, the user is presented with ten rotation problems, similar in function to the match rotation section, with the same shape presented at two different rotations where they must match the shape on the right to be rotated to within twenty degrees of the shape on the left. In this challenge, the rotation buttons have been eliminated narrowing the choice of analytical input methods. Correct answers are rewarded with gold stars of varying color depth and design depending on how close the user comes to the exact rotation. After answering a question the user is presented with an animation of the shape rotating from the answer they supplied to the exact solution. The further off the student's answer, the more rotation can be observed in the animation. This gives the student visual feedback on how the shape should have been rotated.

In the *Match Analogies* section the student is presented with four three-dimensional spaces each containing a single shape. The spaces are organized in a two-over-two layout. The top two spaces contain the same shape, with the shape on the right rotated differently from the shape on the left. The bottom two spaces also have the same shape as each other, but it is a different shape from the top two spaces. The user must rotate the bottom right shape relative to the bottom left shape to the same degree (within 20°) as occurs in the top two spaces. (Figure 7) In both the *Match Rotation* and *Match Analogies* practice sections there are some options the student can set using checkboxes to aid in their visual interpretation of the shapes and the three-dimensional space. The Cartesian X, Y, and Z axes can be turned on and off by the user. A grid can be displayed on the plane of each set of intersecting axes. Additionally a black outline of the edges for the shapes can be turned on and off. (Also Figure 7)

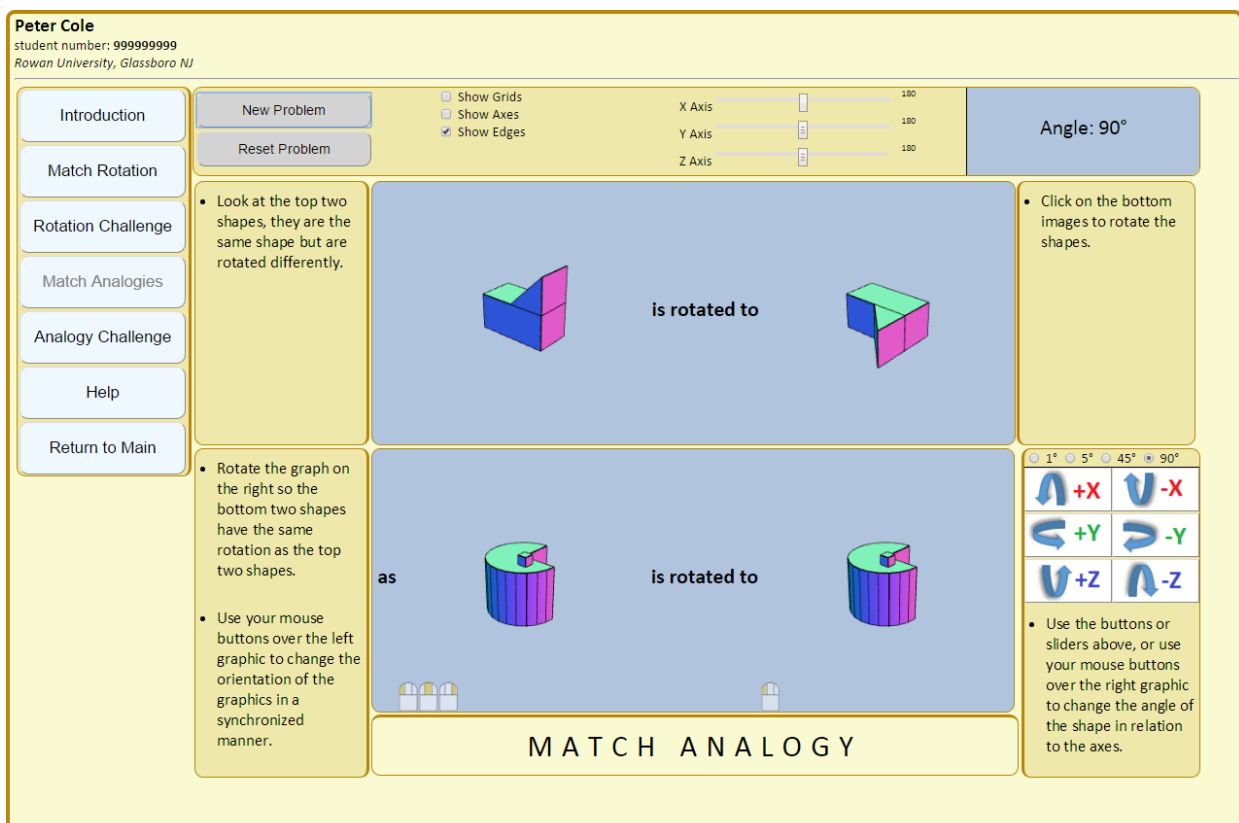


Figure 7 Screen image of the Match Analogy Section of the Intervention Software

In the *Analogy Challenge* section the user must first select a difficulty level, either easy, medium, or hard. The user is then presented with ten rotation analogy problems. On the top of the screen the user is presented with a shape analogy question in the same format as in the

Match Analogies section. On the bottom of the screen the user is presented with a different shape from the shape analogy question and five multiple choice answers, one of the choices has the same rotation as presented in the shape analogy and is the correct answer to the problem. As in the *Rotation Challenge*, after an answer is selected the student is rewarded with a gold star if correct or given a red X if incorrect. Incorrect selections are highlight in a red box and the correct answer is highlighted in a green box. Synchronized animations are performed that demonstrate to the student how the top and bottom shapes should be rotated to match the analogy. (Figure 8)

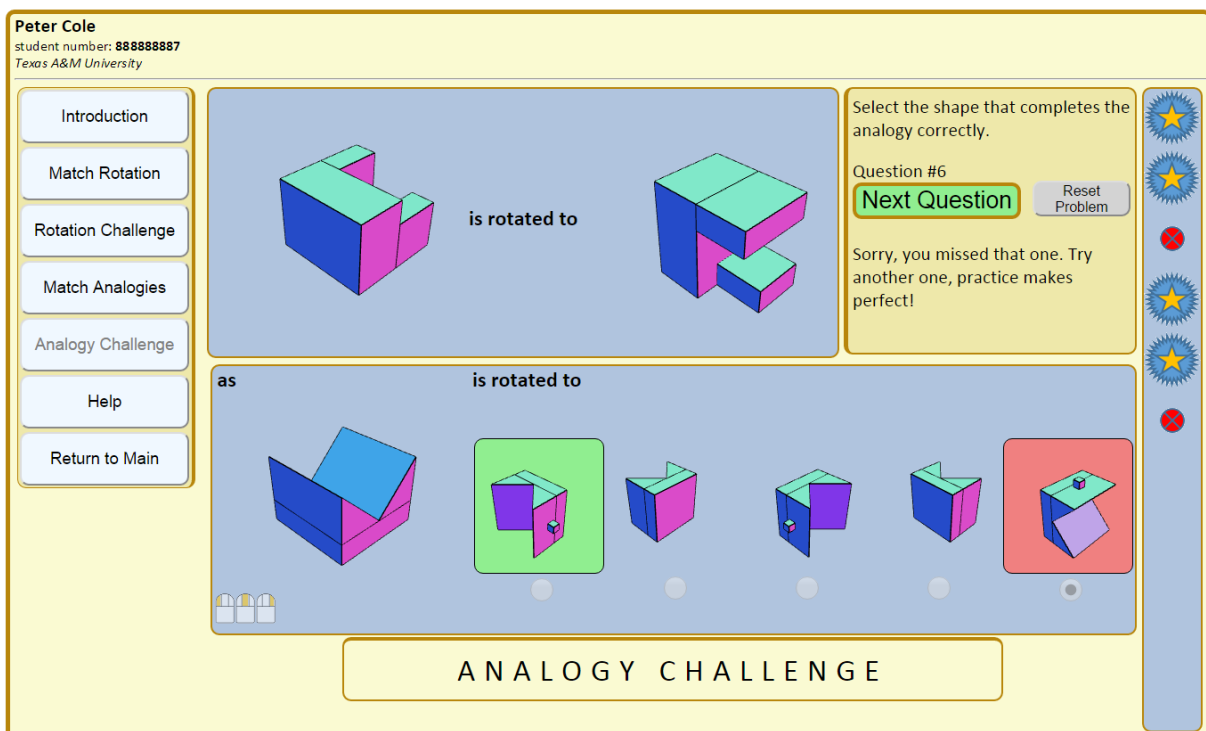


Figure 8 Screen image of the Analogy Challenge Section of the Intervention Software

The difficulty levels affect three properties of the questions presented: the degree of rotation, the colorization of the shapes, and whether the top shape is a random shape. The shape rotation for easy and medium problems is in 90° increments, for hard problems it is in 45° increments. For easy problems all shapes are colorized. For medium problems just the top two shapes that present the analogy are colored. For hard problems none of the shapes are colorized and are displayed in a uniform gray. For the easy level the two analogy shapes on the top are always presented as the same u-shaped object as displayed below (Figure 8). For the medium and hard levels, the shape is randomly selected.

In all sections of the intervention software, shapes are presented in a colorized format. However, the color of the shapes change incrementally as the shape is rotated. For example, the side of a shape may appear blue when rotated to a certain degree but a different shade when rotated to a further degree. This effect was incorporated to deter students from using an analytical approach to solving rotation problems where they are simply remembering the colors of one or two of the sides of the shape. There is no set limit or required number of problems to solve in the practice sections and the user can spend as much time as they need to solve a problem.

As students use the Intervention module their results in the practice and challenge sections are recorded. In all four sections, the number of questions asked is recorded in the database. In the challenges the number of correct answers is also tracked. The amount of time a student spent in each module is also tracked in the database. In the challenge sections, the amount of time spent on each 10-question challenge is tracked.

3.5.2.4 Data Management & Security

Tracking of data, including student scores and certain demographic information such as age and gender were stored in a custom designed relational database. Additional data about the test taking and intervention experience of the students was tracked, such as when the students took the pretests and posttests, for how long a student spent on each question of the test, student pretest and posttest scores, how much time passed between the intervention and the posttesting, and for how long students used the intervention software. All personally identifiable student information is stored in an encrypted format. The database as well as the server on which it resides is secured and password protected at a login level as well as at a transaction level. In Figure 9 a high-level, entity relationship diagram (ERD) for the database implementation shows the major components of the database and their relationships.

3.6 Conclusion

The design and methodology topics of the study were discussed. A PPC experimentation model, which is common to many other spatial skills studies, was implemented in this study. Similarities and differences between the online version and the paper version of the Revised PSVT:R were detailed and discussed from a psychometric perspective. Issues such as test timing, student navigation, test questions, test question order, imaging, and presentation were all examined and the related design decisions were explained and justified. Computer-based testing issues, such as mode effect, presentation latency, test construction and reliability were

addressed. Test question characteristics, such as item discrimination and item difficulty, were examined and the concept of “holistic rotation” was introduced. The online instrument has not been psychometrically validated, but was shown to have internal consistency based on results from a Cronbach’s α calculation. Compliance with human subject research requirements and IRB approval for the study were demonstrated. The system implementation and design was detailed, including the design of the testing portion of the system, the questionnaires used, the implementation of the intervention software, and the management of the system data.

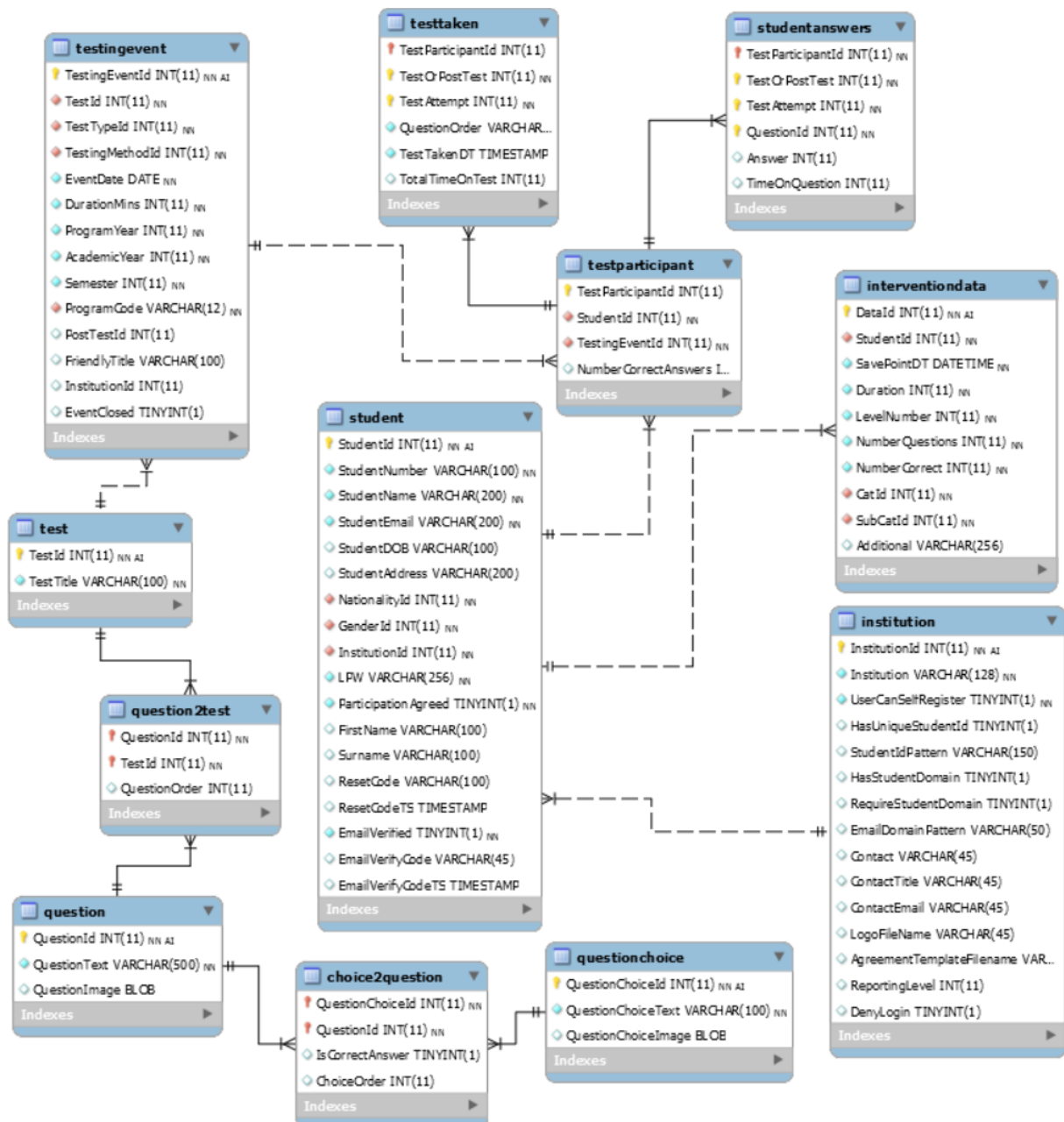


Figure 9 High-Level Entity Relationship Diagram

Chapter 4 Experiment

4.1 Introduction

The study took place at the College of Engineering at Rowan University in Glassboro, NJ, USA throughout the autumn semester, 2015.

All students were required by the engineering department as part of the Freshman Engineering Clinic curriculum to take the online PSVT:R pretest. It was completed by 339 students across 16 sections of the module. Not all students that took the pretest completed the posttest, and therefore were disqualified from inclusion in the study. Some students chose not to participate in the study. All students who did not opt out of participation in the study and who took both the pretest and the posttest are included in the study in either the comparison group or the intervention group. A total population of 285 students took both the pretest and the posttest and were included in the final study results. In the comparison group there were 163 students, in the intervention group there were 122. The sample population included 228 males and 57 females. Of the females included in the study, 28 were in the experimental group and 29 were in the comparison group. Of the males included in the study, 94 were in the experimental group and 134 were in the comparison group. The age range of participating students was from 18 to 20 years.

With the exception of four students who missed the initial pretesting, pretests were administered in the first class session of Freshman Clinic during the first week of the semester. Posttests were taken in the two week period from 25-Nov-2015 through 17-December-2015. The average number of days between administration of the pretest and the posttest was 92 days, approximately 3 months. (Table 2)

All students were given the opportunity to use the intervention software. Those students who were in the comparison group were given an opportunity to use the intervention software after taking the posttest. Their intervention results are not included in the study. The intervention software will remain available to all students for the remainder of the academic year.

Participation in the intervention during class time was decided at the discretion of the instructor for each section of the course. Seven sections participated in the intervention during assigned class time. All students who performed the intervention were able to make use of the intervention software whenever it suited them; they were not just limited to using during class time. Consequently, there were large differences in the amount of time each student spent using

Pretest Administration Dates		
Date	Test Type	Number of Tests
2015-09-01	PRE	118
2015-09-02	PRE	108
2015-09-04	PRE	55
2015-11-10	PRE	4
		285

Posttest Administration Dates		
Date	Test Type	Number of Tests
2015-11-25	POST	1
2015-11-26	POST	1
2015-11-27	POST	2
2015-11-28	POST	4
2015-11-29	POST	3
2015-11-30	POST	33
2015-12-01	POST	24
2015-12-02	POST	66
2015-12-03	POST	23
2015-12-04	POST	21
2015-12-05	POST	1
2015-12-06	POST	2
2015-12-07	POST	3
2015-12-08	POST	97
2015-12-11	POST	3
2015-12-17	POST	1
		285

Table 2 Pretest & Posttest Administration Dates

the software. The minimum time spent doing intervention work by a student in the experimental group was 6.7 minutes, the maximum time was 572.5 minutes, with a δ of 114.9 minutes.

The study tracked the amount of time each student used the intervention software as well as the number of problems attempted in each section. It also tracked the number of questions answered correctly in the challenge sections.

4.2 Intervention Time Tracking

It should be noted that tracking the amount of time a user uses a piece of software is an inherently nebulous undertaking. It is not sufficient to simply record how long a student is logged into the software. Generally, there were two types of issues uncovered in the analysis of the intervention timing data. In the first, students remained logged into the intervention software for long periods of time without making progress, i.e. they appeared to be idle. This most likely occurred when the student left the software running in the browser without logging off. In the second, students appeared to be clicking through problems very quickly without actually spending time solving them, perhaps in an attempt to find practice problems that suited them. As an example, a student could log into the system, enter an intervention section and do some problems, however, they could leave in the middle, possibly for a long time, and then resume working later. This would make it appear in the data that the student was working on the intervention for much longer than they were. It was necessary to make some assumptions when interpreting the amount of time a student spent working on the intervention. A filtering algorithm was developed that attempted to correct the timing given the two issues mentioned.

It attempted to ignore stretches of time where it appeared the student was logged into the system, but was not actively working on solving problems. Likewise it attempted to ignore events where it seemed that the student was addressing problems too quickly. An example of the type of filtering performed is that any record where a student addressed 30 or more problems per minute was filtered out. Similarly, if a student took more than an hour to answer a set of 10 problem, it was assumed that the student had not logged out of the system. Table 3 shows the percentage of students whose intervention timing data was unaffected by filtering of various amounts of minutes. 84.3% of students had fewer than 20 minutes filtered from their overall intervention time, with an average loss of just 3.0 minutes due to filtering. Approximately 13.2% of students had large amounts of intervention time, ranging from 1.3 to 278.7 hours, which were filtered out. In these cases it was determined that the student had clearly stopped using the intervention software and had simply left it running within their browsers. Despite the difficulties in tracking students' time doing intervention work, it should also be noted that the number of problems a student solved is a reliable statistic that was collected indicating the work that was performed.

Fewer than X Minutes Lost to Filtering (<)	Percentage		Number Students Unaffected	Number Students Affected	Average Time Lost (Mins)
	Students Unaffected by Filter	Students			
60	86.8%	105	16	3.9	
20	84.3%	102	19	3.0	
15	81.8%	99	22	2.6	
10	79.3%	96	25	2.2	
5	68.6%	83	38	1.5	

Table 3 Effect of Filtering Intervention Timing Data

4.2.3 Filtering Algorithm

The algorithm for filtering intervention timing records was based on the type of intervention work being performed by the student, either using the introduction, doing one of the practice drills, or playing one of the rotation challenge games. (Figure 10) Tracking records that indicated a student spent less than 1 minute or more than 1 hour reading the introduction screens were rejected from the study. The introduction consists of seven screens that include some minor student interaction. It is very unlikely that a student would spend less than a minute or need more than an hour to go through the screens. Practice records were rejected from the study if students were hyper-clickers who were looking at more than 30 problems a minute. Even a student with very good spatial skills would be unable to sustain such a high rate of execution of problems if they were seriously attempting to answer the questions. Records

```

1 // Filtering Pseudo Code for Intervention Effort
2 if (TrackingRecordType == IntroductionRecord) {
3     if (TimeSpent > 1 Minute and TimeSpent < 60 Minutes) {
4         AcceptRecord()
5     }
6     else {
7         RejectRecord() // Eliminates abandoned introduction sessions and quickie visitors.
8     }
9 }
10
11 if (TrackingRecordType == PracticeRecord) {
12     if (NumberQuestionsPerMinute < 30 and TimeSpent > 5 Seconds and NumberQuestionsAnswered > 0) {
13         AcceptRecord()
14     }
15     else {
16         RejectRecord() // Eliminates Hyper-Clickers and abandoned sessions and quickie visitors.
17     }
18 }
19
20 if (TrackingRecordType == ChallengeRecord) {
21     if (CompletedChallengeGame) {
22         // A completed challenge game consists of 10 questions.
23         if (TimeSpent > 1 Minute and TimeSpent < 60 Minutes) {
24             AcceptRecord()
25         }
26         else {
27             RejectRecord() // Eliminates abandoned games and quickie visitors.
28         }
29     }
30
31     if (!CompletedChallengeGame) {
32         if (SelectedDifficultyLevel) {
33             if (TimeSpent > 1 Minute and TimeSpent < 60 Minutes) {
34                 AcceptRecord()
35             }
36             else {
37                 RejectRecord() // Eliminates abandoned partial games and quickie visitors.
38             }
39         }
40         else {
41             RejectRecord() // Eliminates unstarted challenges.
42         }
43     }

```

Figure 10 Pseudo Code for Intervention Effort Filtering

indicating a student spent less than 5 seconds in a practice session were also rejected from the study. From the collected data, it became apparent that some students would start a practice session but leave after just a few seconds rather than actually practicing rotations. Possible reasons for this behavior could be that they were trying to familiarize themselves with the software or they clicked on the wrong button. Also, tracking records that indicated the student failed to answer a single question were rejected from the study; this does not mean failed to answer a single question correctly rather it means failed to give any answer at all. Challenge game records were rejected from the study under several scenarios. If the student completed a challenge game, i.e. answered all 10 questions, but did so in under 1 minute, or did so in over

an hour, the record was rejected from the study. Some instructors assigned students a homework task to complete a certain number of challenge questions. It is suspected that some students did not make an effort to answer the questions correctly and just wanted to complete the number required by the instructor, therefore they completed challenge games in under a minute. Games that took a student more than an hour, were likely games that the student had stopped playing in the middle and completed later. For students that spent more than 1 minute and less than an hour on a single challenge game, the average time spent to complete all 10 questions was 8.4 minutes for the rotation challenge and 5.1 minutes for the analogy challenge. For challenge games that were not completed by the student, i.e. fewer than 10 questions answered, if the student had not selected a difficulty level or the student spent less than a minute or more than an hour the record was rejected from the study.

4.3 Amount of Intervention

Intervention work was begun by students on 2015-Nov-04 and was completed on 07-Dec-2015 and 13-Dec-2015 by the last female and male respectively. Presented in Table 4 are descriptive statistics for the different population subgroups on the intervention work performed by students in the experimental group. The mean amount of time spent performing intervention work (after filtering as described above) was 105.9 minutes. (Complete descriptive statistics are presented in Table 11 & Table 12 in Appendix 7.6)

Intervention Type	Females		Males		Total Population	
	Mean	Std. Dev. (δ)	Mean	Std. Dev. (δ)	Mean	Std. Dev. (δ)
Overall						
Minutes on Intervention	102.2	136.1	107.0	107.8	105.9	114.9
Aggregate - Rotation Challenge						
Minutes	42.4	60.6	27.8	29.7	31.2	39.5
Number Correct	44.7	61.0	34.1	35.5	36.5	43.0
Number Questions	54.2	79.9	41.0	45.7	44.0	55.7
%					82.9%	
Aggregate - Analogy Challenge						
Minutes	42.8	72.3	67.7	82.2	62.0	80.7
Number Correct	38.2	64.1	81.4	90.5	71.5	87.1
Number Questions	66.0	116.0	139.9	169.0	122.9	161.4
%	57.9%		58.2%		58.2%	

Table 4 Descriptive Intervention Statistics for the Different Population Subgroups

4.4 Student Motivation

After the completion of the experiment for this study, it became apparent that perhaps, some students were not properly incited to try their best on either the testing or the intervention. As

mentioned above, several of the instructors for the module assigned intervention work to the students for which they received a grade; however, most did not. There was a wide variation in the amount of time students spent using the intervention software. After the filtering of the intervention timing data, the minimum time a student spent on the intervention was 6.7 minutes and the maximum time was 9.54 hours. There also appeared to be some discrepancies with the amount of time several students spent answering questions on the PSVT:R. One piece of anecdotal evidence can be seen in the scores of one student who dropped ten points from pretest (24 correct) to posttest (14 correct). His or her time spent on the pretest was 17.1 minutes and 9.8 minutes on the posttest. In itself that total time difference is not remarkable, but the time spent on 13 of the questions was under 2 seconds, of which 9 were incorrectly answered. Another factor that was considered is that after taking the pretest, student's scores were reported to them immediately by the software. It is possible that students who performed well on the pretest may have had low incentive to spend time using the intervention software.

4.5 Conclusion

An experiment to measure spatial skills was executed at Rowan University in the autumn semester of 2015. It tested and trained a sample population of 285 students over a 92 day period. Students participated as part of an experimental or a comparison group. Students in the experimental group were trained using a software-based intervention lesson on mental rotations. Issues with the challenges of accurate timing data collection were discussed and addressed through a filtering algorithm to eliminate outlier tracking records. The average time students trained was approximately 106 minutes. Student motivation was discussed as a potential issue, however no students were removed from the study based on an apparent lack of motivation.

Chapter 5 Evaluation

5.1 Introduction

The relational database was used in conjunction with the R statistical language to analyze the results of the pretests and posttests using quantitative research methods. Descriptive statistics (mean, standard deviation and number of participants) for both pretest and posttest are presented. T-tests were used to test the significance of the difference between pretest and posttest for the entire population as well as broken down into specific group segments, such as by gender and by whether or not the student had failed the pretest. The gain for the experimental group was compared to the gain for the comparison group for each category of student analyzed. Another t-test was used to test for the significance of the difference between groups. Two-way Anova methods were used to assess difference in outcomes for significance when more than one comparison factor was present. The groups were also compared using an effect size (Cohen's d) using pretest and posttest scores to compute gain.

According to Stanley & Hopkins, (1972), any gain due to “*practice effect*” for a test given 3 months apart is typically less than 0.2σ . Practice effect is the gain in score a student is likely to achieve from having taken the test previously. The standard deviation for the pretest for the number of correct responses for the total population was 4.97 giving an expected practice effect gain of 0.99 on the posttest. The mean gain between pretest and posttest for the number of correct responses for the total population was 1.85, which is 0.86 more than the expected practice effect.

5.2 Roadmap

The evaluation performed included a number of statistical tests used to determine if a statistical significance can be demonstrated between various segments of the data in both filtered and unfiltered forms. Notched boxplots showing 95% confidence intervals show where statistical differences are likely to be present. T-tests are performed to demonstrate the presence or absence of statistical differences between pretest and posttest scores and their gains. For segmented data where more than one factor is present two-way Anova tests are performed to show statistical differences in the means. The data are filtered and segmented in five different ways to demonstrate the effect of the intervention on students' scores.

First the analysis is performed on the total population without any filtering or segmenting of the data. T-tests are sufficient because there are means for only the experimental group and the comparison group as the single factor. Second the analysis is performed with a filtered subset

of the total population based on the intervention effort put forth by the students. Recall that there was a wide disparity between individual efforts made using the intervention software. Again the experimental and comparison groups are compared. The final three sets of data analyzed use the intervention effort filtered data as a base. Besides comparing by experimental and comparison group, the data is segmented by gender, then further filtered by students that failed the pretest, and finally segmented by gender again. Because there is more than one factor of comparison two-way Anova tests are performed. Also for each data grouping a Cohen's d effect size is computed for groups where statistical differences in the means could be demonstrated.

For each data segment analyzed, five separate t-tests were performed either comparing different elements between a single segment of the data, e.g. comparing the pretest and posttest number of correct responses for the experimental group; or comparing the same element between different data segments, e.g. comparing the gain between the experimental and comparison groups. The five t-tests are:

- Experimental Group only, Pretest Number Correct to Posttest Number Correct
- Comparison Group only, Pretest Number Correct to Posttest Number Correct
- Experimental Group to Comparison Group, Pretest Number Correct
- Experimental Group to Comparison Group, Posttest Number Correct
- Experimental Group to Comparison Group, Gain

Comparing the mean pretest to the mean posttest number correct in a single data segment makes it possible to demonstrate whether or not there is a statistically significant change in the mean for the group. For example, if there is a significant gain from pretest to posttest for the experimental group, it can be inferred that the intervention given to the experiment group had an effect. (Given that the effect is greater than the expected practice effect.)

Comparing the mean pretest number correct or the mean posttest number correct between two data segments makes it possible to demonstrate that the means of the two groups are statistically the same. For example, in all of the data analyzed it will be shown that there was no statistically significant differences found between the means of the experimental and comparison groups for how they scored on the pretest.

Comparing the mean gains between the experimental and comparison groups makes it possible to demonstrate whether the difference in means of the gain for each group are significant.

Significantly different means in the gain between the groups makes it possible to infer whether the group that received the intervention performed significantly differently, implying the effect of the intervention relative to not having the intervention.

All t-tests were performed with a 95% confidence interval. Population sizes for all data segments analyzed are presented in Table 5.

Datasets with Filtering	Males EG	Males CG	Females EG	Females CG	EG (Both Genders)	CG (Both Genders)	Dataset N
Total Population (No Filtering)	94	134	28	29	122	163	285
Filtered by Intervention Effort (IE)	75	134	19	29	94	163	257
Filtered by IE & Failed Pretest	20	31	6	13	26	44	70

Table 5 Population Sizes for all Populations Analyzed

5.3 Data Analysis

5.3.4 Total Population

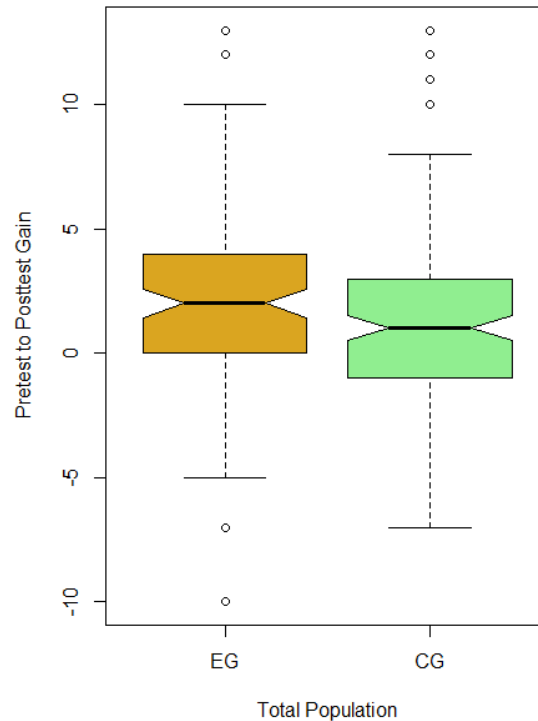
When comparing the experimental group to the comparison group for the entire population (N=285), there was a gain in the mean student score from pretest to posttest for both the experimental group and the comparison group. The mean score for the experimental group improved from 21.31 on the pretest to 23.67 on the posttest, for a gain of 2.36. The mean score for the comparison group improved from 21.57 on the pretest to 23.04 on the posttest, for a gain of 1.47 more questions answered correctly.

Population 1				Population 2				Delta				Degrees of Freedom	
Comparison Value				Comparison Value								P-Value	
Mean	Std. Error	Mean	Std. Error	Delta	CI	CI	T	Degrees of Freedom	P-Value				
EG Pretest Correct	21.31	0.46	EG Posttest Correct	23.67	0.39	2.36	-3.089	-1.632	-6.416	121.00	0.0000000028		
CG Pretest Correct	21.57	0.39	CG Posttest Correct	23.04	0.37	1.47	-1.985	-0.960	-5.677	162.00	0.0000000620		
EG Pretest Correct	21.31	0.46	CG Pretest Correct	21.57	0.39	0.26	-1.436	0.918	-0.433	257.25	0.6651000000		
EG Posttest Correct	23.67	0.39	CG Posttest Correct	23.04	0.37	-0.63	-0.431	1.689	1.168	271.79	0.2437000000		
EG Gain	2.36	0.37	CG Gain	1.47	0.26	-0.89	0.001	1.775	1.973	228.94	0.0496700000		

Table 6 Total Population T-Tests (EG) Experimental Group; (CG) Comparison Group

A paired t-test was performed comparing the experimental group's pretest and posttest means. On average, students who participated in the intervention showed a gain in scores between taking the pretest and the posttest. The gain of 2.36 more questions answered correctly was significant with a p-value < 0.001. That is a gain of 11.1% (2.36 / 21.31) from pretest to posttest. The same test was performed for the comparative group, on average, students who participated in the intervention showed a gain in scores between taking the pretest and the posttest. The gain

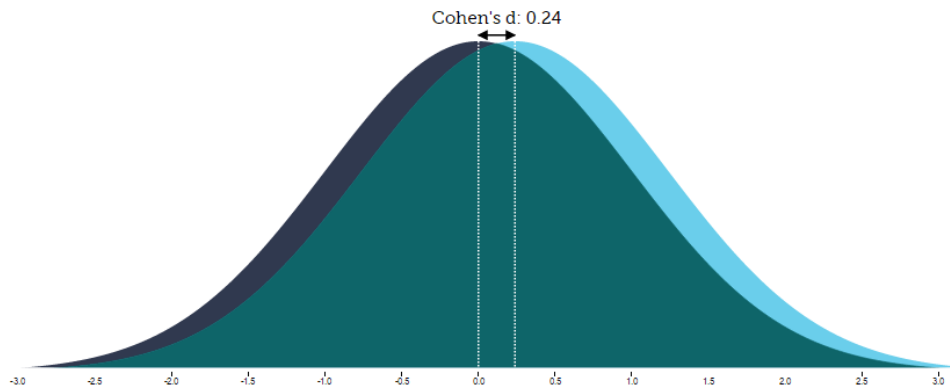
of 1.47 more questions answered correctly, an improvement of 6.8% (1.47 / 21.57), also was significant with a p-value < 0.001. (Table 6 & Plot 1)



Plot 1 Total Population EG Gain vs CG Gain

A two sample t-test was performed comparing the means of the pretest scores for both groups and there was no statistically significant difference found (p-value = 0.665) suggesting that the group performed similarly on the pretest. A comparison of the means of the gains between the groups suggests a statistically significant difference (p-value = 0.049). The comparison group's increase in gain was slightly higher than that anticipated by the computed practice effect of 0.99. The increase in mean score for the comparison group may be due to the expected practice effect plus any effect that may have occurred from students having just completed three months of course work in higher level math and other STEM related courses. A posttest comparison of means between the two groups also showed no statistically significant difference in the means. However, as noted, there was a significant difference in the gain. The discrepancy between the two t-tests is due to the fact that while the posttest means were statistically the same the experimental group had a lower pretest score and a higher posttest score than the comparison group creating enough of a difference for the mean gains to be significantly different.

An effect size (Cohen's d) of 0.24 was computed for the gain between the experimental and comparison groups for the total population. (Plot 2) With a Cohen's d of 0.24, 59% of the intervention group will be above the mean of the comparison group, 90% of the two groups will overlap, and there is a 57% probability of superiority, the chance that a person picked at random from the experimental group will have a higher score than a person picked at random from the comparison group.⁶



Plot 2 Cohen's d Effect Size for Total Population

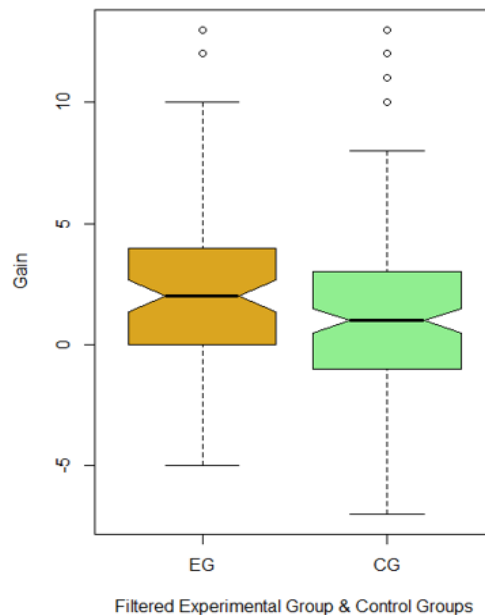
5.3.5 Filtered by Student Intervention Effort

Because of concern that a segment of the student population that participated in the intervention did not perform to the best of their ability the data for the total population was filtered in an effort to more truly reflect that effect of the intervention using only those participants that met a certain threshold of effort when using the intervention software. The filter was created based on two significant pieces of data collected about student's use of the software: the number of complete analogy challenge games completed and the total score, as a percentage, that a student received for all the analogy challenge questions that were answered. The threshold was kept as low as possible in order to keep as many students in the study as possible. Students that completed at least 1 analogy challenge game (a total of 10 questions) and scored at least 30% correct for all the analogy challenge questions they answered were kept in the filtered study dataset. A student that either did not complete a single game, or who performed so poorly that a score of 30% or above could not be maintained was dropped from the dataset. These threshold levels were selected as a minimum indication that the student was making a valid attempt to use the intervention software and improve their rotation analogy abilities. The 30% threshold was selected as a score that was slightly higher than a score of 20% which a student could

⁶ Analysis performed, graphic created, and paraphrased from <http://rpsychologist.com/d3/cohend/>

achieved simply by guessing. The total population for the experimental group was 122 before filtering for intervention effort and 94 afterwards, for a reduction of 28 students or 22.9% of the total experimental group population. The remaining 94 students are the population considered as the experimental group who had completed some level of training.

The same five t-tests were performed on the filtered experimental group against the original comparison group with similar, but slightly improved results. The experimental group showed a statistically significant improvement of 2.72 questions correct from pretest to posttest. This is 0.36 higher than the improvement shown with the total population (2.72 – 2.36). The difference in gain between the experimental group and the comparison group was -1.25, indicating that the comparison group gained 1.25 fewer points than the experimental group. (Note that in the t-test data tables presented in this section, the delta between tests are always calculated as population 2 value – population 1 value.) The experimental group had a gain of 12.7% (2.72 / 21.4) and the comparison had a gain of 6.8% (1.47 / 21.57). As in the total population tests, the pretests for both groups showed no significant difference in the mean of the number of correct answers indicating that both groups started from statistically similar means. (Table 7 & Plot 3)



Plot 3 Filtered EG Population vs CG Gain Pretest to Posttest

Population 1		Comparison Value		Population 2		Comparison Value		Delta				Degrees of Freedom	P-Value
		Mean	Std. Error			Mean	Std. Error		CI	CI	T		
EG	Pretest Correct	21.40	0.51	EG	Posttest Correct	24.13	0.36	2.72	-3.484	-1.963	-7.115	93.00	0.0000000002
CG	Pretest Correct	21.57	0.39	CG	Posttest Correct	23.04	0.37	1.47	-1.985	-0.960	-5.677	162.00	0.0000000620
EG	Pretest Correct	21.40	0.51	CG	Pretest Correct	21.57	0.39	0.17	-1.431	1.098	-0.259	192.77	0.7956000000
EG	Posttest Correct	24.13	0.36	CG	Posttest Correct	23.04	0.37	-1.08	0.072	2.097	2.111	240.76	0.0358400000
EG	Gain	2.72	0.38	CG	Gain	1.47	0.26	-1.25	2.723	0.217	2.706	176.64	0.0074860000

Table 7 (EG) Experimental Group & (CG) Comparison Group Filtered by Intervention Effort

An effect size (Cohen's d) of 0.36 was computed for the gain between the experimental and comparison groups for the total population. With a Cohen's d of 0.36, 64% of the experimental group will be above the mean of the comparison group, 86% of the two groups will overlap, and there is a 60% chance that a person picked at random from the experimental group will have a higher score than a person picked at random from the comparison group. This is a 50% increase in effect size from that of the total population.

5.3.6 Intervention Effort Segmented by Gender

In order to determine if the intervention had a significant effect across gender the data was further segmented. For the same population of students who completed some level of training (N=257) the data were further broken down into subsets by gender giving four groups for comparison: the experimental group for males (N=75) and for females (N=19), and the comparison group for males (N=134) and for females (N=29). The same battery of t-tests as performed for the previous data sets was performed for each gender yielding comparisons within each gender. (Table 8) To compare the means across both gender and study group, a two-way Anova was performed for both factors.

For females in the experimental group there was a statistically significant increase from pretest to posttest of 3.21 additional correct answers, a gain of 15.8% (3.21 / 20.26) that suggests that the intervention had an effect. (Plot 4) For females in the comparison group there was also a significant, although smaller, increase of 6.6% (1.31 / 19.79) from pretest to posttest, again likely due to practice effect and the semester's coursework. The increase for females in the experimental group was 2.45 times greater than that of the comparison group suggesting that females responded well to the intervention software. The average time spent by females on intervention work was 102.2 minutes (Table 4) suggesting that even a modest amount of time using the intervention software proved beneficial.

Population 1		Comparison Value	Mean	Std. Error	Population 2		Comparison Value	Mean	Std. Error	Delta	CI	CI	T	Degrees of Freedom	P-Value
FEMALES															
EG	Pretest Correct	20.26	1.19	EG	Posttest Correct	23.47	0.84	3.21	-5.046	-1.375	-3.674	18.00	0.0017360000		
CG	Pretest Correct	19.79	0.86	CG	Posttest Correct	21.10	0.82	1.31	-2.483	-0.137	-2.289	28.00	0.0298600000		
EG	Pretest Correct	20.26	1.19	CG	Pretest Correct	19.79	0.86	-0.47	-2.518	3.458	0.319	35.44	0.7514000000		
EG	Posttest Correct	23.47	0.84	CG	Posttest Correct	21.10	0.82	-2.37	0.016	4.725	2.030	43.36	0.0485700000		
EG	Gain	3.21	0.87	CG	Gain	1.31	0.57	-1.90	-0.226	4.026	1.819	32.88	0.0780600000		
MALES															
EG	Pretest Correct	21.69	0.57	EG	Posttest Correct	24.29	0.39	2.60	-3.452	-1.748	-6.083	74.00	0.0000000477		
CG	Pretest Correct	21.96	0.42	CG	Posttest Correct	23.46	0.41	1.51	-2.083	-0.932	-5.180	133.00	0.0000008033		
EG	Pretest Correct	21.69	0.57	CG	Pretest Correct	21.96	0.42	0.26	-1.659	1.135	-0.370	153.77	0.7116000000		
EG	Posttest Correct	24.29	0.39	CG	Posttest Correct	23.46	0.41	-0.83	-0.286	1.947	1.467	193.26	0.1440000000		
EG	Gain	2.60	0.43	CG	Gain	1.51	0.29	-1.09	0.070	2.115	2.113	141.58	0.0363700000		

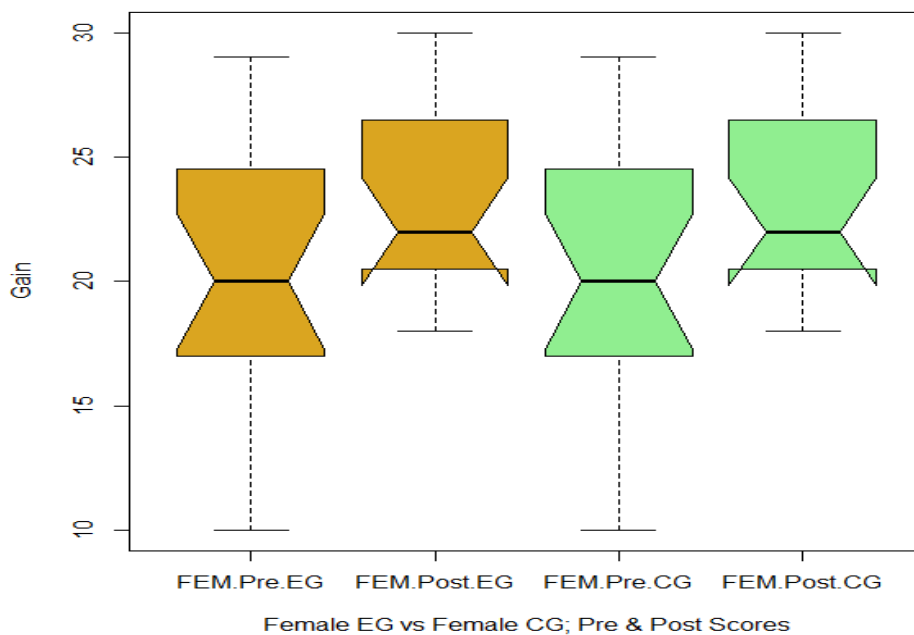
Table 8 EG Filtered by Intervention Effort EG vs CG; Segmented by Gender

For males in the experimental group there was also a statistically significant increase from pretest to posttest of 2.60 additional correct answers, a gain of 12% (2.6 / 21.69) that suggests the intervention also had an effect for males. (Plot 5) For males in the comparison group there was also a significant, although smaller, increase of 6.9% (1.51 / 21.96) from pretest to posttest, again likely due to practice effect and the semester's coursework. The increase for males in the experimental group was 1.72 times greater than that of the comparison group suggesting that males also responded to the intervention software. The average time spent by males on intervention work was 107 minutes (Table 4) suggesting that even a modest amount of time using the intervention software proved beneficial.

A two-way Anova procedure between both study groups and both genders on the pretest score, posttest score, and the gain was performed. There was a significant main effect of gender shown to exist in both the pretest scores, $F(1, 253) = 5.7249$, $p = 0.01746$ and posttest scores $F(1, 253) = 6.6731$, $p = 0.01035$. There was a nonsignificant main effect of the software intervention shown to exist on both the pretest scores, $F(1, 253) = 0.0687$, $p = 0.79352$ and posttest scores $F(1, 253) = 3.8713$, $p = 0.05021$. There was a nonsignificant interaction between the study

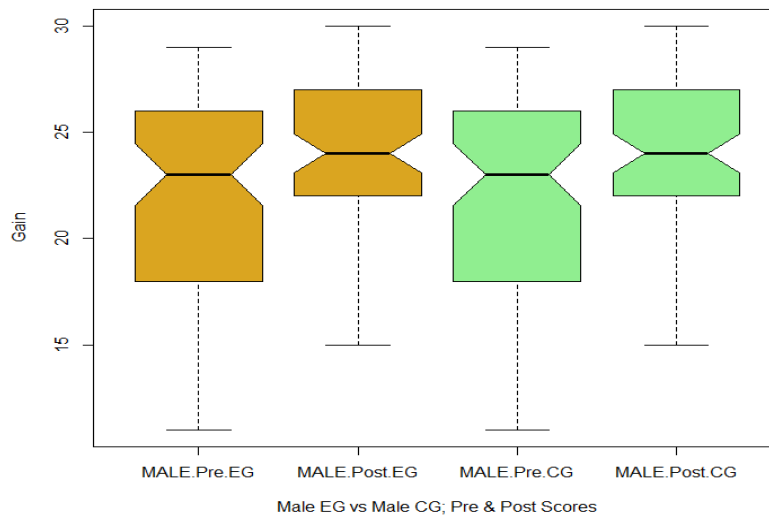
groups and both genders on the pretest scores, $F(1, 253) = 0.2067$, $p = 0.64974$, and the posttest scores, $F(1, 253) = 1.2123$, $p = 0.27193$. These results suggest that based on the means there was not a significant difference between the experiment study group and the comparison study group when they took the pretest. Further it suggests that a significant difference existed between the mean scores for males and for females when they took both the pretest and the posttest.

There was a significant main effect of the intervention shown to exist on the gain, $F(1, 253) = 7.7376$, $p = 0.005816$. There was a nonsignificant effect of gender on the gain, $F(1, 253) = 0.0441$, $p = 0.833771$ and there was a nonsignificant interaction between the study groups and both genders on the gain, $F(1, 253) = 0.5013$, $p = 0.479592$. This suggests that the mean gains for male students was not significantly different from the mean gain for female students, but the difference between main gains for the two study groups was significant. Furthermore, it suggests that the intervention had a significant effect on the means of the gain for both males and females.



Plot 4 Female EG vs Female CG Pre & Post Scores

An effect size (Cohen's *d*) of 0.561 was computed for the gain between the experimental and comparison groups for females in the population filtered by intervention. With a Cohen's *d* of 0.56, 71% of the experimental group will be above the mean of the comparison group, 78% of the two groups will overlap, and there is a 65% chance that a person picked at random from the experimental group will have a higher score than a person picked at random from the comparison group.



Plot 5 Male EG vs Male CG Pre & Post Scores

Finally, an effect (Cohen's *d*) size of 0.313 was computed for the gain between the experimental and comparison groups for males in the population filtered by intervention. With a Cohen's *d* of 0.31, 62% of the experimental group will be above the mean of the comparison group, 88% of the two groups will overlap, and there is a 59% chance that a person picked at random from the experimental group will have a higher score than a person picked at random from the comparison group.

5.3.6.1 Intervention Effort Further Filtered by Failed Pretest

The preceding results were for the entire range of students' test performance, including those students that performed well on the pretest and had little or no room for improvement. Three students had perfect scores on the pretest and had no room for improvement, twelve students had only 1 incorrect response, seven students had 2 incorrect responses, and 21 students had only 3 incorrect responses, leaving little room for gain regardless of the amount or quality of the intervention performed. Additionally, in a longitudinal study covering 15 years at Michigan Technological University it was demonstrated that enrolling students who initially failed the

PSVT:R test in a spatial skills training course showed significant improvement in grades in STEM courses and retention (Veurink & Sorby, 2011). Curricular requirements at the institution where the research was conducted prevented implementation of the training only for students who failed the pretest; however, the following section examines the effect of the intervention for those students who had weak spatial skills initially.

Seventy students in the study “failed” the pretest, where a failure is considered to be a score of 60% or less on the test which translates to a score of 18 or below on the pretest. (27.2% failed, 72.8% passed) Of those who failed, 26 students comprise the experimental group and received an average of 126.6 minutes of intervention and performed an average of 216.6 intervention challenge problems. Forty-four students who failed and received no intervention comprise the comparison group. Of the 44 students in the comparison group 22 students (50%) who failed the pretest also failed the posttest and 22 students (50%) subsequently passed the posttest. Of the 26 students in the experimental group 21 students (80.8%) who failed the pretest subsequently passed the posttest and 5 students (19.2%) failed again on the posttest.

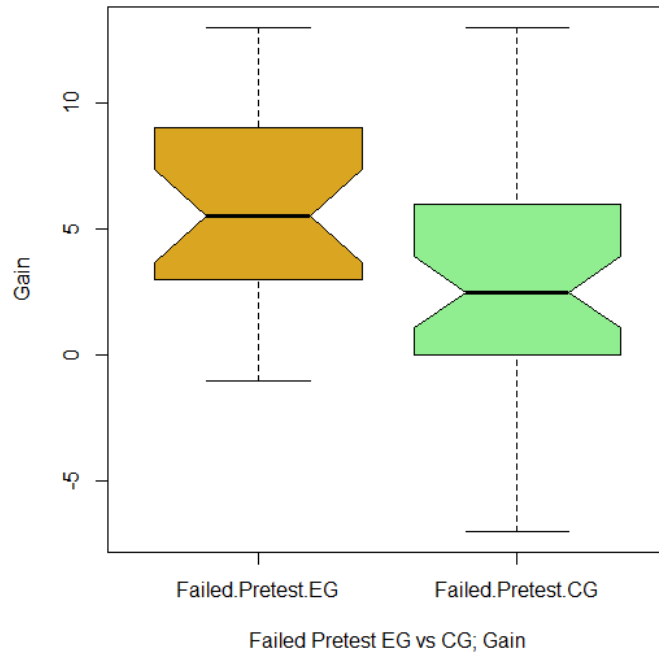
To determine the effect of the intervention on students that failed the pretest, the dataset that had been filtered by intervention effort (N=257) was additionally filtered by pretest scores of less than or equal to 18 reducing the dataset to 70.

The same five t-tests that were performed on the previous datasets were performed on the dataset filtered by failure on the pretest. (Table 9) The experimental group showed a statistically significant improvement of 6.08 questions correct from pretest to posttest. This is 3.36 higher than the improvement shown with the population filtered only by intervention effort (6.08 – 2.72). The difference in gain between the experimental group and the comparison group was -2.94, indicating that the comparison group gained 2.94 fewer points than the

Population 1				Population 2								Degrees of Freedom	
Comparison Value		Mean	Std. Error	Comparison Value		Mean	Std. Error	Delta	CI	CI	T		P-Value
EG	Pretest Correct	14.92	0.51	EG	Posttest Correct	21.00	0.61	6.08	-7.605	-4.549	-8.190	25.00	0.0000000153
CG	Pretest Correct	15.14	0.45	CG	Posttest Correct	18.27	0.73	3.14	-4.390	-1.883	-5.045	43.00	0.0000087300
EG	Pretest Correct	14.92	0.51	CG	Pretest Correct	15.14	0.45	0.21	-1.575	1.148	-0.314	58.84	0.7550000000
EG	Posttest Correct	21.00	0.61	CG	Posttest Correct	18.27	0.73	-2.73	0.838	4.617	2.881	67.51	0.0053090000
EG	Gain	6.08	0.74	CG	Gain	3.14	0.62	-2.94	1.002	4.879	3.038	56.29	0.0036090000

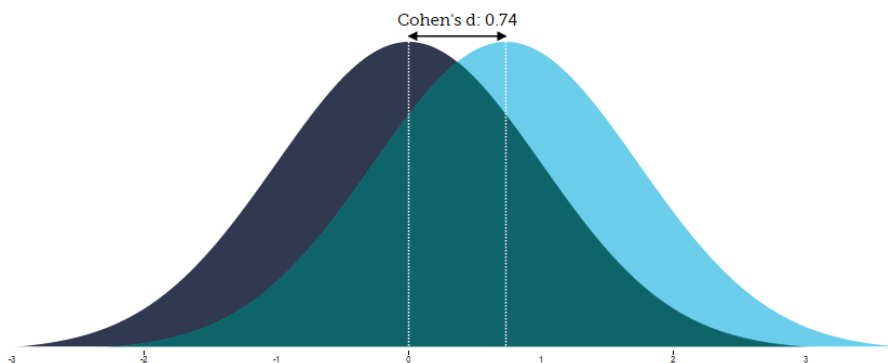
Table 9 EG & CG Filtered by Failure on Pretest

experimental group. The experimental group had a gain of 40.8% (6.08 / 14.92) and the comparison had a gain of 20.7% (3.14 / 15.14). (Plot 6) As in the tests filtered only by intervention effort, the pretests for both groups showed no significant difference in the mean of the number of correct answers indicating that both groups started from statistically similar means.



Plot 6 Failed Pretest EG Gain vs CG Gain

An effect size (Cohen's d) of 0.735 was computed for the gain between the experimental and comparison groups for students who failed the pretest with a score of 18 or fewer in the population filtered by intervention effort. (Plot 7) With a Cohen's d of 0.74, 77% of the



Plot 7 Intervention Effect Size for Student's that Failed the Pretest (≤ 18)

experimental group will be above the mean of the comparison group, 71% of the two groups will overlap, and there is a 70% chance that a person picked at random from the experimental group will have a higher score than a person picked at random from the comparison group.

In the Veurink & Sorby (2011) longitudinal study executed each year from 1993 to 1998 an average overall gain of 26.4% was achieved for students who failed the pretest of the PSVT:R and received the intervention course. The course was 10 weeks long and had 2 hours of lecture and 2 hours of computer lab work each week that used CAD-based computer exercises for a total of 40 hour of intervention time. In that study it does not report how many of the 40 hours of intervention were spent specifically on mental rotations. A full outline of the course shows that there are many spatial skills being taught, sketching isometric pictorials and multi-view drawings of simple objects are taught and reported as, “greatly enhancing spatial skills.” In a more recent study by Duffy *et al.* (2015) that used the same intervention training material developed by Sorby *et al.* and was administered for 2 hours a week for 6 weeks, a 17% gain was reported. It discusses that the delivery of the intervention implementation was “rushed, [and] poorly attended” and had “novice tutors” and was of an overall “poorer quality” than the course presented by Sorby, *et al.* In the current study, for an intervention time of roughly 2 hours using the software-only intervention a gain of 20.3% was achieved.

5.3.6.2 Intervention Effort Further Filtered by Failed Pretest & Segmented by Gender

In order to determine if the intervention had a significant effect across gender within the failed pretest population the data was further segmented. The population remained the same (N=70) but the data was further broken down into subsets by gender giving four groups for comparison: the experimental group for males (N=20) and for females (N=6), and the comparison group for males (N=31) and for females (N=13). Of all males in the sample population filtered for intervention effort, 24.4% failed the pretest. Of all females in the sample population filtered for intervention effort 39.6% failed the pretest, demonstrating that females started the study with weaker mental rotation spatial skills. The same battery of t-tests as performed for the previous data sets was performed for each gender yielding comparisons within each gender. (Table 10) Of those in the experimental group the average amount of intervention time was 116.5 minutes for females and 129.7 minutes for males. Additionally, to compare the means across both gender and study group, a two-way Anova was performed for both factors.

For females in the experimental group there was a statistically significant increase from pretest to posttest of 6.33 additional correct answers, a gain of 43.1% (6.33 / 14.67) that suggests that

Population 1		Comparison Value		Mean		Std. Error		Population 2		Comparison Value		Mean		Std. Error		Delta		CI		CI		T		Degrees of Freedom		P-Value	
FEMALES																											
EG	Pretest Correct	14.67	1.15	EG	Posttest Correct	21.00	0.93	6.33	-10.297	-2.369	-4.107	5.00	0.0092890000														
CG	Pretest Correct	15.92	0.62	CG	Posttest Correct	18.31	0.95	2.38	-4.256	-0.513	-2.776	12.00	0.0167600000														
EG	Pretest Correct	14.67	1.15	CG	Pretest Correct	15.92	0.62	1.26	-4.257	1.744	-0.963	8.12	0.3633000000														
EG	Posttest Correct	21.00	0.93	CG	Posttest Correct	18.31	0.95	-2.69	-0.154	5.538	2.024	14.35	0.0619800000														
EG	Gain	6.33	1.54	CG	Gain	2.38	0.86	-3.95	-0.100	7.997	2.237	8.25	0.0546800000														
MALES																											
EG	Pretest Correct	15.00	0.58	EG	Posttest Correct	21.00	0.75	6.00	-7.816	-4.184	-6.916	19.00	0.0000013560														
CG	Pretest Correct	14.81	0.58	CG	Posttest Correct	18.26	0.96	3.45	-5.100	-1.803	-4.277	30.00	0.0001774000														
EG	Pretest Correct	15.00	0.58	CG	Pretest Correct	14.00	0.58	-1.00	-1.460	1.847	0.236	46.58	0.8148000000														
EG	Posttest Correct	21.00	0.75	CG	Posttest Correct	18.26	0.96	-2.74	0.293	5.191	2.250	48.97	0.0289600000														
EG	Gain	6.00	0.87	CG	Gain	3.45	0.81	-2.55	0.162	4.935	2.151	44.84	0.0369200000														

Table 10 EG vs CG; Males & Females

the intervention had an effect. For females in the comparison group there was also a significant, although smaller, increase of 14.9% (2.38 / 15.92) from pretest to posttest, again likely due to practice effect and the semester's coursework. The increase for females in the experimental group was 2.61 times greater than that of the comparison group suggesting that females responded well to the intervention software. The average time spent by females on intervention work was 116.5 minutes, suggesting that even a modest amount of time using the intervention software proved beneficial.

For males in the experimental group there was also a statistically significant increase from pretest to posttest of 6.0 additional correct answers, a gain of 40% (6.0 / 15.0) that suggests the intervention also had an effect for males. For males in the comparison group there was also a significant, although smaller, increase of 23.3% (3.45 / 14.81) from pretest to posttest, again likely due to practice effect and the semester's coursework. The increase for males in the experimental group was 1.74 times greater than that of the comparison group suggesting that males also responded to the intervention software. The average time spent by males on intervention work was 129.7 minutes, suggesting that even a modest amount of time using the intervention software proved beneficial.

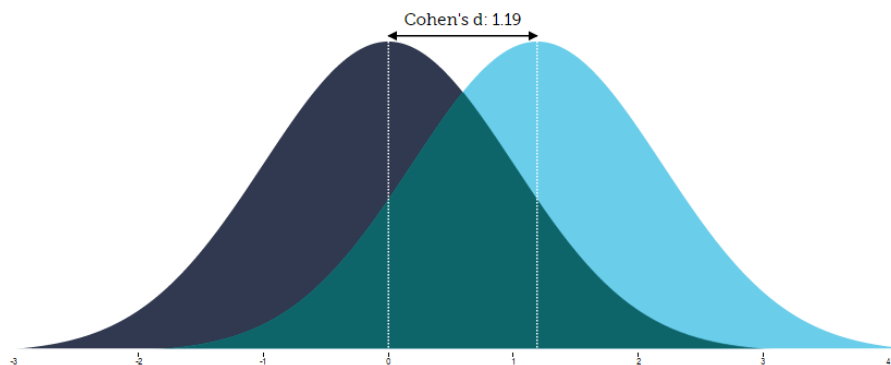
A two-way Anova procedure between both study groups and both genders on the pretest score, posttest score, and the gain was performed. There was a nonsignificant main effect of gender shown to exist in the pretest scores, $F(1, 66) = 0.6661$, $p = 0.4174$ and posttest scores $F(1, 66) = 0.0008$, $p = 0.97755$.

There was a nonsignificant main effect of the software intervention shown to exist on the pretest scores, $F(1, 66) = 0.0903$, $p = 0.7647$ and a significant main effect of the software intervention shown to exist on the posttest scores $F(1, 66) = 6.4664$, $p = 0.01335$.

There was a nonsignificant interaction between the study groups and both genders on the pretest scores, $F(1, 66) = 0.7841$, $p = 0.3791$, and the posttest scores, $F(1, 66) = 0.0004$, $p = 0.98406$. These results suggest that based on the means there was a significant difference between the experiment study group and the comparison study group when they took the posttest but a nonsignificant difference when they took the pretest, implying that the intervention, more so than gender, had an effect. Further it suggests that no significant difference existed between the mean scores for males and for females when they took both the pretest and the posttest.

There was a significant main effect of the intervention shown to exist on the gain, $F(1, 66) = 8.6512$, $p = 0.004504$. There was a nonsignificant interaction of gender on the gain, $F(1, 66) = 0.3013$, $p = 0.584893$ and there was a nonsignificant interaction between the study groups and both genders on the gain, $F(1, 66) = 0.3684$, $p = 0.545952$. This suggests that the mean gain between the genders were not significantly different, but the mean gain between the study groups were significant. Furthermore, it suggests that the intervention had a significant effect on the means of the gain for both males and females.

An effect size (Cohen's d) of 1.192 was computed for the gain between the experimental and comparison groups for females in the population filtered by intervention effort. With a Cohen's d of 1.19, 88% of the experimental group will be above the mean of the comparison group, 55% of the two groups will overlap, and there is an 80% chance that a person picked at random from the experimental group will have a higher score than a person picked at random from the comparison group. (Plot 8)



Plot 8 Effect Size; Females who Failed Pretest

Finally, an effect size (Cohen's d) of 0.597 was computed for the gain between the experimental and comparison groups for males in the population filtered by intervention effort. With a Cohen's d of 0.59, 72% of the experimental group will be above the mean of the comparison group, 77% of the two groups will overlap, and there is a 66 % chance that a person picked at random from the experimental group will have a higher score than a person picked at random from the comparison group.

5.4 Conclusion

Gains for the total population were 1.6 times greater for students in the experimental group than those in the comparison group. For the subset of students who made at least a minimal effort in the intervention, the gain was 1.85 when in the experimental group. Within that subset females gained 2.45 times more and males 1.72 times more when in the experimental group. Students that failed the pretest had 1.94 times a greater gain if they were in the experimental group. Female students who failed the pretest gained 2.66 times more if they were in the experimental group rather than the comparison group and males gained 1.74 times more when in the experimental group. When compared against males in the same segment of data, females gained 1.16 and 1.25 times more than their male counterparts.

With the exception of data that was segmented by gender, all datasets evaluated demonstrated a significant improvement in the mean gain for the experimental group over the comparison group. In the two data subsets segmented by gender, male student gains were all statistically significant (95% confidence interval, $N = 209$, p -value = 0.03637 & $N = 51$, p -value = 0.03692), however female gains, which were even higher, fell just short of being statistically significant (95% confidence interval, $N = 48$, p -value = 0.07806 & $N = 19$, p -value = 0.05468). For the data subset filtered by students who failed the pretest, there were only 19 students included, with only 6 in the experimental group, which may have contributed to the result being

nonsignificant. For both of the female subsets of data analyzed, it should be noted that the degree of statistical significance was just short of the 95% confidence level.

As suggested by Cohen (1977), standard effect sizes of 0.2, 0.5, and 0.8 are small, moderate and large in terms of practical significance. Small effect sizes were noted for the total population and the population filtered by intervention effort. The effect size for males in that population was also small, but it was moderate for females. When the population was further filtered for student's that failed the pretest, effect sizes for gains were moderate for the genders in aggregate, moderate for males alone, and large for females. Although the effect of the intervention for the larger populations examined were small, they were still significant. It is suspected that the short, targeted nature of the intervention is the reason for the small effect size, and that a longer intervention would increase effect size for the total population. Despite the length of the intervention, moderate to large effects were found for females, especially those who failed the pretest. Males who failed the pretest showed moderate effects as well. In the past, other spatial skills interventions, especially those by Sorby (2009b), have focused on students who failed the pretest as they also had the largest effects.

Chapter 6 Conclusions and Future Work

6.1 Introduction

This study examined how software can be used to teach a particular type of spatial ability called mental rotation which is closely linked to success in STEM. It studied the effect of a brief, direct online software intervention on the mental rotation abilities of first year STEM students at a higher education institution. An experiment using a PPC methodology, with experimental and comparison groups, to test students' spatial skills using mental rotation was designed, implementing the Revised PSVT:R standardized test. It created an online spatial rotations testing system, developed a software-based intervention teaching module that taught mental rotations through direct practice and challenge modules, and created a software-based infrastructure to collect, administer, and evaluate data related to student testing, performance and improvement. This project contributes to the body of knowledge about whether software-based interventions can be used to help teach spatial skills to students by demonstrating an increase in student spatial skills. It made original contributions to the field of STEM education by creating a purely software-based, online only, spatial testing and training system. Unlike other research in this domain, the training intervention required no human instructors, students worked independently and were able to train on the system on their own schedule and to the level they desired. The approach taken was different from other spatial skills interventions by designing an intervention that focused primarily on a Gestalt-based, holistic approach to teaching mental rotation skills. It was noticed that existing training intervention focused primarily on more analytical approaches to teaching rotation, that they required instructor involvement, were performed over a much longer period of time and could not be done at an individual's preferred pace and level. Additionally, the improvements shown by students in mental rotation were achieved with a much shorter intervention training period – with admittedly lower gain, than in other research, but with higher gain for the time invested in the intervention.

6.2 Findings

The main finding of this study is that a software-based only, spatial skills intervention can have a statistically significant impact on students' spatial skills scores for mental rotation. Mental rotation has been shown to be the most significant spatial skill linked with success in STEM fields. This study analyzed the data using the total sample population, as well as using several select subsets that filtered or segmented the data in order to make meaningful assessments of the effect of the intervention. With the exception of data that was segmented by gender, all

datasets evaluated demonstrated a significant improvement in the mean gain for the experimental group over the comparison group.

This study demonstrates that software-based only interventions, with no other form of intervention can significantly increase students' spatial skills. One implication of this work is that it could be a more efficient way of increasing spatial skills. A computer-based intervention can reach more students and provide more avenues to improvement with less required manpower and faculty involvement done with a greater ability for students to work independently. This may be particularly significant in STEM degrees at higher level institutions where there are often crowded curricula and time-constrained faculty with little room to expand to include additional training in spatial skills.

One area in which this work departs from other spatial skills interventions is in its attempt to teaching spatial skills by encouraging students to use a more holistic rotation approach. Rather than encourage students to use a strategy for rotating an object that requires an axis-by-axis approach using traditional Cartesian axis, students are asked to perform rotations free of the Cartesian axis. When demonstrating rotations to students in the challenges, the intervention software demonstrates how the object can be rotated in a single step around one axis.

When comparing female gains against male gain in both the experimental and comparison study groups for the subset of students that made at least a minimal effort in the intervention, a statistically significant difference was found between the two study groups; however, a nonsignificant difference was found between genders. The same outcome was found when comparing female gain against male gain for each study group for those who failed the pretest. Before the intervention, a statistically significant gender gap existed between males and females on the pretest scoring. After the intervention the gender gap remained significant on posttest scoring, with both genders achieving mean gains that were not statistically different from each other.

6.3 Research Objects

Research objects for this study were met by demonstrating that even a short, targeted online intervention to teach mental rotation spatial skills to first year STEM students can have a significant, positive impact on students' mental rotation abilities. It demonstrated that the effect is significant for both males and females, but was unable to demonstrate a statistically significant difference in the increases for females over males.

6.4 Limitations

The study was not without limitations in terms of implementation and logistics. One limit of the intervention software is that it used as the primary rotation device a typical mouse pointing device which maps to a two-dimensional space to rotate objects in a three-dimensional space. Anecdotal evidence from students suggest that this method of rotating objects may not be optimal because it proved to be tricky to master for some people. Because of this, the intervention software had to include some input methods that were more analytical, i.e. they were concerned with rotating around traditional Cartesian axes in several steps rather than the true axis of rotation in a single step. An implication of this limitation is that perhaps the intervention taught rotation in a less holistic manner than was intended for those students. This is something that the experiment as it was designed was not able to evaluate. The logistics of running this experiment led to some limitations in how students approached taking the intervention and perhaps even the testing. Because the study dealt with 16 sections of freshman engineering that was taught by 12 different instructors it was difficult to control and assess the amount of significance each instructor placed on participation in the study. Some instructors assigned the intervention as part of the course grade others did not; some used class time to ensure each student worked with the intervention software, others did not. A student who is not properly incentivized to take the intervention or the tests seriously enough, can affect the overall results. This effect of student apathy in taking an extracurricular intervention has been noted in other spatial skills studies (Duffy et al., 2015). In this study, there was no way to specifically isolate this behavior from the data collected, but there were indications that it occurred to some extent. It is difficult to say with certainty to what extent students in this study tried to execute the tests and intervention to the best of their abilities. One factor that may have caused some nonchalance on the part of students is that their pretest scores were reported to them immediately after taking the test. Students that performed very well might not have seen the point of the intervention; likewise students who were aware of the fact that they performed poorly might have been discouraged and become disinterested in the intervention work.

It is more difficult to monitor students when they are given free range over their own intervention. It was recognized that the timing data for student intervention work may not accurately reflect the amount of intervention work performed because of certain unavoidable limitations of monitoring software usage, e.g. a student might decide to multitask by answering a newly arrived email or simultaneously be watching a video in another window. For this reason the intervention tracking data had to be filtered because student behavior could not be

ascertained with certainty. An intervention performed in a classroom where student effort could be monitored by an instructor likely would not suffer from the same difficulty.

A final point that came up anecdotally from several test takers is that they accidentally were trying to solve the rotation analogies, in both the PSVT:R tests and in the analogy intervention sections, by mentally rotating the object presented on the right to the position of the object presented on the left. This is an incorrect approach; however, it will yield a correct response under certain specific rotations, e.g. if the object is rotated 180° around a single axis. The correct approach to solving the analogies presented in the intervention and on the PSVT:R test is to rotate the object presented on the left to the position of the object presented on the right. Because mental rotation analogies are not always commutative in nature which led to confusion (the expected answer is not one of the choices present) or mistakes being made (the wrong answer is selected) despite correct mental rotations (from the right side object to the left side object) being performed.

6.5 Future Work

This study has looked at first year students in an engineering program at a higher learning institution and attempted to train students and to improve their ability to mentally rotate objects in a three-dimensional space. While it has concluded that a software-based only spatial skills intervention can help students improve, there are many avenues of exploration closely related to this work that can be explored. The concept of teaching spatial skills using a more holistic and Gestalt-based method was of significant concern to this study, but it may have fallen short of achieving that intention. The study would be improved had the pointing devices available for student's to train with been more suited to that end, for example, a mouse that maps directly to three-dimensional spaces or a device with a tracking ball could be used to make it easier for students to think in a holistic manner as they train. Other input devices such as virtual reality rooms could be used to more closely simulate real three-dimensional experiences, or other less esoteric technology that is more available such as smart phones which include accelerometers and gyroscopes could be used as input devices to make it easier for students to experience rotating objects using just a single axis. It was not uncommon to see students who were taking the PSVT:R test using their hands to simulate rotation. It was clear that one strategy used by students to solve rotation problems involved imagining that they had the object to be rotated in their hand and mentally mapping the rotation of their hand to the solution of the problem. If students who used this strategy could hold a smart phone in their hand as they performed this mapping, they might be able to more easily understand, practice and solve rotation problems.

In general, if students were given better options for which input device they train with during the intervention, it would be possible to remove some of the input methods currently used in the software, such as sliders, buttons, and the keyboard that all work by mapping directly to Cartesian axes, hopefully improving students' approaches to solving rotation problems.

There are many ways in which the experiment in this study could be altered to provide different insights in how spatial skills can be developed in students. Several students noted after taking the PSVT:R test that two-dimensional, isometric representations of the images in the test were not as clear to them as they might have been had they been shaded to more closely represent how we naturally see objects, i.e. with light reflection and color. The intervention software in this study used difficulty levels in the challenges where the color and shading of rotation objects were removed as the students advanced through the levels. The PSVT:R test is a test of rotation, but clearly, a skill that is required to perform well on the test is to understand the three-dimensional shape that is being represented by two-dimensional isometric images. It would be an interesting direction of study if the experiment was run with a modified PSVT:R that used shapes that might be more easily processed, i.e. with more realistic shading as studied by Yue (2009). Another improvement to the experiment would involve figuring out a method to more accurately track the students' use of the intervention software and account for situations that as is prevented a meaningful analysis of the data based on the amount of time students spent doing intervention work, such as when a student left their computer for long period of time in the middle of an intervention challenge.

The timing data collected on students in this study showed that many students completed the tests far more quickly than the allotted 20 minutes given. It raised the question as to whether the test is clearly differentiating between students who are holistic mental rotators vs. those who are using other analytical skills that do not rely as much on spatial skills intelligence. Future studies could run this experiment with a much shorter time allotment for the test in an attempt to make this differentiation.

Effect sizes were small to large depending on the segment of the population examined. Future work should examine if a longer intervention would improve the practical significance of the intervention.

While the online version of the Revised PSVT:R used for this experiment showed internal consistency using Cronbach's α that was similar to that found when the paper version was psychometrically validated, it has not yet been psychometrically validated.

Future work should take steps to ensure students understand the ramifications of not following the left-to-right direction of the analogy problems presented in the intervention.

Finally, for future work, the intervention software should be enhanced to do a better job of tracking student effort in the intervention.

Bibliography

- Alexander, G. M., & Evardone, M. (2008). Blocks and Bodies: Sex Differences in a Novel Version of the Mental Rotations Test. *Hormones and Behavior*, *53*(1), 177–184.
<http://doi.org/10.1016/j.yhbeh.2007.09.014>
- Bethell-Fox, C. E., & Shepard, R. N. (1988). Mental Rotation: Effects of Stimulus Complexity and Familiarity. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(1), 12.
- Black, A. A. (2005). Spatial Ability and Earth Science Conceptual Understanding. *Journal of Geoscience Education*, *53*(4), 402.
- Bodner, G. M., & Guay, R. B. (1997). The Purdue Visualization of Rotations Test. *The Chemical Educator*, *2*(4), 1–17.
- Branoff, T. (1998). The Effects of Adding Coordinate Axes to a Mental Rotations Task in Measuring Spatial Visualization Ability in Introductory Undergraduate Technical Graphics Courses. *Engineering Design Graphics Journal*, *62*(2), 16–34.
- Bridgeman, B., Lennon, M. L., & Jackenthal, A. (2003). Effects of Screen Size, Screen Resolution, and Display Rate on Computer-Based Test Performance. *Applied Measurement in Education*, *16*(3), 191–205.
http://doi.org/10.1207/S15324818AME1603_2
- Burke, R. J., & Mattis, M. C. (2007). *Women and Minorities in Science, Technology, Engineering, and Mathematics: Upping the Numbers*. Edward Elgar Publishing.
- Caplan, P. J., MacPherson, G. M., & Tobin, P. (1985). Do Sex-Related Differences in Spatial Abilities Exist? A Multilevel Critique with New Data. *American Psychologist*, *40*(7), 786–799. <http://doi.org/10.1037/0003-066X.40.7.786>
- Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*. Academic Press.

- Connor, J. M., Schackman, M., & Serbin, L. A. (1978). Sex-Related Differences in Response to Practice on a Visual-Spatial Test and Generalization to a Related Test. *Child Development, 49*(1), 24–29. <http://doi.org/10.1111/1467-8624.ep10400905>
- Deno, J. A. (1995). The Relationship of Previous Experiences to Spatial Visualization Ability. *Engineering Design Graphics Journal, 59*(3), 5–17.
- Duffy, G., Farrell, S., Harding, R., Behan, A., Mac Raighne, A., Howard, R., ... Bowe, B. (2015). The Effects of Spatial Skills and Spatial Skills Training on Academic Performance in STEM Education. In *Proceedings of the 2015 REES*. Dublin, Ireland.
- Gardner, H. (1984). *Frames of Mind: The Theory of Multiple Intelligences* (2nd Edition). Fontana Press.
- Geiser, C., Lehmann, W., & Eid, M. (2006). Separating “Rotators” from “Nonrotators” in the Mental Rotations Test: A Multigroup Latent Class Analysis. *Multivariate Behavioral Research, 41*(3), 261–293.
- Gerson, H. B., Sorby, S. A., Wysocki, A., & Baartmans, B. J. (2001). The Development and Assessment of Multimedia Software for Improving 3-D Spatial Visualization Skills. *Computer Applications in Engineering Education, 9*(2), 105–113.
- Guay, R. B. (1980). Spatial Ability Measurement: A Critique and an Alternative.
- Halpern, D. F. (1986). A Different Answer to the Question, “Do Sex-Related Differences in Spatial Abilities Exist?” *American Psychologist, 41*(9), 1014–1015.
<http://doi.org/10.1037/0003-066X.41.9.1014>
- Harris, J., Hirsh-Pasek, K., & Newcombe, N. S. (2013). Understanding spatial transformations: similarities and differences between mental rotation and mental folding. *Cognitive Processing, 14*(2), 105–115.

- Ho, C.-H., Eastman, C., & Catrambone, R. (2006). An Investigation of 2D and 3D Spatial and Mathematical Abilities. *Design Studies*, 27(4), 505–524.
<http://doi.org/10.1016/j.destud.2005.11.007>
- 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.
<http://doi.org/10.1002/j.2168-9830.1997.tb00278.x>
- Humphreys, L. G., Lubinski, D., & Yao, G. (1993). Utility of Predicting Group Membership and the Role of Spatial Visualization in Becoming an Engineer, Physical Scientist, or Artist. *Journal of Applied Psychology*, 78(2), 250.
- Just, M. A., & Carpenter, P. A. (1985). Cognitive Coordinate Systems: Accounts of Mental Rotation and Individual Differences in Spatial Ability. *Psychological Review*, 92(2), 137.
- Kail, R., Carter, P., & Pellegrino, J. (1979). The Locus of Sex Differences in Spatial Ability. *Perception & Psychophysics*, 26(3), 182–186. <http://doi.org/10.3758/BF03199867>
- Kline, P. (1986). *A Handbook of Test Construction: Introduction to Psychometric Design*. Methuen.
- Kozhevnikov, M. (2012). An Overview of Tests of Cognitive Spatial Ability. Presented at the 66th EDGD Mid-Year Conference Proceedings 97.
- Leeson, H. V. (2006). The Mode Effect: A Literature Review of Human and Technological Issues in Computerized Testing. *International Journal of Testing*, 6(1), 1–24.
http://doi.org/10.1207/s15327574ijt0601_1
- Leopold, C., Gorska, R. A., & Sorby, S. A. (2001). International Experiences in Developing the Spatial Visualization Abilities of Engineering Students. *Journal for Geometry and Graphics*, 5(1), 81–91.

- Linn, M. C., & Petersen, A. C. (1985). Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Development*, 1479–1498.
- Lubinski, D. (2010). Spatial Ability and STEM: A Sleeping Giant for Talent Identification and Development. *Personality and Individual Differences*, 49(4), 344–351.
<http://doi.org/10.1016/j.paid.2010.03.022>
- Maeda, Y., & Yoon, S. Y. (2013). A Meta-Analysis on Gender Differences in Mental Rotation Ability Measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT: R). *Educational Psychology Review*, 25(1), 69–94.
- Maeda, Y., Yoon, S. Y., Kim-Kang, G., & Imbrie, P. K. (2013). Psychometric Properties of the Revised PSVT: R for Measuring First Year Engineering Students' Spatial Ability. *International Journal of Engineering Education*, 29(3), 763–776.
- Morris, S. B. (2008). Estimating Effect Sizes from Pretest-Posttest-Control Group Designs. *Organizational Research Methods*, 11(2), 364–386.
<http://doi.org/10.1177/1094428106291059>
- Newcombe, N. S. (2010). Picture This: Increasing Math and Science Learning by Improving Spatial Thinking. *American Educator*, 34(2), 29.
- Parsons, T. (2004). Sex Differences in Mental Rotation and Spatial Rotation in a Virtual Environment. *Neuropsychologia*, 42(4), 555–562.
<http://doi.org/10.1016/j.neuropsychologia.2003.08.014>
- Pedrosa, C. M., Barbero, B. R., & Miguel, A. R. (2014). Spatial Visualization Learning in Engineering: Traditional Methods vs. a Web-Based Tool. *Journal of Educational Technology & Society*, 17(2), 142–157.
- Sorby, S. A. (2006). Developing 3-D Spatial Skills for Engineering Students. In *Proceedings of the 17th Annual Conference of the Australasian Association for Engineering*

- Education: Creativity, Challenge, Change; Partnerships in Engineering Education* (p. 679). Australasian Association for Engineering Education.
- Sorby, S. A. (2009a). Assessment of a “New and Improved” Course for the Development of 3-D Spatial Skills. *Engineering Design Graphics Journal*, 69(3).
- Sorby, S. A. (2009b). Educational Research in Developing 3-D Spatial Skills for Engineering Students. *International Journal of Science Education*, 31(3), 459–480.
- Sorby, S. A., & Baartmans, B. J. (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(3), 301–307.
- Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The Role of Spatial Training in Improving Spatial and Calculus Performance in Engineering Students. *Learning and Individual Differences*, 26, 20–29. <http://doi.org/10.1016/j.lindif.2013.03.010>
- Stanley, J. C., & Hopkins, K. D. (1972). *Educational and Psychological Measurement and Evaluation*. Prentice-Hall.
- Tartre, L. A. (1990). Spatial Orientation Skill and Mathematical Problem Solving. *Journal for Research in Mathematics Education*, 21(3), 216–229.
<http://doi.org/10.2307/749375>
- Uttal, D. H., Meadow, N. G., Tipton, E., Hand, L. L., Alden, A. R., Warren, C., & Newcombe, N. S. (2013). The Malleability of Spatial Skills: A Meta-Analysis of Training Studies. *Psychological Bulletin*, 139(2), 352–402.
<http://doi.org/10.1037/a0028446>
- Vandenberg, S. G., & Kuse, A. R. (1978). Mental Rotations, a Group Test of Three-Dimensional Spatial Visualization. *Perceptual and Motor Skills*, 47(2), 599–604.
<http://doi.org/10.2466/pms.1978.47.2.599>

- Veurink, N., & Hamlin, A. J. (2015). Comparison of On-line Versus Paper Spatial Testing Methods. Presented at the 122nd ASEE Annual Conference & Exposition, Seattle, Washington, USA: American Society for Engineering Education.
- Veurink, N. L., & Sorby, S. A. (2011). Raising the Bar? Longitudinal Study to Determine Which Students Would Benefit Most from Spatial Training. In *Proceedings of the ASEE Annual Conference, Vancouver, BC, CD-ROM*.
- Voyer, D. (2010). Time Limits and Gender Differences on Paper-and-Pencil Tests of Mental Rotation: a Meta-Analysis. *Psychonomic Bulletin & Review*, *18*(2), 267–277.
<http://doi.org/10.3758/s13423-010-0042-0>
- 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.
- 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.
- Yoon, S. Y. (2011). *Psychometric Properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (The Revised PSVT-R)*. ProQuest LLC.
- Yoon, S. Y. (2015, March 4). RE: Revised PSVT:R (Dr. Stephanie Farrell, 2015, pers. comm., 4 March).
- Yue, J. (2002). Spatial Visualization Skills at Various Educational Levels. Presented at the Proceedings of the 2002 American Society for Engineering Education Annual Conference & Exposition, American Society for Engineering Education.
- Yue, J. (2004). Spatial Visualization by Orthogonal Rotations. In *Proceedings of the American Society for Engineering Education (ASEE) Annual Conference and Expositions*.

Yue, J. (2006). Spatial Visualization by Isometric Drawing. In *Proceedings of the 2006 IJME-INTERTECH Conference, Union, New Jersey.*

Yue, J. (2009). Spatial Visualization by Realistic 3D Views. *Engineering Design Graphics Journal*, 72(1).

Yue, J., & Chen, D. M. (2001). Does CAD Improve Spatial Visualization Ability? In *Proceedings of the 2001 ASEE Annual Conference & Exposition*. Albuquerque, New Mexico.

Chapter 7 Appendices

7.1 Institutional Review Board (IRB) – Spatial Skills Study Rowan University

** This is an auto-generated email. Please do not reply to this email message. The originating e-mail account is not monitored. If you have questions, please contact your local IRB office **					
DHHS Federal Wide Assurance Identifier: FWA00007111					
IRB Chair Person: Harriet Hartman					
IRB Director: Sreekant Murthy					
Effective Date: 12/10/2015					
Study Expiration Date:					
eIRB Notice of Approval					
STUDY PROFILE					
Study ID:	Pro2015000515				
Title:	Spatial Skills of First Year Engineering Students				
Principal Investigator:	Stephanie Farrell	Study Coordinator:	None		
Co-Investigator(s):	None	Other Study Staff:	Peter Cole		
Sponsor:	Department Funded Internal / Institutional Funding	Approval Cycle:	Not Applicable		
Risk Determination:	Minimal Risk	Device Determination:	Not Applicable		
Review Type:	Exempt	Exempt Category:	1 2		
CURRENT SUBMISSION STATUS					
Submission Type:	Modification (Mod201500000196)	Submission Status:	Approved		
Approval Date:	12/10/2015	Review Type:	Exempt		
Pregnancy Code:	No Pregnant Women as Subjects Not Applicable	Pediatric Code:	Not Applicable No Children As Subjects	Prisoner Code:	Not Applicable No Prisoners As Subjects

7.2 Institutional Review Board (IRB) – Certification

In order to perform Social/Behavioral Research with human subjects using students at Rowan University it was necessary to complete a 14-module certification course online. This was completed as a Rowan University affiliate at: <https://www.citiprogram.org>.

COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM) COURSEWORK REQUIREMENTS REPORT*

* NOTE: Scores on this Requirements Report reflect quiz completions at the time all requirements for the course were met. See list below for details. See separate Transcript Report for more recent quiz scores, including those on optional (supplemental) course elements.

- **Name:** Peter Cole (ID: 5214823)
- **Email:** cole2000@gmail.com
- **Institution Affiliation:** Rowan University (ID: 2336)
- **Institution Unit:** Chemical Engineering
- **Phone:** 2155742003

- **Curriculum Group:** Social & Behavioral Research - Basic/Refresher
- **Course Learner Group:** Group 3: Social, Behavioral and Education Research
- **Stage:** Stage 1 - Basic Course
- **Description:** Choose this group to satisfy CITI training requirements for Investigators and staff involved primarily in Social/Behavioral Research with human subjects.

- **Report ID:** 17916632
- **Completion Date:** 11/15/2015
- **Expiration Date:** 11/14/2018
- **Minimum Passing:** 80
- **Reported Score*:** 94

REQUIRED AND ELECTIVE MODULES ONLY	DATE COMPLETED	SCORE
Belmont Report and CITI Course Introduction (ID: 1127)	11/15/15	3/3 (100%)
History and Ethical Principles - SBE (ID: 490)	11/15/15	5/5 (100%)
Defining Research with Human Subjects - SBE (ID: 491)	11/15/15	5/5 (100%)
The Federal Regulations - SBE (ID: 502)	11/15/15	5/5 (100%)
Assessing Risk - SBE (ID: 503)	11/15/15	5/5 (100%)
Informed Consent - SBE (ID: 504)	11/15/15	5/5 (100%)
Privacy and Confidentiality - SBE (ID: 505)	11/15/15	4/5 (80%)
Research with Prisoners - SBE (ID: 506)	11/15/15	4/5 (80%)
Research with Children - SBE (ID: 507)	11/15/15	5/5 (100%)
Research in Public Elementary and Secondary Schools - SBE (ID: 508)	11/15/15	4/5 (80%)
International Research - SBE (ID: 509)	11/15/15	5/5 (100%)
Internet-Based Research - SBE (ID: 510)	11/15/15	5/5 (100%)
Conflicts of Interest in Research Involving Human Subjects (ID: 488)	11/15/15	4/5 (80%)
Unanticipated Problems and Reporting Requirements in Social and Behavioral Research (ID: 14928)	11/15/15	5/5 (100%)

For this Report to be valid, the learner identified above must have had a valid affiliation with the CITI Program subscribing institution identified above or have been a paid Independent Learner.

CITI Program
 Email: citisupport@miami.edu
 Phone: 305-243-7970
 Web: <https://www.citiprogram.org>

**COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM)
COURSEWORK TRANSCRIPT REPORT****

** NOTE: Scores on this Transcript Report reflect the most current quiz completions, including quizzes on optional (supplemental) elements of the course. See list below for details. See separate Requirements Report for the reported scores at the time all requirements for the course were met.

- **Name:** Peter Cole (ID: 5214823)
- **Email:** cole2000@gmail.com
- **Institution Affiliation:** Rowan University (ID: 2336)
- **Institution Unit:** Chemical Engineering
- **Phone:** 2155742003

- **Curriculum Group:** Social & Behavioral Research - Basic/Refresher
- **Course Learner Group:** Group 3: Social, Behavioral and Education Research
- **Stage:** Stage 1 - Basic Course
- **Description:** Choose this group to satisfy CITI training requirements for Investigators and staff involved primarily in Social/Behavioral Research with human subjects.

- **Report ID:** 17916632
- **Report Date:** 11/15/2015
- **Current Score**:** 94

REQUIRED, ELECTIVE, AND SUPPLEMENTAL MODULES	MOST RECENT	SCORE
History and Ethical Principles - SBE (ID: 490)	11/15/15	5/5 (100%)
Defining Research with Human Subjects - SBE (ID: 491)	11/15/15	5/5 (100%)
Belmont Report and CITI Course Introduction (ID: 1127)	11/15/15	3/3 (100%)
The Federal Regulations - SBE (ID: 502)	11/15/15	5/5 (100%)
Assessing Risk - SBE (ID: 503)	11/15/15	5/5 (100%)
Informed Consent - SBE (ID: 504)	11/15/15	5/5 (100%)
Privacy and Confidentiality - SBE (ID: 505)	11/15/15	4/5 (80%)
Research with Prisoners - SBE (ID: 506)	11/15/15	4/5 (80%)
Research with Children - SBE (ID: 507)	11/15/15	5/5 (100%)
Research in Public Elementary and Secondary Schools - SBE (ID: 508)	11/15/15	4/5 (80%)
International Research - SBE (ID: 509)	11/15/15	5/5 (100%)
Internet-Based Research - SBE (ID: 510)	11/15/15	5/5 (100%)
Unanticipated Problems and Reporting Requirements in Social and Behavioral Research (ID: 14928)	11/15/15	5/5 (100%)
Conflicts of Interest in Research Involving Human Subjects (ID: 488)	11/15/15	4/5 (80%)

For this Report to be valid, the learner identified above must have had a valid affiliation with the CITI Program subscribing institution identified above or have been a paid Independent Learner.

CITI Program
 Email: citisupport@miami.edu
 Phone: 305-243-7970
 Web: <https://www.citiprogram.org>

7.3 Invitation to Participate in Intervention

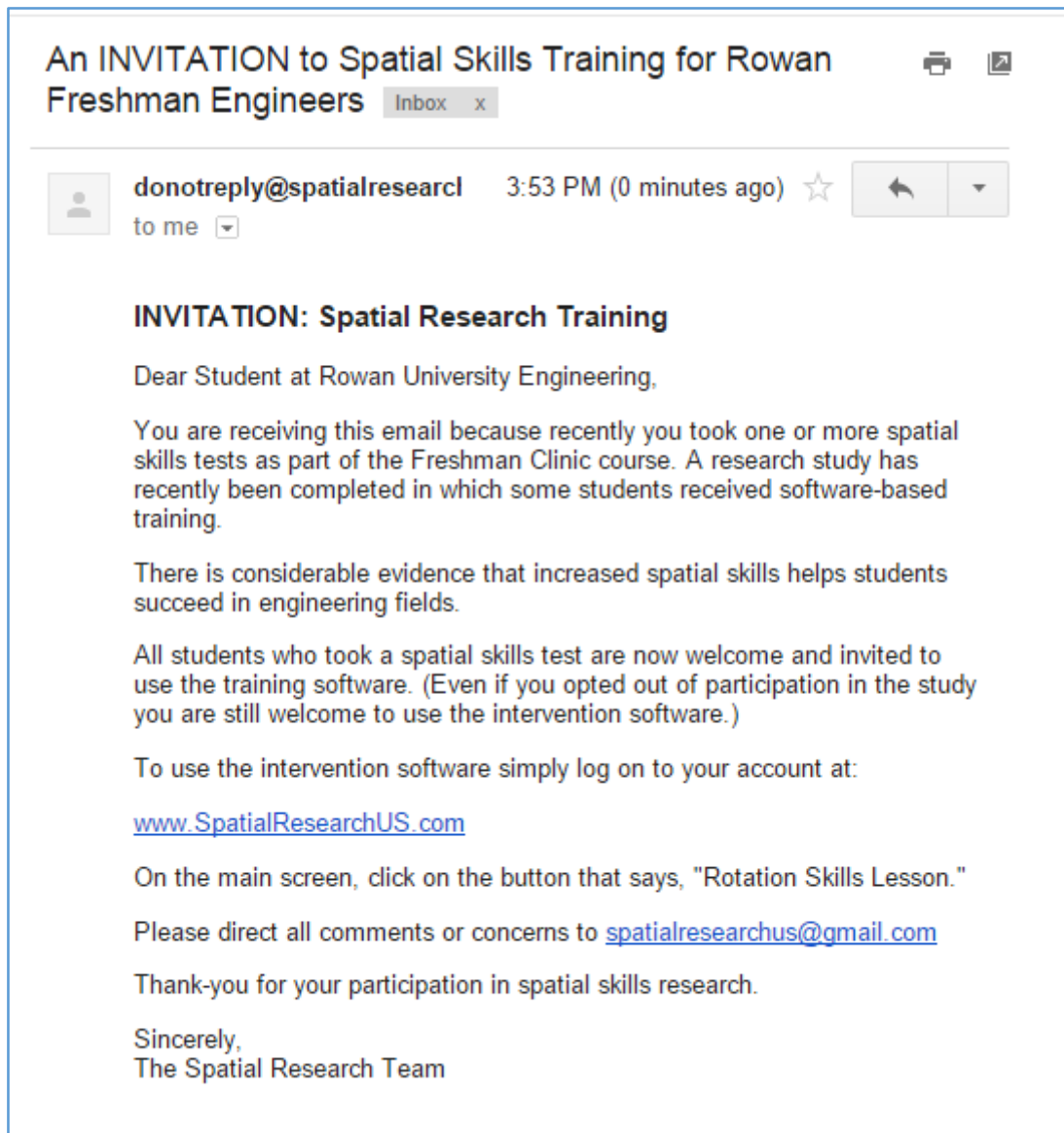


Figure 11 Invitation to All Students to Participate in the Intervention

7.4 Study Participation Agreement



Study Participation Agreement

Invitation: You are invited to participate in a study entitled “Spatial Skills of Engineering Students”. All first year students at Rowan University are invited to participate in this study.

What you will do: This test will take approximately 25 minutes. All first year students will take this spatial skills test as part of Freshman Engineering Clinic. If you choose to participate in the research, your test results may be used in our study.

Purpose: The purpose of the research study is to know more about the spatial skills of engineering students and how they can be improved.

Risks and benefits: There are no risks or benefits associated with participating in the study. However, by participating in this study, you will help us understand more about the spatial skills of engineering students.

Confidentiality: Your test results will be kept confidential. We will store the data in a secure computer file and the file will be destroyed once the results have been published. Your individual information will never be used in any part of the study that is published.

Questions: If you have any questions about the study, you can contact the researcher at the address provided below, but you do not have to give your personal identification.

Dr. Stephanie Farrell | Chemical Engineering | Rowan University | 201 Mullica Hill Rd | Glassboro, NJ 08028
Farrell@rowan.edu | (856) 256 5315

Participation: You may choose whether your results may be used in the research study. Your participation in the study is voluntary. You may indicate your choice by clicking the appropriate button at the bottom of this page. Refusal to participate in the study will not involve a penalty or loss of benefits to which the participant is otherwise entitled. A participant can discontinue participation at any time without penalty or loss of benefits to which the participant is otherwise entitled.

To participate in this study, you must be 18 years or older. Place a check box here if you are 18 or older

Click the appropriate box to indicate whether or not you agree for your results to be used in the study

Yes, my results may be used in the study No, Please do not use my results in the study

Figure 12 Study Participation Agreement Screen Capture from Website

7.5 Questionnaire Questions

7.5.1 Pretest Questionnaire

Do you have prior experience in vocational areas such as carpentry, electrical, plumbing or auto repair? (Yes, No)
How many high school (secondary school) academic courses have you completed in art (drawing or painting)? (0, 1, 2, more than 2)
How many high school (secondary school) academic courses have you completed in art (sculpture)? (0, 1, 2, more than 2)
How many high school (secondary school) academic courses have you completed in chemistry? (0, 1, 2, more than 2)
How many high school (secondary school) academic courses have you completed in computer graphics? (0, 1, 2, more than 2)
How many high school (secondary school) academic courses have you completed in computer programming? (0, 1, 2, more than 2)
How many high school (secondary school) academic courses have you completed in descriptive geometry? (0, 1, 2, more than 2)

How many high school (secondary school) academic courses have you completed in mechanical drawing or drafting? (0, 1, 2, more than 2)
How many high school (secondary school) academic courses have you completed in physics? (0, 1, 2, more than 2)
How often do you engage in assembling electrical circuits? (Frequently, Occasionally, Seldom, Never)
How often do you engage in building models? (Frequently, Occasionally, Seldom, Never)
How often do you engage in building train sets, race car sets, etc.? (Frequently, Occasionally, Seldom, Never)
How often do you engage in carpentry projects? (Frequently, Occasionally, Seldom, Never)
How often do you engage in cooking or baking? (Frequently, Occasionally, Seldom, Never)
How often do you engage in creating computer graphics? (Frequently, Occasionally, Seldom, Never)
How often do you engage in drawing or painting? (Frequently, Occasionally, Seldom, Never)
How often do you engage in dressmaking or sewing using pattern pieces? (Frequently, Occasionally, Seldom, Never)
How often do you engage in driving a car? (Frequently, Occasionally, Seldom, Never)
How often do you engage in playing a musical instrument? (Frequently, Occasionally, Seldom, Never)
How often do you engage in playing strategy board games (e.g. chess, checkers, backgammon)? (Frequently, Occasionally, Seldom, Never)
How often do you engage in playing video games? (Frequently, Occasionally, Seldom, Never)
How often do you engage in playing with building sets? (Frequently, Occasionally, Seldom, Never)
How often do you engage in plumbing projects? (Frequently, Occasionally, Seldom, Never)
How often do you engage in repairing household appliances (radio, stereo, etc.)? (Frequently, Occasionally, Seldom, Never)
How often do you engage in repairing machines (bicycles, mopeds, cars, equipment, etc.)? (Frequently, Occasionally, Seldom, Never)
How often do you engage in riding a moped or motorcycle? (Frequently, Occasionally, Seldom, Never)
How often do you engage in sketching house plans or machine designs (e.g., cars)? (Frequently, Occasionally, Seldom, Never)
How often do you engage in solving 3-Dimensional puzzles similar to Rubiks™ cube? (Frequently, Occasionally, Seldom, Never)
How often do you engage in solving logic puzzles such as Sudoku? (Frequently, Occasionally, Seldom, Never)
How often do you engage in using hand tools? (Frequently, Occasionally, Seldom, Never)

How often do you engage in using power tools? (Frequently, Occasionally, Seldom, Never)
How often do you engage in playing 3-Dimensional action video games? (Frequently, Occasionally, Seldom, Never)
How often do you spend time actively using computer software (but not including watching video or using social media such as Facebook)? (Frequently, Occasionally, Seldom, Never)
How often have you studied academic course materials using computer software? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in ball-and-bat or ball-and-stick team games such as cricket, baseball, hurling, lacrosse, or hockey? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in contact field sports such as football, rugby? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in court sports such as basketball, kabadi, volleyball? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in cross-country, track or jogging? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in cycling? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in dance? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in gymnastics? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in orienteering? (Frequently, Occasionally, Seldom, Never)
In high school (secondary school), how often did you engage in racquet sports such as tennis, ping pong, or badminton? (Frequently, Occasionally, Seldom, Never)
Indicate the degree to which you agree with the following statement: "I prefer to learn academic course material by using computer software before using a classroom setting."(Agree, Somewhat agree, Neither Agree nor Disagree, Somewhat Disagree, Disagree)
Indicate the degree to which you agree with the following statement: "I prefer to learn by reading a book rather than reading online."(Agree, Somewhat agree, Neither Agree nor Disagree, Somewhat Disagree, Disagree)
What best describes your home environment? (Urban, Rural, Suburban)
What is the highest degree your father holds in a technical field? (None, BS - Bachelor of Science or BTech - Bachelor of Technology, MS - Master of Science or MTech - Master of Technology, PhD - Doctor of Philosophy)
What is the highest degree your mother holds in a technical field? (None, BS - Bachelor of Science or BTech - Bachelor of Technology, MS - Master of Science or MTech - Master of Technology, PhD - Doctor of Philosophy)
Which answer best describes <i>a desire to help others</i> as a factor for your deciding to study engineering? (Strong, Moderate, Weak, None)

Which answer best describes *easy to study* as a factor for your deciding to study engineering? (Strong, Moderate, Weak, None)

Which answer best describes *family expectations* as a factor for your deciding to study engineering? (Strong, Moderate, Weak, None)

Which answer best describes *good job opportunities* as a factor for your deciding to study engineering? (Strong, Moderate, Weak, None)

Which answer best describes *personal interest and nonacademic experience* as a factor for your deciding to study engineering? (Strong, Moderate, Weak, None)

Which answer best describes *prior academic coursework* as a factor for your deciding to study engineering? (Strong, Moderate, Weak, None)

7.5.2 Posttest Questionnaire

How would you describe your comfort level at using a browser (Chrome, Firefox, Safari, etc.) to complete tasks on a computer? (Strong, Moderate, Weak)

How often do you use a computer browser in your daily life? (Frequently, Occasionally, Seldom, Never)

7.6 Intervention Full Descriptive Statistics

Intervention Type	Females		Males		Total Population	
	Mean	Std. Dev. (δ)	Mean	Std. Dev. (δ)	Mean	Std. Dev. (δ)
Overall						
Minutes on Intervention	102.2	136.1	107.0	107.8	105.9	114.9
Aggregate - Rotation Challenge						
Minutes	42.4	60.6	27.8	29.7	31.2	39.5
Number Correct	44.7	61.0	34.1	35.5	36.5	43.0
Number Questions	54.2	79.9	41.0	45.7	44.0	55.7
%					82.9%	
Aggregate - Analogy Challenge						
Minutes	42.8	72.3	67.7	82.2	62.0	80.7
Number Correct	38.2	64.1	81.4	90.5	71.5	87.1
Number Questions	66.0	116.0	139.9	169.0	122.9	161.4
%	57.9%		58.2%		58.2%	
Introduction						
Minutes	4.3	4.7	3.2	4.7	3.5	4.7
Match Rotation Practice						
Minutes	8.2	6.5	5.6	5.3	6.2	5.7
Number Questions	8.3	7.1	6.4	6.7	6.8	6.9
Match Analogies Practice						
Minutes	4.5	6.0	2.7	3.2	3.1	4.1
Number Questions	4.2	5.1	3.8	5.2	3.9	5.2

Table 11 Descriptive Intervention Statistics for the Different Population Subgroups

Intervention Type	Females		Males		Total Population	
	Mean	Std. Dev. (δ)	Mean	Std. Dev. (δ)	Mean	Std. Dev. (δ)
Rotation Challenge (Easy)						
Average Minutes	8.1	8.6	5.2	3.9	5.9	5.5
Average Number Correct	11.7	9.1	8.0	5.8	8.9	6.9
Average Number Questions	12.6	9.7	8.3	6.0	9.3	7.3
%	92.9%		96.5%		95.4%	
Rotation Challenge (Medium)						
Average Minutes	15.9	23.1	9.0	11.9	10.6	15.5
Average Number Correct	14.6	21.3	9.8	13.0	10.9	15.5
Average Number Questions	17.8	27.1	12.3	17.3	13.6	20.1
%	82.1%		79.3%		80.2%	
Rotation Challenge (Hard)						
Average Minutes	18.5	31.8	13.6	18.9	14.7	22.6
Average Number Correct	18.4	32.6	16.3	23.1	16.8	25.6
Average Number Questions	23.8	45.5	20.4	29.6	21.2	33.9
%	77.2%		79.7%		79.1%	
Analogy Challenge (Easy)						
Average Minutes	12.6	19.1	15.3	16.6	14.6	17.3
Average Number Correct	10.0	10.9	16.8	15.7	15.3	15.0
Average Number Questions	15.5	17.0	25.1	23.1	22.9	22.2
%	64.6%		66.8%		66.5%	
Analogy Challenge (Medium)						
Average Minutes	11.4	20.3	21.3	27.5	19.1	26.3
Average Number Correct	12.5	22.9	28.4	35.2	24.8	33.5
Average Number Questions	19.1	36.2	46.0	64.7	39.8	60.5
%	65.6%		61.8%		62.3%	
Analogy Challenge (Hard)						
Average Minutes	18.8	37.7	31.1	43.8	28.3	42.8
Average Number Correct	15.6	31.5	36.2	46.8	31.5	44.6
Average Number Questions	31.3	64.6	68.7	90.4	60.1	86.6
%	49.8%		52.6%		52.3%	

Table 12 Descriptive Intervention Statistics for Population Subgroups for Challenges by Level

7.7 Cronbach α Results Summary

Cronbach α results from the online PSVT:R pretest for N=285 students included in the study.

Reliability analysis

Call: alpha(x = PPCStudy.Pretest.285, check.keys = TRUE)

raw_alpha std.alpha G6(smc) average_r S/N ase mean sd

0.86 0.85 0.94 0.16 5.6 0.015 159 7.4

lower alpha upper 95% confidence boundaries

0.83 0.86 0.88

Reliability if an item is dropped:

raw_alpha std.alpha G6(sm) average_r S/N alpha se

Q01	0.85	0.83	0.93	0.15	5.0	0.015
Q02	0.86	0.85	0.94	0.17	5.8	0.015
Q03	0.86	0.85	0.94	0.17	5.9	0.015
Q04-	0.86	0.85	0.94	0.17	5.8	0.015
Q05	0.85	0.83	0.93	0.15	5.0	0.015
Q06-	0.86	0.85	0.94	0.17	5.8	0.015
Q07	0.85	0.83	0.93	0.15	5.0	0.015
Q08	0.86	0.85	0.94	0.17	5.8	0.015
Q09	0.85	0.84	0.93	0.15	5.1	0.015
Q10	0.86	0.85	0.94	0.17	5.8	0.015
Q11	0.85	0.84	0.93	0.15	5.3	0.015
Q12-	0.86	0.85	0.94	0.17	5.7	0.015
Q13	0.85	0.83	0.93	0.15	5.0	0.015
Q14-	0.86	0.85	0.94	0.17	5.8	0.015
Q15-	0.86	0.85	0.94	0.17	5.9	0.015
Q16	0.85	0.84	0.93	0.15	5.1	0.015
Q17	0.85	0.83	0.93	0.15	5.0	0.015
Q18	0.85	0.84	0.93	0.15	5.1	0.015
Q19-	0.86	0.85	0.94	0.17	5.8	0.015
Q20	0.86	0.85	0.94	0.17	5.8	0.015
Q21	0.85	0.84	0.93	0.16	5.3	0.016
Q22	0.85	0.84	0.93	0.16	5.4	0.016
Q23	0.84	0.83	0.93	0.15	5.1	0.016
Q24	0.85	0.84	0.93	0.16	5.3	0.016
Q25	0.84	0.83	0.93	0.14	4.9	0.016
Q26	0.83	0.83	0.93	0.15	5.0	0.017
Q27	0.84	0.84	0.93	0.16	5.4	0.016
Q28	0.84	0.84	0.93	0.16	5.4	0.016
Q29	0.85	0.84	0.93	0.16	5.3	0.016
Q30	0.85	0.84	0.93	0.16	5.4	0.016

Item statistics

n raw.r std.r r.cor r.drop mean sd

Q01 285 0.526 0.734 0.7569 0.5067 99 5.90
 Q02 285 0.054 0.103 0.0657 0.0265 103 6.17
 Q03 285 0.029 0.067 0.0067 0.0253 108 0.84
 Q04- 285 0.046 0.103 0.0314 0.0148 132 6.91
 Q05 285 0.523 0.726 0.7487 0.4994 119 7.15
 Q06- 285 0.058 0.144 0.0809 0.0563 122 0.49
 Q07 285 0.528 0.733 0.7564 0.5017 131 7.83
 Q08 285 0.180 0.154 0.1286 0.0996 134 17.96
 Q09 285 0.480 0.665 0.6860 0.4376 140 11.88
 Q10 285 0.055 0.134 0.0672 0.0509 146 0.81
 Q11 285 0.353 0.499 0.4885 0.2890 150 15.53
 Q12- 285 0.087 0.162 0.0962 0.0820 91 1.11
 Q13 285 0.525 0.731 0.7543 0.4932 159 9.46
 Q14- 285 0.048 0.094 0.0202 0.0038 82 9.85
 Q15- 285 0.028 0.066 0.0167 -0.0175 78 10.10
 Q16 285 0.485 0.669 0.6905 0.4317 175 14.79
 Q17 285 0.525 0.733 0.7561 0.4885 178 10.65
 Q18 285 0.491 0.675 0.6964 0.4353 182 15.39
 Q19- 285 0.069 0.092 0.0175 0.0187 58 11.26
 Q20 285 0.065 0.132 0.0688 0.0607 194 0.89
 Q21 285 0.585 0.476 0.4897 0.5477 199 11.91
 Q22 285 0.557 0.446 0.4491 0.5002 204 17.22
 Q23 285 0.684 0.694 0.6978 0.6168 208 24.83
 Q24 285 0.594 0.480 0.4951 0.5546 214 12.80
 Q25 285 0.787 0.854 0.8813 0.7517 219 18.46
 Q26 285 0.829 0.728 0.7402 0.7791 221 29.59
 Q27 285 0.676 0.444 0.4394 0.5903 226 30.22
 Q28 285 0.701 0.457 0.4643 0.6187 232 31.05
 Q29 285 0.699 0.463 0.4608 0.6048 235 34.54
 Q30 285 0.676 0.442 0.4362 0.5845 241 32.31

Message:

In alpha(PPCStudy.Prestest.285, check.keys = TRUE) : Some items were negatively correlated with total scale and were automatically reversed. This is indicated by a negative sign for the variable name.