INVESTIGATIONS IN THE IMPACT OF VISUAL COGNITION AND SPATIAL ABILITY ON STUDENT COMPREHENSION IN PHYSICS AND SPACE SCIENCE

by

XIMENA CLARA CUICATL CID

Presented to the Faculty of the Graduate School of

The University of Texas at Arlington in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

THE UNIVERSITY OF TEXAS AT ARLINGTON

May 2011

ACKNOWLEDGEMENTS

"Sigue adelante, nunca para atrás!" – Gramma Cid

Thank you mom and dad for never doubting me. I have traveled all over the country and have known that no matter where I was or what I was doing you would always support me. For that I am blessed. Dad, I know that you are not here to watch me walk across the stage, but I know you are always watching. Mom, you continue to be my biggest inspiration. Sue Sue thanks for always making sure I was ok. Naiders, Mazatl, and Miguel, love you guys. Alex, girl you have kept my life interesting, here's to another ten years! Heather, girl thanks for being my writing buddy, it has been intense.

I couldn't have asked for a better research group. Ramon Lopez, thank you for taking a chance on me. You have shown me what it means to be a scientist and a mentor. You have seen me through some tough times and have always given me space to grow as a person and as a scientist and I thank you for that. Betsey and Robert thanks for always being there for me. Couldn't have made it so far without your daily support.

I am grateful to acknowledge that this dissertation is based upon work supported by National Science Foundation grants DUE-0856796, GEO-0803981, GEO-0823929, and GEO-0607195 with additional support from CISM, which is funded by the STC Program of the National Science Foundation under Agreement Number ATM-0120950 and US Department of Education GAANN Grant # P200A090284.

Life is a battle, but if you don't pick yourself up and keep going, you miss all the fun adventures.

April 22, 2011

ABSTRACT

INVESTIGATIONS IN THE IMPACT OF VISUAL COGNITION AND SPATIAL ABILITY OF STUDENT COMPREHENSION IN PHYSICS AND SPACE SCIENCE

Ximena Clara Cuicatl Cid, Ph.D.

The University of Texas at Arlington, 2011

Supervising Professor: Ramon Lopez

Physics and Space Science examine topics that are highly spatial in nature. Students are required to visualize a system, manipulate that system, and then solve a given problem. Doing all of this, simultaneously, can lead to a cognitive overload where the student is unable to correctly solve the problem. Some difficulties may be rooted in conceptual difficulties, whereas other difficulties may arise from issues with spatial intelligence and visual cognition. In some cases, students might have created an incorrect mental image of the problem to begin with, and it's this misconception, not the lack of content knowledge, that has caused an incorrect answer. It has been shown that there is a correlation between achievement in Science Technology Engineering and Mathematics (STEM) fields and spatial ability. My work focuses on several discrete investigations in topics that relate to student learning in physics and space science and the relationship to spatial ability.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iii

ABSTRACT iv

LIST OF ILLUSTRATIONS viii

LIST OF TABLES ix

Chapter Page

1. BACKGROUND 1

1.1 What is Physics Education Research (PER)? 1

1.2 Cognitive Load 6

1.3 Visual Cognition and Spatial Ability 8

1.3.1 The Mental Rotation Test 11

1.4 Experts vs. Novices 12

1.5 Relationship Between Visual and Spatial Abilities and Working Memory 15

1.6 Spatial Abilities in Other STEM Fields 17

1.6.1 Chemistry 17

1.6.2 Mathematics 19

1.6.3 Geoscience and Earth Sciences 20

1.6.4 Engineering 22

1.7 Discussion 23

2. COLOR AS A THIRD DIMENSION 24

2.1 Introduction 24

2.2 Background for Study 24

2.3 Methodology 27

2.4 Findings and Supporting Data 32

2.4.1 Finding 1: Preconceived Interpretation of Figures 32
2.4.2 Finding 2: Preference for Figure Order 34
2.4.3 Finding 3: Distractions 37
2.4.4 Finding 4: Embedded Color Scheme 39
2.4.5 Finding 5: Previous Experience Has no Effect 41
2.5 Discussion 42
3. REPRESENTATIONS OF LUNAR PHASES 44
3.1 Introduction 44
3.2 Background for Study 44
3.3 Methodology 48
3.3.1 Lunar Phases Concept Inventory and Misconceptions in Astronomy 50
3.4 Lunar Phases Lab and Data 54
3.5 Discussion 56
4. CURRENT AND FUTURE WORK 58
4.1 Introduction 58
4.2 Spatial Ability Comparison of STEM Majors 60
4.2.1 Methodology 61
4.2.2 Data 62
4.2.3 Results 64
4.2.4 Discussion 65
4.3 Test ReTest Effect of MRT 66
4.3.1 Methodology 67
4.3.2 Data and Results 67
4.3.3 Discussion 69
4.4 Correlation Between Spatial Ability and Mathematics Comprehension 69
4.4.1 Methodology 69

4.4.2 Data 70
4.4.3 Results 71
4.5 Correlation Between Spatial Ability and Engineering 72
4.5.1 Methodology 72
4.5.2 Data for Total Engineering 72
4.5.3 Data for Subset of Engineering 73
4.5.4 Results 75
4.6 Correlation Between Spatial Ability and Physics Comprehension 75
4.6.1 Methodology 75
4.6.2 Data for Physics 1443 76
4.6.3 Data for Physics 1444 78
4.6.4 Results 79
4.7 General Conclusions 80
5. GENERAL DISCUSSION AND FUTURE WORK 82
5.1 General Conclusions 82
5.2 Future Work 85
APPENDIX
A. SURVEYS FROM CHAPTER 2 87
B. SCANTRON USED FOR FORCE CONCEPT INVENTORY 91
REFERENCES 93
BIOGRAPHICAL INFORMATION 103

LIST OF ILLUSTRATIONS

Figure Page

1.1 Hake Plot 5

1.2 Shepard and Metzler Matched and Unmatched Pairs 9

1.3 Example of MRT. 11

1.4 Example of Molecular Diagrams 18

1.5 Molecule Rotation Task. 18

1.6 Example of PSVT: R. 21

2.1 Topographic Map Depicting a Volcano 30

2.2 Topographic Map Depicting a Crater with Spike at Center 30

3.1 Example of Elongated Elliptical Orbit 45

3.2 GeoWall 47

3.3 AstroWall Software 49

3.4 Example of LPCI Question 51

4.1 Graphical Representation of the Pre-Post MRT Data 64

4.2 Graphical Representation of the Pre-Post MRT Data for Engineering and Physics 65

4.3 Graphical Representation of Mean Differences for Engineering, Physics, and Test ReTest 68

4.4 Correlation Table for Pre MRT and Final Grades in Calculus 1 71

4.5 Correlations for Total Engineering Pre MRT and Final Grades 73

4.6 Correlations Between MRT and Subset Engineering Final Grades 74

4.7 Correlations for MRT, FCI, and Final Grades for Physics 1443 78

4.8 Correlations for Physics 1444 Final Grades and MRT 79

LIST OF TABLES

Table Page

2.1 Student Preference for Figure by Color Scheme....................36

2.2 Data for Student Preference in Order of Presentation 37

3.1 Gender and Major 55

3.2 Age and Classification....................55

3.3 Previous Astronomy of Physics Courses 55

3.4 Means and Standard Deviation for Pre and Post LPCI Data 55

4.1 Sample Size for Pre-Post MRT STEM Comparison....................63

4.2 Means and Standard Deviation for Pre-Post MRT 63

4.3 Means and Standard Deviation for Engineering, Physics, and Test ReTest 68

4.4 Means and Standard Deviation for Pre-MRT and Final Grades for Calculus 1 70

4.5 Means and Standard Deviation for Total Calculus Sample....................71

4.6 Means and Standard Deviation for Total Engineering Data....................73

4.7 Means and Standard Deviation for Subset of Engineering Data 74

4.8 Means and Standard Deviation for MRT, FCI, and Final Grades for Physics 1443 77

4.9 Means and Standard Deviation for Physics 1444....................78

CHAPTER 1

BACKGROUND

1.1 What is Physics Education Research (PER)?

The opening lines of many Physics Education Research (PER) review articles is a quote from F. K. Richtmyer's original article "Physics is Physics" from the American Physics Teacher [1933]. He states, "Teaching, I say, is an art and not a science." This statement is controversial in that it expresses that science and education are completely separate. If that were the case one could question why we have physics education research being conducted in physics departments. This quote has been at the heart of many of these debates, and an important question is how to categorize PER in relation to the rest of physics.

Redish [1999] discusses how the creation of a body of knowledge in science is like creating a map. No one person creates the map, but the collective bodies of researchers contribute to the map and the community as a whole agrees that the conceptual ideas are correct. He discusses that this map creation is agreed upon by peer-reviewed publications and, as in the case of other disciplines, PER should be considered a science. He states that confusion comes because "people sometimes forget the role of the mind in doing physics," and that "in order to do the best physics education research, we not only have to create an understanding of how people think, thereby possibly creating new cognitive science, we have to rethink/reformulate elements we take for granted but which our students lack." The idea of incorporating the mind is crucial for me because my work focuses on how students use their minds to create or manipulate mental images of abstract physics concepts.

Over recent decades there have been considerable advances in the field of PER. Many of these advances and changes in physics departments' culture and acceptance are credited to Lillian McDermott's group at the University of Washington. McDermott, in her Oersted Lecture paper [2001], argues that the PER community collects and reports findings in peer-reviewed journals and professional meetings similar to other science. Because of the nature of PER, she describes The University of Washington's Physics Education Group's structure of research as "an empirical applied science." She also makes the claim that science education research should be conducted by science professors in science departments, which is known as discipline-based education research. The main reason that this should be the case is that science professors have the content knowledge that is need in order to truly understand the misconceptions created by the students that they are trying to educate. This view, that PER should be done by physicists in physics departments, was endorsed in a 1999 American Physical Society Council statement that clearly identified PER as a branch of physics, whose practitioners are subject to the same metrics as any other branch of physics [http://www.aps.org/about/governance/committees/popa/1999.cfm].

The University of Washington's PER group focuses on identifying student misconceptions and creating interventions to reduce the perpetuation of these misconceptions (for more information see McDermott and Redish [1999] for a list of about 115 misconceptions in physics). They use conceptual change strategies in the design of instructional materials to address known difficulties that students have, and they evaluate the effectiveness of those materials both with pre-post-testing and student interviews.

Many of the studies on misconceptions led to the development of concept inventories. A concept inventory is a multiple-choice test designed to explore a student's comprehension of a given topic. The creation of a concept inventory is extremely tedious. One cannot simply create a multiple-choice test and have it automatically be considered a concept inventory; there are a number of steps involved in the creation of concept inventories. In his paper on the creation of the Test of Understanding Graphs in Kinematics (TUG-K), Beichner [1994] proposed a model for creating research-oriented multiple choice tests which can be used as diagnostic tools and for formative and summative evaluations of instruction. When creating a concept inventory you need to verify the content validity and reliability of the assessment. He suggests, based on the field of educational assessment, that a test have a mean of 50% in order to maximize the spread of scores. He also explains that validity is considered accuracy, meaning does the test measure what you think it measures? Reliability, on the other hand, is how precise the measurement is. He discusses that validity is not really calculated but is established, where as there are several different types of statistics that can be used to assess the reliability. The most common statistical test is the KR-20 coefficient (named after the statisticians Kuder and Richardson who developed it). If a test has a KR-20 ≥ 0.7 it is generally considered to be reliable.

In mechanics there are four main concept inventories that are used in the literature: the Mechanics Diagnostic Test [Halloun and Hestenes, 1985], the Mechanics Baseline Test [Hestenes and Wells, 1992], the Force Concept Inventory (FCI) [Hestenes et. al, 1992], and the Force and Motion Conceptual Evaluation (FMCE) [Thornton and Sokoloff, 1998]. There are now over 30 inventories that have been created on concepts in physics and space science (see http://www.ncsu.edu/per/TestInfo.html). The two main concept inventories that are used in this work are the FCI and the Lunar Phases Concept Inventory (LPCI) [Lindell and Olsen, 2002].

The FCI is a 30 question multiple-choice assessment that has been widely used throughout the literature, since its inception in 1998. Hestenes and colleagues [1992] claim that "[the] FCI requires a forced choice between Newtonian Concepts and commonsense alternatives." There are five main topics that are covered on the FCI. Kinematics has six questions, Newton's 1st Law has eight questions, Newton's 2nd Law has four questions, Newton's 3rd Law has four questions, Superposition Principle has four questions, and Kinds of Forces has sixteen questions (forces include: Solid Contact, Fluid Contact, and Gravity). The FCI is widely used in the literature so reliability and validity has been verified.

The FCI was chosen to be used for our research because of its validity, reliability, and continual use in the literature, including a study conducted by Hake [1998]. Hake [1998] collected a six-thousand student survey of pre-post FCI data. The 62 introductory physics courses that comprised the study were high school, community college, and university courses. The plot that represents the basic findings of Hake [1998] showed those classes that utilized interactive engagement (see Figure 1.1) had a larger conceptual gain than those that did not. In addition, Mazur [1997] found that interactive engagement did not diminish scores on conventional exams. Thus evidence exists that research-based, active engagement pedagogy results in improved conceptual understanding in physics classes with no loss of traditional problem-solving skills.

For the study [Chapter 3] conducted on Lunar Phases, the LPCI [Lindell and Olsen, 2002] was chosen because it is the only concept inventory that assesses student comprehension on Lunar Phases (please see Lindell and Olsen [2002] for details regarding validity and reliability). The LPCI is a multiple choice assessment with 20 questions devoted to orbital period, phase period, direction of rising and setting, main phases (New Moon, 1st quarter, Full Moon, 3rd quarter), eclipses, and phases if viewed from different locations on Earth. There are an additional 9 multiple choice questions devoted to collecting student demographics and attitudinal information which included the following: confidence for answers chosen, major, age, home community (i.e. urban suburban, rural, etc.), ethnic background, highest math level achieved, confidence in math, and confidence in science. In chapter 3 details about our study will be further described along with the motivation for the study and the outcomes, so I will refrain from going into details here.

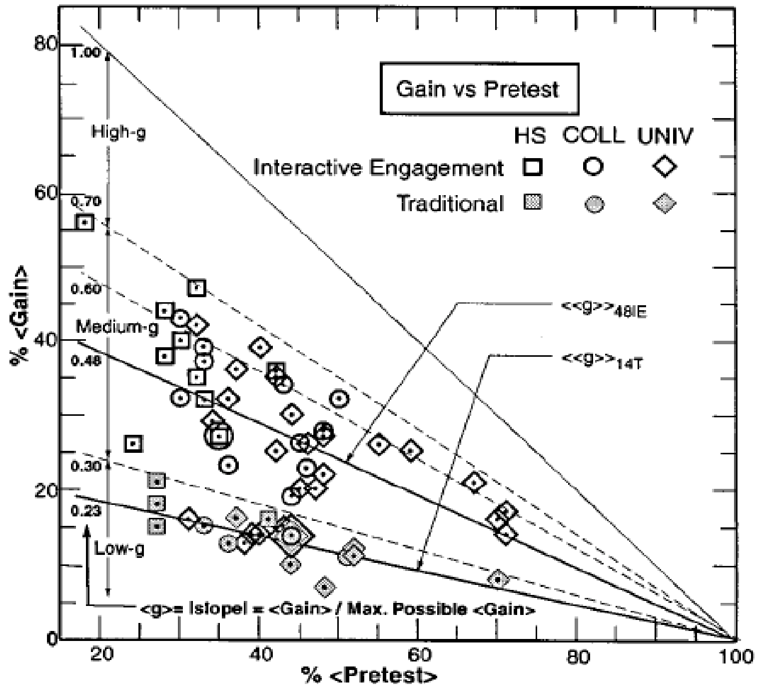


Fig. 1. %⟨Gain⟩ vs %⟨Pre-test⟩ score on the conceptual Mechanics Diagnostic (MD) or Force Concept Inventory (FCI) tests for 62 courses enrolling a total $N=6542$ students: 14 traditional (T) courses ($N=2084$) which made little or no use of interactive engagement (IE) methods, and 48 IE courses ($N=4458$) which made considerable use of IE methods. Slope lines for the average of the 14T courses $\langle\langle g\rangle\rangle_{14T}$ and 48 IE courses $\langle\langle g\rangle\rangle_{48\mathrm{IE}}$ are shown, as explained in the text.

Figure 1.1 Hake Plot. Figure and text taken from Hake [1998], which clearly shows that interactive engagement has a greater effect on student comprehension, as measured by the FCI, than non-active engagement.

The field of PER has evolved tremendously and has branched out in a multitude of directions. As briefly stated above, my research interests are centered on how students' manipulation of mental visualizations of abstract physics and space science concepts, or other visual processing of information, impacts student learning. Because of this interest in how the mind works and how students process information, we use a cognitive science/psychological approach. Due to this approach, in the following sections, a brief description of the relevant aspects of cognitive science will be discussed.

1.2 Cognitive Load

In Cognitive Psychology it has been well established that a person has a limited amount of working memory that can be accessed while trying to complete a task [e.g. Sweller, 1988]. In the purest sense, psychologists have determined that people can hold a finite amount of verbal memory (words or sounds) and a finite amount of visual-spatial (I will go into more detail on visual-spatial skills in a future section) objects in their mind at once, as long as these object or sounds are unrelated [e.g. Baddeley, 2003 and references therein]. Thus 7, 1, 6, 7 would generally be treated as four discrete elements in a random string of numbers while 1776 is a single mental object.

Working memory is vital for multi-tasking. It is the ability to hold a phone number in your mind while, getting your keys out of your pocket, while walking back to your car. The more working memory a person has, the more capable a person is to complete each task simultaneously [e.g., König et al., 2005]. Engle [2002] argues that a person's working memory is not necessarily about the power of their actual memory, but more related to how well a person can maintain attention on multiple tasks without getting distracted. In my opinion, this is more relevant to studies in the sciences because of the nature of problem solving in the sciences. Students go to lecture and listen to the instructor describe a topic, this is followed by reading the book (if the student is diligent), followed by problem solving. From my own experiences, with my students, keeping attention focused on the task at hand, without getting distracted is where they have issues. As a person increases the number of tasks they are trying to focus on, they are increasing what is called their cognitive load.

Cognitive load has been studied extensively in psychology in many different domains, but this research is going to focus on cognitive load as it applies to problem solving in physics, and in particular the cognitive load imposed by processing visual information. The field of physics uses diagrams to represent complex systems in order to reduce the cognitive load of the students. If there is a diagram presented to the student, the student does not have to use their mind to create the image, though there may be additional mental processing needed to manipulate and use it. Meltzer [2005] claims that there is no purely abstract understanding of a physical concept, instead we continue to create representations for abstract concepts in order to aid in the comprehension. Furthermore, he discusses the large number of previous studies that suggest that developing the ability to understand and manipulate these representations is vital to understanding concepts in physics (see Meltzer [2005] and references therein). Although these representations are created for the purpose of increased conceptual understanding, these representations can create new misconceptions for students. For example, in electricity and magnetism (E&M), diagrams used in most textbooks that are intended to represent an E&M plane wave and aid understanding, appear to be the source of some additional misconceptions about E&M waves that must be addressed by subsequent instruction [Ambrose et al., 1999].

There is a difference, however, in conceptual misconceptions and misconceptions that are created from incorrect manipulations of mental images. For example, there is evidence that the mental manipulation of 3-D images plays a role in increasing cognitive load when dealing with the relationship between magnetic fields and currents [Lopez and Hamid, 2004]. Students were given a standard textbook figure of a complex current system that occurs in near-Earth space during a geophysical disturbance called a substorm, and they were to determine the magnetic perturbations produced by the current system. Lopez and Hamid [2004] found that the conceptual knowledge of the students were not the source of the difficulty in the task presented to the students. It was the manipulation of the mental images that caused the students to come to wrong conclusions about the magnetic perturbations produced by the current. Thus the issue was caused by the creation of an incorrect mental image based on the diagram and the rotation (a spatial task) of that mental image. In the next section, details regarding the relationship between visual cognition and spatial ability, and how they relate to problem solving in physics, will be discussed.

1.3 Visual Cognition and Spatial Ability

Cognitive scientists have a number of ways of classifying visual cognition and spatial abilities, but for the purposes of this document, visual cognition will be defined as the ability to create a mental image or representation and then manipulate that image with the mind [Kosslyn, 1995]. Spatial ability, as described by Salthouse and colleagues [1990], can be thought of as the mental manipulation of spatial information to determine how a given spatial configuration would appear if portions of that configuration were to be rotated, folded, repositioned, or otherwise transformed. For my research, I am primarily interested in the rotation aspect of this definition, as it is a foundational spatial ability [e.g. Kosslyn, 1995; 2005]. There are slight differences in these two definitions, but in this work they will be thought of as an intertwined pair.

Shepard and Metzler [1971] conducted a study in which they gave participants 1600 pairs of tetris-like objects (see Figure 1.2). The pairs were either matched pairs or mirror images of each other. The matched pairs were either rotated in what the authors classify as “picture frame” or “depth” (see Figure 1.2 for author description). The subjects were instructed to pull a right lever if the pairs were matched or to pull a left lever if the pairs were not matched. The subjects were timed in how long it took them to determine if they were matched pairs or un-matched pairs; pulling the lever stopped the time. When the authors compared the time it took to make a decision with the angle the pairs were rotated through, they found a linear relation. This means that when an object is rotated through more and more degrees, the processing time for the brain to understand the association is increased, hence the cognitive load for a task that requires a larger mental rotation is increased.

But how does this relate to students learning physics? When students are presented with a 2-dimensional image, say in a textbook, often they have to use their visual and spatial abilities to imagine that system in 3-dimensions in order to solve problems. If they are given one snap shot of an object drawn from a single perspective and asked to solve a problem that requires a different view of the object, or something that is perpendicular to the object, then they have to rotate their mental image, which increases their cognitive load [e.g. Lopez and Hamid, 2005]. There is also historical evidence showing that visualizations are essential to thought experiments such as the ones performed by Einstein and Newton [e.g. Botzer and Reiner, 2005, and references therein].

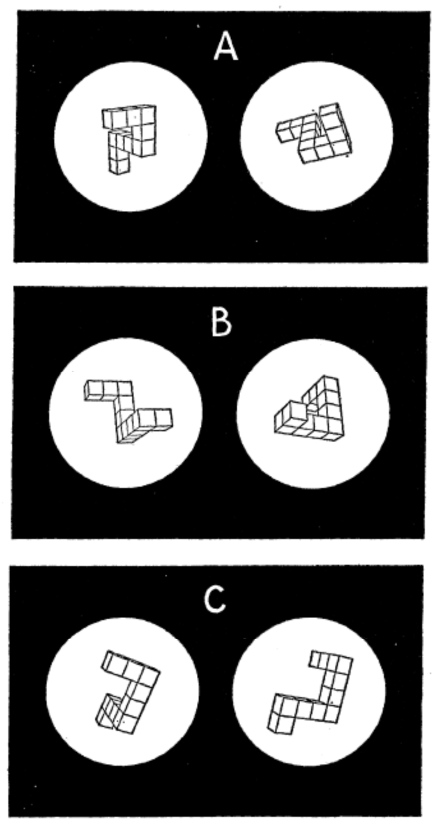


Fig. 1. Examples of pairs of perspective line drawings presented to the subjects. (A) A "same" pair, which differs by an 80° rotation in the picture plane; (B) a "same" pair, which differs by an 80° rotation in depth; and (C) a "different" pair, which cannot be brought into congruence by *any* rotation.

Figure 1.2 Shepard and Metzler Matched and Unmatched Pairs. Figure and text taken from Shepard and Metzler [1971]. Image A shows a matched pair with rotation in the picture plane, image B shows a matched pair with rotation in depth, and image C shows an unmatched pair.

It has been suggested that science students have a higher spatial ability than non-science students, but physics students have the highest spatial ability of all the sciences followed by geoscience students [Siemakowski and MacKnight, 1971]. Pallrand and Seeber [1984] suggested that visual spatial ability is needed to succeed in courses like introductory physics. They also found that the simple act of taking a course like introductory physics courses improves your visual spatial ability. It is not known, however, if students taking physics have a high spatial ability because they are taking physics, or if physics students already have high spatial ability and hence this is one of the reasons they are in physics. In other words, there is no evidence to decide if students are self-selecting themselves to be in physics or if it is the act of taking the courses that is improving their spatial ability.

Unfortunately, there has been little research devoted to the role that spatial ability has on student comprehension in physics, and essentially no work in space science. Kozhevnikov and colleagues [2001] have suggested that in kinematics-type problems there is a correlation between individual differences in visual-spatial ability and problem solving. They suggest that high visual-spatial students might have the ability to grasp concepts from real-world situations easier than low visual-spatial students. They suggest that both the high and low visual-spatial students start a problem with the same misconceptions, but high visual-spatial students have the ability to internalize the conceptual information faster and therefore come to more correct answers on the conceptual assessments. In their studies, Kozhevnikov and colleagues [2001] also suggested that once conceptual knowledge has been gained, visual-spatial ability is no longer a predictor of performance on kinematic problems. This is interesting because in other Science, Technology, Engineering, and Mathematics (STEM) fields, it has been suggested that visual-spatial ability is a predictor of overall success and retention in introductory courses [e.g. Sorby, 2005]. Details of spatial ability in other STEM fields will be discussed in Chapter 4.

1.3.1 The Mental Rotation Test

In the studies mentioned above, there are several assessments that can be used to measure a student's visual-spatial ability. The one that is used in this work is the Mental Rotation Test (MRT) [Vandenberg and Kuse, 1978]. The MRT was chosen because of its extensive use in the literature, its availability, its validity, and its reliability. It must be noted that there are studies that show that there are individual differences (e.g. gender differences) present on the MRT [e.g., Linn and Petersen, 1985; Bors and Vigneau, 2011]. In terms of gender, males tend to score higher than females, but because we are not interested in individual differences, we are not worried about these types of biases.

We use the paper based MRT because it is easier to take into a classroom and eliminates a location threat. The paper based MRT has two parts. Each of the parts has 10 questions and it is timed with a limit of 3 minutes. In between the two parts the students are given a 1-minute break. If students finish a section before the 3-minute time expires, they are allowed to check their answers for that part only.

The questions are structured where there is an image on the left and they are to choose two of four images presented at the right that correctly match the image on the left (see Figure 1.3). In order to get credit for each question, the students must correctly identify both images that match the image on the left.

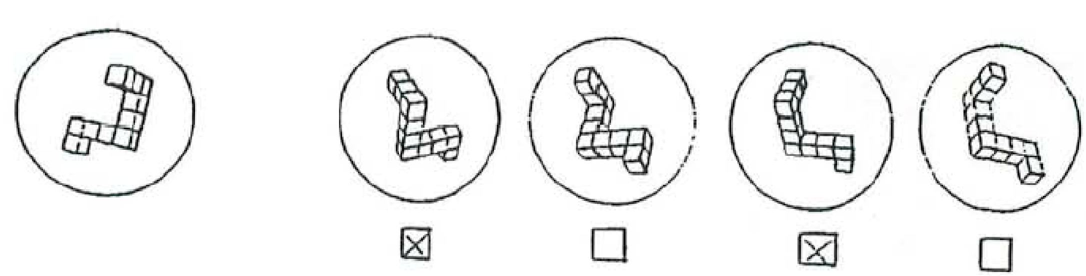


Figure 1.3 Example of MRT. Figure taken from actual paper based MRT [Vandenberg and Kuse, 1971] and depicts an image on the left and four images on the right. The subjects are to choose two of the four images on the right that correctly match the image given on the left and in the figure above, the correct images are the first and third images, as marked by x's. Figures two and four are mirror images of the image given on the left.

1.4 Experts vs. Novices

It has been known for the past few decades that experts store information in their minds differently than novices. Section 1.2 discussed cognitive load and working memory. When information is transferred from a person's working memory to long-term memory, it is stored in such a way as to facilitate easy access. Sweller [1994] gives an example of how people read in terms of experts and novices. He discusses that novices, who have just learned to read, focus on the order of individual letters and order of words. Adults, experts at reading, focus on the meaning of the document they are reading instead of the words that are actually written. They have switched to what Sweller describes as "automatic processing" instead of "conscious processing." Conscious processing is where the subject holds information in working memory whereas automatic processing is where the subject is pulling information that has been stored in long-term memory and uses their working memory for other processes. The meaning of words and the order of letters that form words are information that is stored in an adult's long-term memory so they can devote their working memory to the meaning of the sentences and topics instead of the individual letters and words.

The experts use a technique called "chunking" in order to comprehend the meaning of the document. The term "chunking", in cognitive sciences, means that chunks of familiar or repeated information are used as a single unit of information [Larkin et al., 1980], therefore reducing the working memory cognitive load. For example, an expert understands a word, or a group of words, and utilizes the information stored in their long-term memory associated with that particular word or group of words. Their working memory is not used to comprehend individual words but instead is used to understand chunks of words. A novice, who is barely learning to read, on the other hand, uses individual letters as the information and therefore has to use more working memory to process the document they are reading. They have not had the experience and/or practice to have transferred the meaning of chunks of letter (that make up words) to their long-term memory therefore they are increasing their cognitive load in order to understand the document. A classic example in cognitive science, describing the difference between experts and novices, is chess masters.

Studying chess masters in cognitive science was one of the first breakthroughs that demonstrated the difference in memory use between novice and experts [e.g. De Groot, 1978; Chase and Simon, 1973a; 1973b; Chi et al 1981; Mestre, 1991]. Chess masters can look at a board with a chess game in progress for a few seconds and exactly recreate where each chess piece was placed. This is because chess masters have chunked countless opening moves, attack moves, defense moves, and closing moves in their long-term memory. They can easily access their long term memory and look at a chess board and know what moves could have been used to create the game structure and they know what moves should be preformed next in order to win the game. The novice on the other hand does not have countless moves stored in their long-term memory to rely on. They look at the chessboard and try to remember where each individual piece is at on the board. Remember, a person can only hold a certain amount of information at one time in their working memory, therefore the more a person tries to remember, the higher the cognitive load. This simple experiment shows that experts have the ability to pull chunks of information from long-term memory and use it as individual bits of information in their working memory. Novices do not have the same knowledge stored in their long-term memory so they do not have the ability to chunk information as experts can.

Larkin and colleagues [1980] devised a study to explore how experts and novices solve basic kinematics physics problem. Experts access information similar to how we use an index in a book. They have created a stored database of information based on experiences and practice that they can utilize when needed. Novices have not had enough experience to store information with a sufficiently flexible access mechanism in their long-term memory for fluent, on-the-fly, access when needed.

Larkin and colleagues [1980] have also found that the experts tend to have a “working forward” strategy to solving problems. If you know the value of all of the independent variables, you have a more complete understanding of the problem/system. They do not use sub-goals to accomplish the task. Novices do not fully understand the kinematics equations that they have written down; therefore they approach the problem with a "working-backwards" approach. Novices have a desired outcome and searching through the equations in order to find the ones that will solve for the unknown variable. If they do not have an equation to directly solve for the desired quantity, they set sub-goals to accomplish the task. The expert has solved the problem with every unknown variable in mind, from known equations, and only plug in numbers at the end. They have a framework that they can rely on. It has been built over years of practice and exposure, so they can look at a problem and chances are they have solved that type of problems before, and, like the chess masters, they have a store of knowledge to pull from.

The novice, on the other hand only knows how to look for specific equations that will define the given unknown and if future problems ask for other unknowns, they will solve for them again by searching through equations for the specific unknown variable. For example, if a problem asks to solve for "V" they will look through their equation list and pull every equation that has "V" in it, but "V" could mean velocity, potential, volume, voltage, etc. The novice does not have a store of knowledge with a flexible conceptual framework on which to rely that allows them to easily distinguish, in an automatic fashion, between equations with "V" that are appropriate and those that are inappropriate for the situation at hand. In fact, the expert will not consciously consider equations with the inappropriate "V", and mention of such equations (like $V = IR$ in a problem about volume) would elicit surprise, and perhaps amusement, from the expert. Novices, on the other hand, have an overwhelming amount of unstructured information in front of them so they look for surface features (i.e., any equation that includes "V") to guide them.

Chi et al. [1981] found that physics professors have the same ability to chunk information that chess masters have. When novices and experts were asked to group problems into categories, the time it took to group the problems was about the same, but the way problems were grouped was different between the two groups. Novices tended to group problems together based on their surface features. For example, all problems with a block on an inclined plane were grouped into one category, where as all problems with pulleys and strings were grouped into another category. Experts, on the other hand, grouped problems together based on the concepts used to solve the problem. For example, all problems that use the concept of conservation of energy were grouped into one category, whereas another category would consist of all problems that use Newton's Second law ($\mathbf{F} = m\mathbf{a}$). So the time that the experts used in grouping the problems was for recalling conceptual ideas and identifying which concept belonged to which problem. The novices used their time to look at each of the problems and group them in terms of their surface features. Because the novice does not have the experience (or practice) in solving problems based on their conceptual ideas, they rely on how the problems look. The expert has conceptual information stored in their long-term memory so they focus on the solutions of the problems instead of how they look.

This issue of expertise has a clear connection to how students process visual information. When presented with some visual representation of information a novice will be trying to organize the visual by surface similarities. Thus they may be distracted by perceptually salient, but conceptually unimportant details. The expert, however, will be sorting conceptually important information from the visual, and may not even be aware of other features that attract the attention of the novice. In fact, one of the features of expertise is an inability of the expert to think like the novice any more [*How People Learn,* 1999]. This has important implications for the design of instruction and the creation of visual representations.

1.5 Relationship Between Visual and Spatial Abilities and Working Memory

In cognitive science and psychology, there is a difference between working memory and short-term memory. Miyake et al [2001] describes short-term memory as simple storage oriented tasks with no specific processing requirements. Working memory is considered to be complex tasks with both a storage component and a specific processing component. For the purposes of this research they will be considered equivalent and be used interchangeably.

Miyake et al. [2001] conducted a study to explore the relationship between spatial abilities and working memory. In their paper, Miyake and authors had three different categories for spatial abilities: Spatial Visualization, Spatial Relations, and Visuospatial Perception. Spatial Visualization is described to be tasks that require the mental manipulation of spatial forms. Spatial Relations is described to be tasks similar to Spatial Visualization tasks but often include planer rotations and speed. The third category, Visuospatial Perception, is described as simple tasks that require the identification of similar, identical, or hidden patterns within a given set of objects. Often Visuospatial Perception tasks are also timed. Miyake et al. [2001] found moderate to high correlations between these three tasks. Though there are differences between these tasks, the subtle differences are not relevant to this work. More importantly the spatial tasks that are most closely related to this work are Spatial Visualization and Spatial Relations, therefore these two tasks will be referred to generally as “spatial abilities” or “visual cognition” and be used in conjunction with each other or interchangeably. Please note that in their findings, these two spatial tasks had the highest correlation with working memory.

Miyake et al. [2001] claim that spatial relations tests use working memory and short term memory because in order to solve a spatial task, the target figure must be mentally represented in order to accomplish the transformation required (specifically rotation) to solve the problem. It was also expressed that spatial tasks are highly correlated with working memory [see references therein], which means that the more complex the spatial task, the more working memory one must use thereby increasing the cognitive load.

STEM fields often rely on visual mental images and the mental manipulation of these images to process information. Novices in particular are taught a multitude of visualization skills in order to aid in the comprehension of abstract concepts. The next section will highlight some of the different STEM fields that have explored the relationship between spatial abilities and subject comprehension.

1.6 Spatial Abilities in Other STEM Fields

From the previous sections, we know that cognitive load plays a role when students are trying to solve problems. In the STEM fields it is particularly important because of the abstract nature of the information. Students, in the STEM fields, constantly have to use their minds to imagine a system, manipulate that system (i.e. rotation, time progression, etc.), and then solve a problem based on the mental image they just created. Trying to do this simultaneously can increase a student's cognitive load which can lead to misconceptions that are not related to content knowledge, as was the case found by Lopez and Hamed [2004] when dealing with electromagnetic current systems in near-Earth space. Until this point, there hasn't been too much research exploring the nature of cognitive load created by spatial abilities and visual cognition, and student comprehension in physics and space science. There have, however, been studies exploring this topic with relation to other STEM fields. This next section is going to explain some of these studies and present a foundation for the purpose of this work.

1.6.1 Chemistry

It has been suggested that having strong spatial visual skills is vital to learning and problem solving in some aspects of chemistry. Stieff and colleagues [e.g., Stieff et. al 2005] have done extensive research on the role of spatial abilities and the comprehension of molecular diagrams. Most chemistry majors are required to take at least one Organic Chemistry course, which relies heavily on the comprehension of molecular diagrams. Molecular diagrams can be represented in a variety of different ways (three of which are depicted in Figure 1.4). Stieff [2007] claims that students are taught both analytic and visuospatial problem solving techniques. Similar to the MRT, Stieff [2007] conducted an experiment to test whether students were favoring analytical problem solving methods over mental rotations. He found that students were using mental rotations for molecular diagrams and similarly he found a linear relation for the time it took to solve the problem and the angular disparity (degree) of the matching molecules (See Figure 1.5 for sample of molecular test).

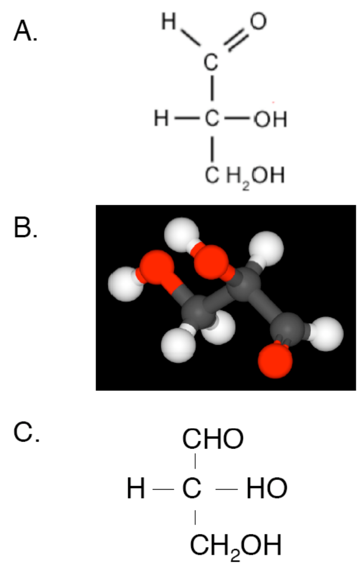


Figure 1.4: Examples of Molecular Diagrams. These are just three different representations of the molecule Glyceraldehyde. The actual molecule is not important, but is intended to give the reader an idea of how molecules can be represented with different diagrams.

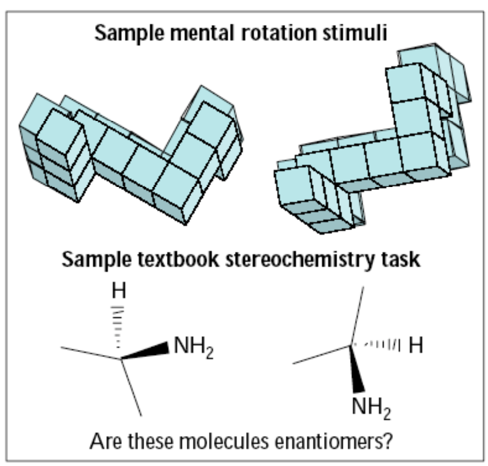


Fig. 2. Experimental stimuli from mental rotation tasks and chemistry assessments.

Figure 1.5: Molecule Rotation Task. Figure and text taken from Stieff [2007]. Both the blocks and the molecular representation require a spatial task of rotation through a given angle.

Though Stieff [2007] found a relationship between molecular diagrams and spatial ability, it must be noted that he did not find a relationship between all aspects of chemistry. Instead, the correlation was between specific aspects of chemistry and spatial ability. Organic chemistry is one of the subfields of chemistry that relies heavily on molecular diagrams so organic chemistry is one aspect that a correlation could be found.

1.6.2 Mathematics

It has been well established that there is a difference in rotational spatial abilities between males and females [e.g. Bors and Vigneau, 2011]. Casey and colleagues [1995] devised a study to test if the spatial ability gender differences, as measured by the Mental Rotation Test could be a predictor for the mathematics gender differences, as measured by the SAT-M, and therefore a predictor of why females are not entering STEM fields. Casey and colleagues [1995] were primarily interested in identifying a correlation for females' spatial ability and mathematics performance, but they also took data from males in order to do a more broad comparison. They took data from four different categories. They chose college students from all majors (not just STEM majors) from two liberal arts colleges in the Northeast, high scoring SAT-V college-bound high school students from a middle-income suburban high school in the Northeast who had elected to take the SAT's, low scoring SAT-V college-bound high school students from a middle-income suburban high school in the Northeast, and talented 7^{th} and 8^{th} grade students. Talented 7^{th} and 8^{th} grade students refer to students who scored high on the SAT-V and SAT-M (or ACT) and were invited to participate in a summer program for gifted students.

Casey and colleagues [1995] found that for females in all groups there was a correlation between the MRT and scores on the SAT-M. For males, both high scoring and low scoring college level students had a correlation between high spatial abilities and high scoring

SAT-V score, but for high scoring college bound and talented males, there was no correlation found. The authors acknowledge that spatial ability is not the primary cause for differences in math scores, but they suggest that by improving female spatial abilities could possible reduce the gap for scores between males and females.

1.6.3 Geoscience and Earth Sciences

Pallrand and Sieber [1984], in their comparison of STEM fields, found that Geoscience students have the second highest spatial abilities. Black [2005] created a concept inventory called the Earth Science Concept test (ESC) and did a correlation study with different types of spatial ability. The ESC is a 20 question multiple choice assessment with questions focused on the following topic: causes of the seasons, moon phases, tides, phase changes as related to KMT, geologic blocks diagrams, map projection, movements in the Sun/Earth/Moon system, relative distances within the universe, interpretation of aerial photos and topographic maps, relative distances to the magma that produces features such as volcanoes, and three-dimensional drawings of weather fronts and air masses. The sample population polled for this study were non-science students from a midwestern public university. Theses students were mainly pre-service teachers taking courses in the departments of geoscience, chemistry, physics and biology. Of the three different spatial abilities tested, Black [2005] found that mental rotation was the best predictor on the ESC. The mental rotation test used was the Purdue Spatial Visualization Test: Rotation (PSVT: R, see Figure 1.6) [Guay, 1977]. The PSVT: R is similar to the MRT but uses a different type of figure that is rotated.

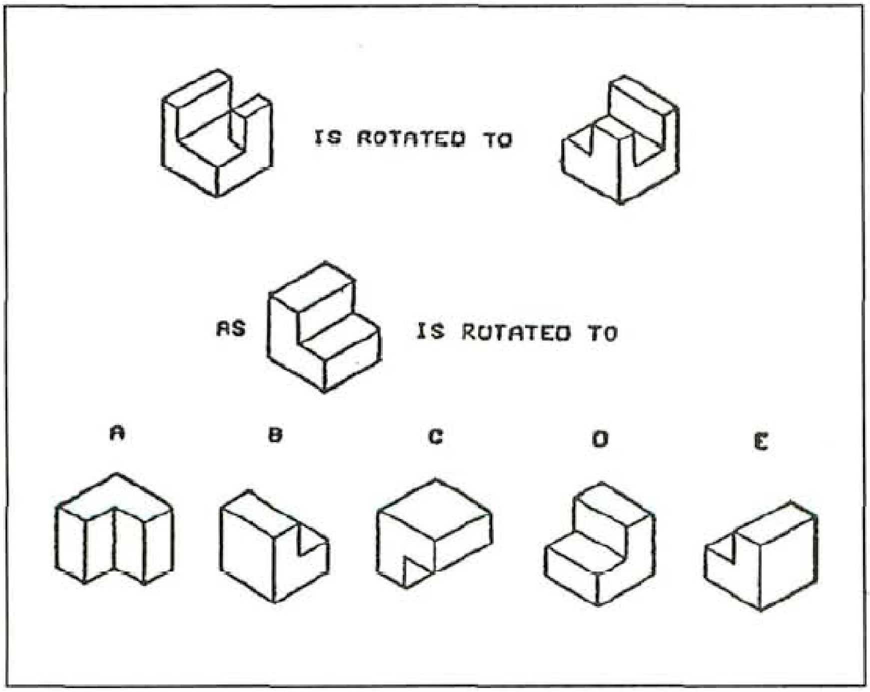


Figure 1.6: Example of PSVT: R. Figure was taken from Purdue Spatial Visualization Test: Rotation [Guay, 1977].

In terms of correlations between spatial-visualization abilities and success in Geoscience course, Orion and colleagues [1997] found a high correlation. Orion and colleagues took a sample of first-year geology undergraduates who completed a one-year introductory geology course in the Hebrew University of Jerusalem. These students also took courses in Earth Sciences, Chemistry, Physics, and Mathematics. Though there is research that suggests all of these fields have some correlation with spatial abilities [e.g. Black, 2005; Stieff 2005; Kozevnikov and Hegarty 2007; Casey et al., 1995] Orion and Colleagues [1997] claim that these courses did not attribute to the correlation between spatial visualization skills and achievement in geology. Achievement in this study was measured with final grades in the one-year geology course[1].

1.6.4 Engineering

The strongest support of correlations between spatial abilities and success in STEM courses comes from the work done by Sheryl Sorby in engineering at Michigan Technological University (Michigan Tech). Sorby has been primarily interested in improving the retention rates of female students in engineering [Sorby, 1999] and has suggested that the gender differences in spatial ability are one reason for the attrition of females in engineering. Sorby and Baartmans [1996] created a spatial visualization course for students who scored low on the PSVT: R (see Figure 1.6). Of those female students who optionally enrolled in the course there was a retention rate of 63.6% as opposed to 53.1% for those students who did not enroll in the course. For male students who optionally enrolled in the course there was a retention rate of 69.2% as opposed to 62.5% for those students who did not enroll in the course.

When Michigan changed from a quarter system to a semester system, the course that was originally created was restructured [Sorby, 2005]. One of the main critiques for implementing a course like this at other institutions was the time commitment required by faculty. The new course significantly reduced the number of faculty involvement hours (20 hours in a 10 week session to 4 hours in a 15 week session) by using undergraduate T. A.'s and NSF funded developed software and lesson plans [Gerson et al., 2001]. With the new course, similar retention rates were recorded for both males and females, but the new data also did follow up comparisons of grade point averages of students who scored low on the PSVT: R and enrolled in the course and students who did not enroll in the course. Sorby [2005] found that students who optionally enrolled in the spatial course had higher grades in their introductory Engineering courses, Calculus 1, Chemistry 1, and Physics 1, with the highest gains in the

[1] Though the authors did find a high correlation between achievement in a geoscience course and spatial abilities, the authors noted that their findings could not be generalizable to larger populations because of the unique characteristics of their population.

Engineering courses[2]. On top of improving course grades, through longitudinal studies, Sorby [2009] found that retention rates of both males and female engineering students who enrolled in the spatial course drastically improved.

1.7 Discussion

In Physics Education Research there are a lot of factors that can be considered when trying to improve student comprehension and retention. Our group has an interest in studying how students' visual spatial skills correlate with comprehension in physics and closely related fields. As previously stated, physics has a variety of topics that are abstract and spatial in nature. Students are forced create visual mental representations of visual, graphical, conceptual topics and manipulate those representations. This manipulation, such as rotation, can increase the cognitive load and introduce new misconceptions, thereby preventing the students from coming to the right conclusions. With the previous work done in other STEM fields, particularly Sheryl Sorby's work at Michigan Tech, we have devised several studies to test whether our physics students also have correlations between their spatial abilities and comprehension in subject matter.

[2] Sorby [2005] noted, however, that even though students who voluntarily enrolled in this course earned higher grades in their STEM courses it is unknown if these gains were truly due to the course or if these students were more dedicated to their studies in general.

CHAPTER 2

COLOR AS A THIRD DIMENSION[3]

2.1 Introduction

This chapter presents material from one of two studies [Cid, Lopez, and Lazarus, 2009] published in peer-reviewed journals, and which form a significant fraction of the research presented in this dissertation. This study focused on the difference between novices and experts and the misconceptions that can arise with visual representations of information, in this case the use of color to indicate a third dimension on a two-dimensional display. This particular study arose from our participation in an interdisciplinary course (discussed in detail in section 2.4.3) at Florida Institute of Technology. Chapter 1 laid the conceptual foundation for this research. This chapter is one of several discrete investigations that focuses on visual cognition and spatial abilities.

2.2 Background for Study

As was described in Chapter 1, there are differences between how an expert comprehends information and how novices comprehend information. As experts, we are trained to understand color schemes used in visualizations in our respective scientific fields. Unfortunately, we tend to forget how complicated graphics can be when viewed for the first time. While visual representations of concepts and information can be a powerful tool for learning science under the proper circumstances [e.g., Mayer, 1997], understanding visual representations is a learned skill [e.g., Land and LoPerfido, 1993; Beichner, 1996].

The representations of complex three-dimensional objects using two-dimensional

[3] Information and text from this chapter has been taken from Cid, X. C., Lopez, R. E., and Lazarus, S. M., 2009, "Issues Regarding Student Interpretation of Color as a Third Dimension on Graphical Representations", *Journal of Geoscience Education, v. 57(5), pp. 372.*

representations can prove problematic for students [Lopez and Hamed, 2004], as can the relationship between maps and what they represent [e.g., Kastens et al., 2001; Liben, 2006; Rapp et al., 2007]. Sometimes a graphical representation of a system can introduce misconceptions among students, as in the case of what causes the seasons (as documented in the video *A Private Universe*), or in the case of student understanding of the spatial extent of electric and magnetic fields in plane, transverse electromagnetic waves [Ambrose et al., 1999]. In previous studies in both physics and geography, misconceptions occur because a student is missing critical information about a particular problem that results in a mental model based on irrelevant features, which ultimately leads to an incorrect solution [e.g., Anderson and Leinhardt, 2002].

In university classrooms, professors routinely present images to aid in the explanation of abstract ideas as well as depend on textbook images to help communicate ideas. Regardless whether these visual representations are used in a manner that might be deemed effective [e.g., Mayer, 1997], issues remain. For example, individuals who generate these images are subject matter experts, who have lost the ability to see things as a novice might see them [Mestre, 1994; *How People Learn*, 1999]. In addition to having more content knowledge than novices, experts possess knowledge that is organized and stored in memory in such a way as to facilitate quick access to sophisticated problem-solving strategies [Anderson and Leinhardt, 2002]. As discussed in section in section 1.4, Sweller [1994] describes that experts and novices also chunk information differently. Experts have experience that allows them to recall stored information from their long-term memory. Novices do not have experience or practice to have detailed information stored in their long-term memory. Thus, what might seem self-evident in a visual representation to the expert might not be so obvious to the novice [Rapp et al., 2007].

In this study we examine a commonly used technique: the use of color to indicate a third dimension. From discussions in section 1.5 we know that visual information presented in three dimensions increases a student's cognitive load and can introduce misconceptions that are not content based.

In the geosciences, topographic maps are frequently used to depict landforms, atmospheric geopotential heights, temperatures, etc. On most topographical maps, contour lines provide two main types of information. Quantitative information is provided about absolute height of the land, qualitative information can be obtained about the shape of the landscape by visually integrating the contour lines [Barrell and Cooper, 1986], and it is well known that novice students often have difficulty with such representations [e.g., Rapp et al., 2007, Kastens et al., 2001]. More over the shape of an object seems to be the most important feature in recognizing an object and, presumably a representation [Biederman and Ju, 1988]. Therefore, color might provide secondary information to a novice viewer while providing primary information to an expert who might not even recognize that he/she has categorized an image due to shape.

Often, these maps depend on the use of color to represent vital information. In this study, we looked at the use of color to determine if there is a preference in color schemes when rendering a three-dimensional geographic landform onto a two-dimensional surface. We also examined the use of color as a representation of temperature, because 1) temperature is generally represented with red being high and blue being low, and 2) this color scale is often used within different scientific fields to express a scale for other quantities. Thus scientists generally use a color scale where red is high and blue is low and they probably do not think much about it when viewing figures that use color to represent a quantity.

In fact, most users of the red/blue color scale probably think of it as a natural representation, similar to bars and lines that indicate an increasing or decreasing quantity. Some elements of "cognitive naturalness" appear to be supported by experiments with, for example, the slopes of lines being associated with trends and bars being associated with discrete quantities [Zacks and Tversky, 1999]. However, in an unfamiliar setting, the use of color might be a less obvious and natural representation and more of a "cognitive collage" as described by Tversky [1993]. Some information may be distorted by the perception/interpretation of the viewer, a perception that may be grounded in some feature of the representation that was not intended by the creator. On the other hand, there is evidence that students who have experience with both contour and grayscale representations of temperature do better on standard contour map assessments [Taylor et al., 2004]. Regardless, this does not answer the question of how “natural” such representations might be to the novice viewer.

In particular, the widespread use in science of the red/blue color scale raises the question of to what extent do novices transfer a color scale from one domain to another. Are there other visual clues in a representation that might dominate a novice interpretation of the representation, despite the use of a color scale that is seen as “natural” by the expert? As experts, do we overlook the fact that novices might have a different, embedded understanding of what the colors imply? Do these notions transfer between different representations? We address these issues in this study by examining students’ responses to the use of color as a scale height (elevation) on a basic topographic map, as well as the use of color as a representation of temperature.

2.3 Methodology

The target population for this study included undergraduate students taking a physical science course. The accessible population included undergraduate physics students enrolled in a small private university in east central Florida. The sample population for this study was drawn from undergraduate students enrolled at Florida Institute of Technology (Florida Tech), a small (2,500 students), independent, technologically oriented university on the Atlantic coast just south of Kennedy Space Center. During the spring and summer semesters in 2007, we administered surveys regarding the use of color to fifty-four undergraduates from two different physics courses and a Research Experience for Undergraduates (REU) program funded by the

National Science Foundation (NSF) (which brought students from other universities to Florida Tech to do research for a six-week period in the summer of 2007). While the data were collected at Florida Tech, the bulk of the analysis, the writing, submission, and revision of the paper resulting from this work [Cid, Lopez, and Lazarus, 2009] was done at UT Arlington.

The method used for selecting the sample was semi-convenient. We used two intact classes from the physics and space science department, in addition to four REU students. The intact classes used were an introduction to physics course composed mainly of first year undergraduate students from a variety of STEM majors, and an electromagnetic theory course composed mostly of junior and senior physics majors. The REU students were also STEM majors. In total, there were fifty-four students in the sample, with ages ranging from 18-39 (even though there is a large age range, only three students were over the age of 21). Of the Florida Tech students, thirty were Freshmen, five were Sophomores, nine were Juniors, and six were Seniors. The REU students consisted of one Sophomore and three Juniors.

Fifteen of these students were female, which is 27.7% of the sample. According to statistics posted on the American Institute of Physics website, in 2005, about 22% of Bachelor's degrees in physics were earned by females. So, even though male students outnumbered female students, statistically our sample was consistent with the national average of females obtaining their Bachelor's in physics. It should also be noted that we conducted this study without taking into account if a student was colorblind, however, no students raised this issue.

A sample size of 54 students is sufficient for a random sample distribution [e.g., Gravetter and Wallau, 2007]. However, this sample is too small to subdivide by gender or by year in school and still have statistically viable sub-samples. Therefore we will consider the behavior of the population as a whole. In the future, studies with larger data sets may be conducted to determine if there are gender differences, for example, in the way that students approach the use of color scales to interpret information given in visual representations. However, those kinds of questions are beyond the scope of this study.

We used two topographic surveys and a temperature survey. We also collected demographic information from the participants (i.e., age, gender, major, year in school) as well as their prior experience with map reading. Both topographic surveys were based on a commonly used color map, namely one in which red indicates high values of a quantity and blue indicates low values of a quantity [e.g., Tufte, 1997]. This color scale is often used in many fields when a third dimension is required, such as in space physics where the flux of energetic particles hitting a detector on a spinning spacecraft might be color coded, with the X and Y axes representing time and the direction in which the detector is pointing at a given time on a spinning spacecraft, respectively [e.g., Lopez et al., 1993].

The first survey comprised seven questions and one figure. Basic questions included whether or not the subject knew what a topographic map was, and whether they had ever used one in the past. The participants were then asked to indicate high and low points on the figure, and to identify the landform. The figure depicted the topography of a basic volcano (see Figure 2.1). The volcano was created using nearly concentric rings with slight variations in some of the rings to create a small amount of complexity. Instead of using a numbering system to differentiate the elevations, we used a basic coloring system, which was composed of the following colors: dark blue, blue, light blue, very light blue, dark green, light green, yellow, orange, and red. The dark blue of the innermost ring represented the deep depression on the top of the volcano. We deliberately omitted a legend because we wanted to evaluate how students would orient themselves to the color scheme.

The figure used in the second survey was identical to the first but the color scheme was reversed. The outermost portion of the figure was orange, then yellow, light green, dark green, very light blue, light blue, blue, then dark blue. The red of the innermost ring indicated a spike in the middle of a crater (see Figure 2.2). Again, the legend was intentionally omitted and will be discussed in more detail later.

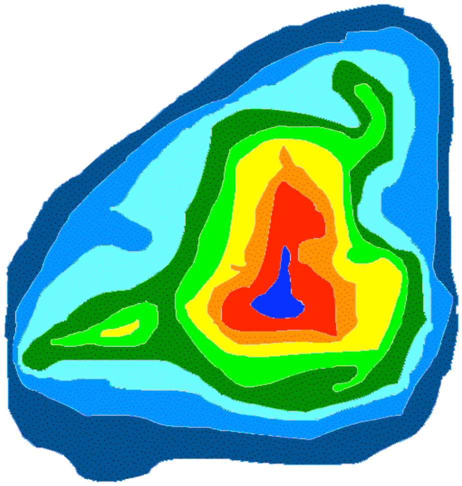


Figure 2.1: Topographic Map Depicting a Volcano. Figure taken from Cid et al. [2009]. If blue indicates a low value and red indicates a high value, this image would depict a volcano. The blue in the center would be the depression in the center of a volcano.

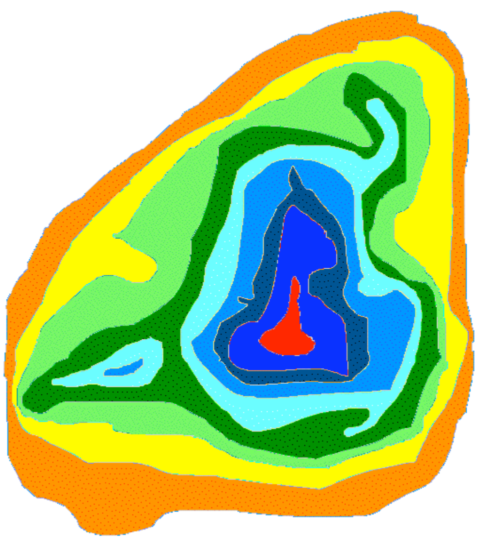


Figure 2.2: Topographic Map of a Crater with a Spike at Center. Figure taken from Cid et al. [2009]. If we maintain the same color scheme as figure 2.1, blue indicates a low value where red indicates a high value; this image would depict a crater with a spike in the center similar to those found on the moon.

Half of the students, when presented with the first survey, were given Figure 1 while the other half of the students were given Figure 2. In the second survey the order was reversed, i.e., students who received Figure 1 in the first survey were then given Figure 2 and vice-versa. The third survey was a temperature survey in which the subjects were asked what blue and red indicated in terms of temperature, why, and where they learned this association. There were no figures associated with this survey [Please see appendix for surveys].

Of these fifty-four students, eight students were chosen to participate in follow-up interviews (conducted by author) designed to investigate in more detail the responses provided. The interviews were structured using a think-aloud protocol in which the students were asked to explain their particular survey answers. The interviews, which lasted between 16 and 22 minutes, were recorded and viewed by members of the study, who met several times to discuss their findings and to come to consensus. Because of the qualitative nature of the interview data, a Grounded Theory approach was followed [e.g. Strauss and Corbin, 1990].

Grounded Theory is a qualitative approach to data. Grounded Theory has a backwards approach to comprehension than the scientific method. The first step in a grounded theory approach is to collect data and then to allow the overall theme to guide the researcher. Grounded Theory methodology allows for a theory to develop by reviewing data that is collected. No preconceived hypotheses or ideas are generated before data, including interviews, are collected, but as a hypothesis develops, the original data are searched to provide evidence to support the hypothesis. Once a theme starts to emerge, the research can reprocess the data and can apply quantitative methods if needed.

After analyzing the interviews several times with the other two members of the study, we developed our interpretation of the students' responses. Once we had come to consensus about what the major findings were, we went back through the data (both interviews and responses on surveys) to find supporting evidence to support our conclusions. In light of this approach, the findings are presented first and then discussion of data that support those findings follows.

2.4 Findings and Supporting Data

This section has five findings that will be discussed in following separate subsections.

2.4.1 Finding 1: Preconceived Interpretation of Figures

The first finding that we had is that when viewing the topographic map, the students had a preference to see a hill. The two topographic maps used in this study were created to represent a volcano and a crater with a spike in it. Depending on the choice of a color scheme for encoding height information (red = high or red = low) one could decide which image was the volcano and which was a crater, assuming the same color scheme was used for both images [please see Appendix A for questions regarding interpretation of figures].

The question, "What might you call such a landform if you saw it in real life?" led to a variety of answers and supporting reasons as to why the students referred to the landform in a particular manner. If, in their responses, students mentioned a type of figure that rose out of the ground and came up to a point, then we called that figure a "hill". For instance, students referred to the figure as a volcano, mountain, hill, mesa, and mound. Instead of classifying the figure as a volcano (as was intended by the members of the study) we decided to classify the figure as a hill because there were more students who referred to the figure as a hill instead of a volcano. For this same reasoning we classified Figure 2.2 as a depression instead of a crater (as was intended by the members of the study).

In the first survey, 19 out of 27 students who viewed Figure 2.2 first called the landform a hill. Of the students who viewed Figure 2.1 first, 22 out of 27 called the landform a hill. In the second survey, 16 out of 27 students who viewed Figure 2.1 second called the image a hill. Of those students who viewed Figure 2.2 second, 11 out of 27 called the image a hill. Please note, these numbers are total numbers and do not give information regarding specific students or color scheme used. An interesting note, 16 students maintained a constant color scheme, meaning they chose a color scheme based on the first image that they viewed and proceeded to stay with that color scheme when viewing the second image. Of those students who did not maintain a consistent color scheme, 21 students called both images a hill. This finding suggests that some students did not use the color scheme to interpret the image, rather they created the image first in their mind and then imposed conditions on the color schemes to support their original idea of what the figure should represent. When presented with something they are told is a topographic map, most students assumed that it was a hill and created a color scheme to fit.

Because only 16 students maintained a consistent color scheme, it indicates students were using other cues to interpret the figures, and thus the number of responses identifying a hill versus a depression was unequal. Even if the students were not applying a consistent color scheme, because we presented half of the students with Figure 2.1 first and Figure 2.2 second and the other half of the students Figure 2.2 first and Figure 2.1 second, we would have expected a consistent interpretation. However, the students' responses did not correlate with the color scheme of the figure in a statistically significant fashion. This reinforces the conclusion that students are not utilizing the color scale in these images to the extent one might naively expect. Instead, other cues drove the interpretation.

The idea that the color mapping was actually a secondary process was quite clear in several interviews. For example, consider the following interview with one particular student (referred to here as Student 1) who saw Figure 2.2 first and considered the landform to be a volcano.

> ***Interviewer:*** *[3:02][4] How did you decide that blue was high elevation?*
>
> ***Student 1****: [3:08] Um…To be perfectly honest I…It was almost an arbitrary decision. I just looked at it and said 'Ok, this looks like this is coming up to a peak.' At first glance the colors really didn't…um…affect my decision, but as soon as I had it in my head that…ok this is coming to a peak, then I just associated the colors to the different elevation.*

[4] The numbers in square brackets indicate the time [min:sec] in the interview.

When the student was given Figure 2.1 and asked what landform was being represented, the response was the same, a volcano.

> ***Interviewer:*** *[11:12] How did you use the colors to help you?*
>
> ***Student 1:*** *[11:16] Well…um…again once I looked at it and used the boundaries to…kinda, get a mental picture of what I thought it was…um then I guess I automatically associated colors to different elevations and so…in this one [Figure 2.1], I just associated my [pause] my dark to light blues as lower elevations and then my greens to mid-level and then just getting higher with yellow orange and red.*
>
> ***Interviewer:*** *[11:56] And what gave you that association that blue was low in this case and the oranges and reds were high How did you choose the colors of blue low and orange high in this figure versus the way you had them in figure one [Figure 2.2]?*
>
> ***Student 1:*** *[12:14] Um…I think I associated the colors to elevations after I got an idea of what it was. So I didn't really use the colors to…figure out what the actual feature was until I actually got a picture of it in my head and um…then it's like "ok…well if this is a hill then blue must be low, red must be high." And then I looked at it and…well if blue is low and I've got blue at the tip the maybe it's not a tip, maybe it's a divot. And so I…I kinda used the colors as a second guess or like a…like a…like a second check.*

Student 1 mentioned that when Figure 2.1 was presented, there was an inclination to use the same color scheme as Figure 2, but once the student realized that the figures were the same, the colors were used as a secondary guide. When prompted to use the same color scheme used in Figure 2.2 for Figure 2.1, this is what was said:

> ***Interviewer:*** *[14:46] If you do use the same color scheme for figure two [Figure 2.1] what do you think that [Figure 2.1] landform would be?*
>
> ***Student 1:*** *[14:52] I think it would be the exact same thing.*
>
> ***Interviewer:*** *[14:55] Even with the same color scheme as orange being, the orange and warmer colors being high and the blues being lower?*
>
> ***Student 1:*** *[15:02] Mm Hm [yes]*

2.4.2 Finding 2: Preference for Figure Order

The second finding that we had is that our subjects had a preference for the figure they saw first. Remember half the students were given Figure 2.1 first and then Figure 2.2 second, whereas the other half of the students were given Figure 2.2 first and then Figure 2.1 second.

The reason for this was to eliminate any bias the figures might have.

The hypotheses for the quantitative aspect of this study is that there is no preference for either figure, where "preference" is defined by a student response to the question "Which of the two figures did you find easiest to understand?" If there were no preference then we would expect that half of the students would choose Figure 2.1 and half would choose Figure 2.2. Therefore the statement that preferences in the population are equally divided among the figures is our null hypothesis, referred to as H_0.

H_o:

Figure 2.1	Figure 2.2
0.5	0.5

Our alternative hypothesis, H_a, is that preferences in the population were not equally divided among the figures.

In education research it is common practice to state research questions in the form of null hypotheses. We want to test if our intervention has a statistical effect on our sample, but there are multiple outcomes for a specific effect. Instead of testing each effect individually, we test the null hypothesis. If we can statistically show that the null hypothesis has failed, we can reject the null hypothesis and accept an alternative hypothesis. By rejecting the null hypothesis, we are stating that we are confident that our results are not due to chance. Our confidence level is defined by alphas and reported as a "p" value. So as long as our "p" value is less than 0.05 than we can say that we are 95% (or above) confident that we have viable results.

The following data indicate student preference for figure by color scheme (Figure 2.1 vs. Figure 2.2):

Table 2.1: Student Preference for Figure by Color Scheme[5].

	Figure 2.1	**Figure 2.2**
Observed Frequencies	28	21
Expected Frequencies	27	27

Because there was not a significant result with χ^2 (1, n = 54) = 1.367, p = .24, there is insufficient evidence to reject the null hypothesis. Thus, there is insufficient evidence to warrant the rejection of the claim that preferences in the population are equally divided among the figures and so they did not have a preference for Figure 2.1 versus Figure 2.2.

Although students did not appear to have a preference for the color scheme used, the majority of students (37 out of 54 students) claimed that the first figure that was viewed was the easiest to understand. To demonstrate this, we examine the hypothesis that students had no preference for the first figure viewed versus the second figure viewed when asked which was easier to understand. Again, our null hypothesis would be that half of the students would claim that the first figure viewed was easier to understand and half would state that the second figure viewed was easier to understand.

H_o:

First Figure Viewed	Second Figure Viewed
0.5	0.5

Our alternative, hypothesis, H_a, is that preferences in the population were not equally divided by order of presentation. Table 2.2 presents the data for the preference indicated by the students.

[5] Please note there were five abstentions

Table 2.2: Data for Student Preference in Order of Presentation[6]

	1st Figure Viewed	2nd Figure Viewed
Observed Frequency	37	12
Expected Frequency	27	27

Because there was a significant result with χ^2 (1, n = 54) = 12.03, p = 5.2E-4, there is enough evidence to reject the null hypothesis and accept the alternative hypothesis. The Chi-Squared test indicates that there was statistical evidence to support that students preferred the first figure they viewed.

2.4.3 Finding 3: Distractions

Our students focused on and became distracted by aspects of Figure 2.1 and 2.2 that were unintentional or seemingly unimportant to the instructor. In order to explain this finding, an explanation of how this study came about is first needed.

A pre-study, conducted to investigate the use of visual representations in geoscience education, underscored the importance of figure color. We observed student reaction to the use of a number of visual representations in the Whole Earth Course (WEC), an interdisciplinary course taught at Florida Tech by six professors on aspects of the Earth system. WEC has been the subject of previous research in student learning due to the interactive qualities of the professors and students [Eason, 2000]. WEC covers six different subjects including the Biosphere, Atmosphere, Anthroposphere, Cosmosphere, Geosphere, and Hydrosphere, each of which are taught by a different professor from their respected field. During one of the lectures, a professor presented a slide on global sea surface temperatures. The slide was a simple graphic intended to illustrate that the ocean surface temperatures were, in general, warmer near the equator than at the poles. In order to illustrate this, contour lines were labeled according to the

[6] Please note there were five abstentions

temperature (in degrees Celsius) along with color bands between the contours, which also represented temperatures. Unfortunately, there was a 'gray' color filled region in the middle of the image that created some confusion in the classroom. The 'grey' color was used to represent two separate temperature ranges. Even though the contour lines were clearly labeled, the students thought that the two 'grey' color bands represented the same temperature. The out of order 'gray' color demonstrated how unimportant information to an instructor can be salient in the mind of a novice.

Just as there was confusion produced by a small discrepancy in the WEC figure, our study presented us with a similar issue. We created simple topographic diagrams, without a legend and with colors to differentiate the different heights. When we printed the images out, the printer that was used created dots that were not present when we viewed the figures on the computer screen (see Figure 2.1 and Figure 2.2). The different shades of blue, the different shades of green, and the orange had dots in their color bands. As the creators of the figures, we knew that the dots did not mean anything and did not pay attention to their presence.

During the interviews, however, it became obvious that the dots may have created confusion for the students. Four out of the eight students interviewed mentioned the presence of the dots. Some of the students simply asked if the dots were significant, and upon our reassurance that they were insignificant, they moved on with the discussion. One student tried to interpret the dots as part of the representation of varying height in the figure. The student looked for a pattern in terms of which colors had dots and was confused when no clear pattern could be discerned. Another student (referred to here as Student 2), when asked what was confusing about the surveys, specifically mentioned the dots as a source of confusion. For example,

> ***Interviewer:*** *[22:57] Was there anything confusing about the survey overall? In terms of colors, in…in those diagrams, in terms of temperature?*
>
> ***Student 2:*** *[23:03] Just the diagrams because when you have…just because*

there are dots and there are no legends to follow and you don't have…know what the dots could possibly represent so you end up just making a picture of your own and hoping for the best.

Because the dots were unintentional, we did not include a question in the surveys about the dots. The student interviews demonstrated the importance of perceptually salient, but conceptually unimportant details when a novice observes an image for the first time. Because half of the students interviewed mentioned the dots, we speculate that many other students, who were not interviewed, probably had some issues or questions about what the dots represented. Since experts tend to ignore conceptually unimportant information, it is possible that they could construct a visual representation that inadvertently contains a distraction for novices, as in our case with dots, or in the case of the figure in the Whole Earth Course.

2.4.4 Finding 4: Embedded Color Scheme

It is known from studies in cognitive science, that the transfer of information from one domain to another domain is extremely difficult [*How People Learn,* 1999]. With this in mind, we asked whether or not our students would transfer their experience with temperature and color to scale height. Our subjects had an embedded color scheme for color when representing temperature, with red being high temperature and blue being low temperature. However, they did not automatically transfer this idea that red represents 'high' altitude and blue represents 'low' altitude to the colored topographic map.

As one might expect, the majority (48 out of 54) of students associated blue with low temperature and red with high temperature. This association is often formed early, as one student said in an interview *"When I was really young, when you are learning your colors"* in answer to the question *"When did you learn this?"* Interestingly enough, that student also referred to sunburns as red, showing that the association can extend to other things perceived to be "hot", even if temperature is not really a factor. However, as we have seen above, the association that blue represents low and red represents high did not automatically assert itself when students were confronted with a colored, topographic map. In fact, the preference to see a hill, and a preference to use whichever color scheme matched the first image viewed were more important to the students than the temperature color map.

Astronomers, on the other hand, reverse the popular color scale when representing temperature. This is a more accurate interpretation of the physical representation because blue photons have shorter wavelengths than red photons and thus blue photons are more energetic than red photons. Some students adopt this color scale for temperature if they have had exposure to modern physics (photons) and/or astronomy. In our sample only a few students (6 out of 54) indicated that blue was hot and red was cold. One student (in the interview) explicitly stated that blue photons have higher energy than red photons, a concept the student claimed was learned in high school. But this reversed color scale interpretation didn't appear to transfer to our topographical images since the two students (out of six) who said that they actually used the astronomical color scale for temperature did not transfer the idea that red represented 'low' altitude and blue represented 'high' altitude to their interpretation of the figures. However, as expected, a large majority of the students have an embedded color scheme for temperature whereby red represents hot and blue represents cold. Despite this, we found no evidence that there was an embedded color scheme when color was applied to height, as discussed above. The primary factors in the interpretation of the color information were the preference to see the hill and the preference for whatever figure the student saw first.

This issue of seeing what you want to see is one that is very evident in the literature on expertise [*How People Learn*, 1999]. In section 1.4 it was discussed that experts have the ability to chunk information in a way that allows easy access from long-term memory. The chunks of information act as single bits of information in working memory. From practice and experience, experts have the ability to develop conceptual frameworks for organizing information, and the ability to confront seemingly disparate facts and to organize them and make sense out of them. This is one of the main benefits of expertise. At the same time, expertise makes it difficult to see things that are not in accord with expert paradigm and it also makes it difficult for experts to see things as a novice would. This issue, of course, extends beyond a narrow discussion of expertise and touches on broader issues of paradigms and paradigmatic change in science [Kuhn, 1996]. From the point of view of instruction, when dealing with non-expert or novices, there may exist, for what ever reason, a hidden paradigm or conceptual framework that drives student understanding of a visual representation and inhibits the communication of what the expert had intended; the expert having forgotten what it was like to think like a novice.

2.4.5 Finding 5: Previous Experience Has no Effect

In Survey 1 Part A[7], students were asked if they had experience with topographic maps. Based on that self-reported data, we concluded that previous experience with topographic maps does not seem to have influenced the results.

The surveys included a question asking the subject to report what experience they had had previously with topographic maps. There were three students who reported that they used topographic maps quite a bit. Two of those students used a consistent color scheme. There were 13 students who said they sometimes used a topographic map (i.e., skiing, camping, orienteering, and a few said they had used topographic maps in a previous class). Of these 13 students only 4 used a consistent color scheme. Of the remaining students, 33 students said they rarely used a topographic map. Of the 33 students who reported that they had rarely used a topographic map, 10 students used a consistent color scheme. We recognize that these are very small samples and not statistically sufficient to establish normal distributions. However, there does not seem to be anything suggestive in the data to indicate that the 16 students who reported experience with topographic maps were significantly more disposed to use a consistent color scheme to interpret topographic information than the students who did not report such

[7] Surveys can be found in Appendix A

background knowledge.

2.5 Discussion

Color is widely used as a means of representing a third dimension for two-dimensional figures. However, students, as novices, might not necessarily impose the most widely used color scale, where red is a ‘high’ value and blue is a ‘low’ value of the quantity represented. Other cues, such as the representation being a topographic map, may elicit other responses, such as a preference to see a hill. In fact, our subjects tended to adapt whatever color scheme they were given to what they wanted to see. Also, aspects of a figure (such as a misplaced color or a printer artifact) that would be typically ignored by an expert may loom large in the mind of a novice. These results underscore the importance of carefully constructing and testing visual representations, the pitfalls of erroneous image association, and the false presumption that certain elements of the images such as an implicit color scale will automatically be applied in the manner intended by the instructor.

Our results have some obvious implications for instruction. In geoscience and space science education, color is often used to encode information. It is generally assumed that students will be able to process such information appropriately using the “red = high, blue = low” color scale. This study shows that that is not necessarily the case and that other visual cues may predominate in the minds of students. Moreover, it can be assumed that color was less important. Shape was the more dominant cue, as might be expected in general studies of object recognition [Biederman and Ju, 1988]. Thus, instructors should take care to determine that students are properly interpreting the representation, perhaps by including a conceptual question based on the representation in a peer-instruction sequence [Mazur, 1998] in class. Also, when creating a new classroom representation using color to convey information, a test with a few students might be a useful exercise. Once well established, the use of a color scale along with contours in a representation should result in a more robust understanding on the part of students [e.g., Taylor et al., 2004], but until it is clear that students are properly interpreting the color scale, and using it as a primary information source, one cannot be certain that this will be the outcome. One specific item we encountered was the preference by students to see a “hill” when confronted by a topographic map. This should be of interest to geoscience educators who use topographic representations, whether they are of landforms or not. For example, atmospheric science uses topographic maps. One common example is the map of the altitude of the 500 mb level (approximately the midpoint of the atmosphere). One of the collaborators (S. M. Lazarus from FIT) has repeatedly encountered difficulty and subsequent confusion experienced by students as they try to grasp the concept of a sloping (i.e., three-dimensional) isobaric surface. It could be that an embedded desire to see a “hill” is causing an obstruction to learning the illustrated concept. Such cognitive conflict could occur at other times when contour representations are used to illustrate complex concepts that might seem straightforward to the professor, but which mystify students. The desire to see a “hill” might be part of the mystery. Further research on this topic seems warranted. However, based on our results we suggest that instructors using contour representations of any quantity be aware of the possible unintended misconceptions that might arise in the minds of novices.

We as experts also tend to have this idea that our students will be able to process the visual diagrams that we use, but as it was shown, students might be relying on other cues for their own interpretations. These other cues have the ability to manifest themselves in an incorrect mental model. Lunar phases are a classic example of incorrect mental models produced by novices. The next chapter is devoted to student misconceptions about lunar phases.

CHAPTER 3

REPRESENTATIONS OF LUNAR PHASES[8]

3.1 Introduction

This chapter presents the research that formed the second published study from this dissertation [Cid and Lopez, 2010]. Students have significant difficulties comprehending the abstract nature of Lunar phases and there have been studies that suggest that concepts that are 3-dimensional in nature have the greatest effect on student comprehension if it is presented in 3-dimensions. The idea is that if you are reducing the cognitive load devoted to mentally creating or manipulating the concept, then the students can devote more energy to comprehending that material. This chapter explores this idea.

3.2 Background for Study

Physics and Astronomy have topics that are highly spatial in nature (i.e. Electricity and Magnetism, Seasons, Lunar Phases, etc.). These topics are usually presented using 2-D representations, some of which might reinforce preconceived ideas or introduce new misconceptions about the 3-D system [e.g., Ambrose et al. 1999]. As discussed in Chapter 2, students often have preconceived understandings of diagrams used in class instruction, which might introduce and propagate misconceptions. A classic example of this problem is seasons, where textbook artists try to illustrate the path of the Earth around the Sun and often draw elongated ellipse to present the Earth's elliptical orbit from a perspective view (see Figure 3.1). It is possible, then, that this 2-D representation has now introduced a misconception that the Earth's orbit is an elongated ellipse when, in fact, it is almost circular [DeBuvitz, 2009]. In

[8] Information and text from this chapter has been taken from Cid, X. C. and Lopez, R. E., "The Impact of Stereo Display on Student Understanding of Phases of the Moon", *Astronomy Education Review,* 2010, *v. 9(1),* 010105, 10.3847/AER2009044.

physics, there is evidence that the mental manipulation of 3-D images plays a role in increasing cognitive load when dealing with magnetism and the relationship to currents [Lopez et al. 2004]. Similarly, in astronomy there are many topics in which 3-D relationships are crucial to understanding the concept, thus we would expect that the manipulation of mental images would be an important source of cognitive load when trying to acquire a concept or accomplish a certain task that required the application of that astronomical content.

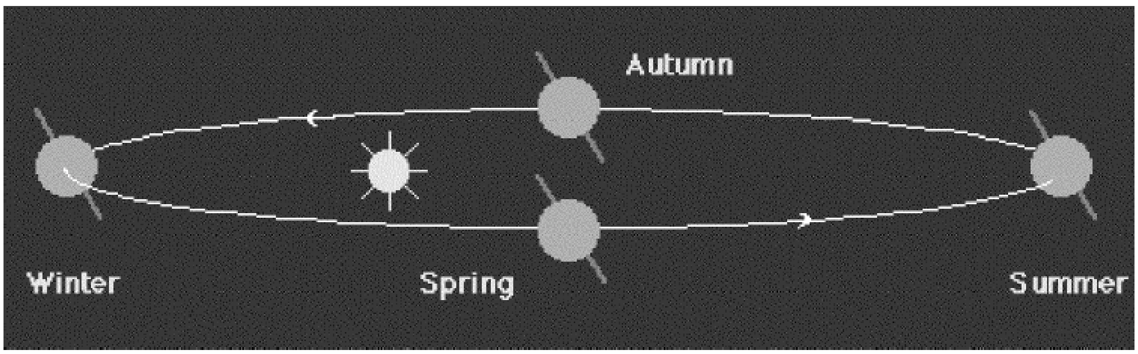


Figure 3.1: Example of Elongated Elliptical Orbit. This is an example of a typical depiction of seasons showing that the Earth is actually closer to the Sun during winter, therefore the seasons cannot be cause by the Earth's distance from the Sun. However, the dominant perceptual feature of the diagram is that the orbit of the Earth is an elongated ellipse.

In this study we examine student understanding of the phases of the Moon, the impact of instruction in a single lab on that understanding, and the importance of an explicitly 3-D representation of the topic. Misconceptions about phases of the moon by children and adults alike have been a topic of study for decades [e.g., Schnepps and Sadler 1987; Trumper, 2000; Trundle et al., 2002; Lindell and Olsen 2002; Bailey and Slater 2003]. Given that the 3-D spatial relationship between the Earth, Moon, and Sun is critical to understanding the phenomenon of phases, one might *a priori* consider that instruction that provides a 3-D view of the system would hold an inherent advantage in instruction.

Pedagogy [e.g. McDermott, 1996], based on active engagement techniques, has shown to improve student comprehension in phases of the moon [e.g. McDermott 1996; Trundle el al, 2002; Trundle et al., 2007], but it has been suggested that computer simulations can be more effective than some aspects of interactive experiences [e.g., Bell and Trundle 2008; Winn et al. 2006]. More specifically, it has been suggested that 3-D computer simulations can increase student comprehension of the phases of the Moon [Hansen et al. 2004; Küçüközer 2008; Küçüközer et al., 2009]. One might expect such an outcome given that it is essential to understand the Sun-Earth-Moon system in 3-D in order to fully understand lunar phases (and eclipses). There are studies indicating that a stereo display can be much more effective in communicating information for a particularly 3-D intensive tasks [e.g., Ware and Franck 1998; Volbracht et al. 1997]. There is also evidence that the effectiveness of a stereo display is task specific [Hubona et al. 1999], and that 2-D perspective representations may, in certain circumstances, be just as effective as 3-D representations [Cockburn and McKenzie 2002].

When considering whether or not to use a 3-D display in instruction, there will be a trade-off between pedagogical effectiveness and cost and ease of use. A low-cost stereo projection system that has become popular in recent years is the Geowall system (www.geowall.org), which is comprised of two projectors stacked on top of each other fitted with polarized filters, one produces a diagonally polarized image from upper left hand corner to bottom right hand corner, while the other produces a diagonally polarized image perpendicular to the first image (see Figure 3.2). Two images of the same scene, but with a slight separation, are projected onto a special silver matted screen that allows for the light to retain its polarization when it is reflected. Viewers wear glasses where each eye looks through a filter that matches the polarity of each projector in order to get image separation and the stereo effect. This inexpensive stereo visualization system has been extensively used in undergraduate geoscience classrooms [Morin et al. 2003]. Moreover, there is a freely available software package, called AstroWall, for viewing lunar phases and seasons (www.geowall.org/astrowall.html), and some evidence has been presented that use of the

AstroWall had positive effects for student comprehension on phases of the moon [Turner et al. 2003]. What is not known, however, is if the use of a stereo display in the context of teaching lunar phases provides a significant pedagogical advantage that makes the use of a special projection system worth the time and effort.

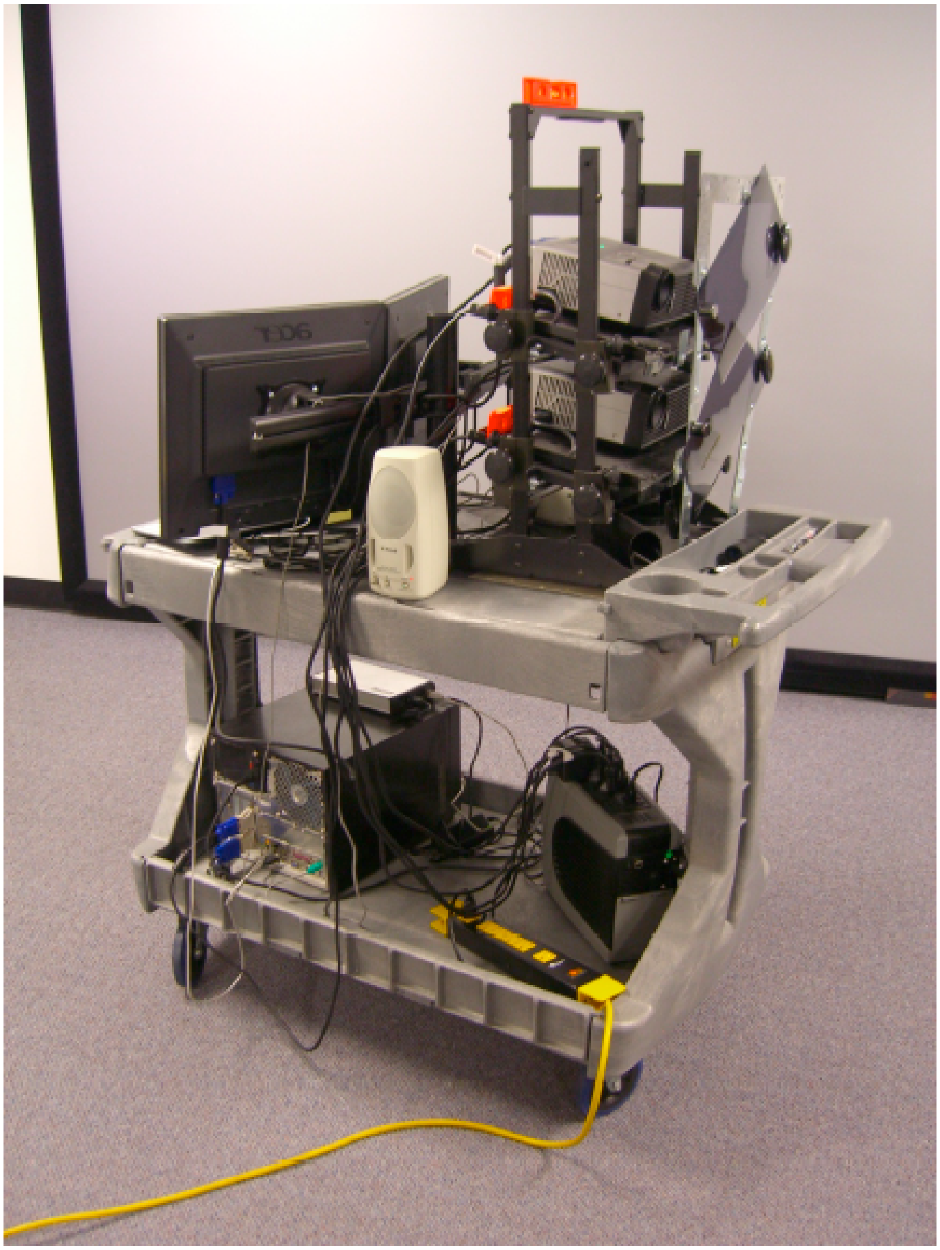


Figure 3.2: GeoWall. The portable GeoWall stereo projection system used in this study (portable silver-matted screen not shown).

3.3 Methodology

The target population for this study included undergraduate students taking an introductory astronomy science course. The accessible population included undergraduate astronomy students enrolled in a public university in North Texas. The sample population for this study was drawn from undergraduate students enrolled at the University of Texas at Arlington (UT Arlington), a large (>28,000 students), comprehensive doctoral/research based university in Arlington, Texas (centrally located between Dallas and Fort Worth).

The method used for selecting the sample was semi-convenient. During the Fall 2008 semester, we created and conducted a Lunar Phases Lab in place of the normal Moon Lab for the Introduction to Astronomy courses that semester. We used all fifteen intact labs for the four Introduction to Astronomy lectures. Students from each of the four astronomy courses are not restricted to a particular lab, therefore they enroll in whichever lab fits their schedule. In this manner the labs retain their homogeneity. The phases of the moon lab was created by the author and Ramon Lopez using the AstroWall software as the principle display tool. The author, Ramon Lopez, and three other lab instructors conducted two approximately 2-hour training sessions on how to teach the lab.

All fifteen of the lab sections used the GeoWall system using the AstroWall software (see Figure 3.3). Eight lab sections had the projection done in full stereo using both projectors. The remaining seven lab sections were conducted with the exact same lab protocol and structure but without stereo. The 2-D labs used the same equipment, but one of the projectors was covered up and the students did not wear the 3-D glasses. Labs were divided based on the following requirements:

1. In order to reduce instructor bias, we had each instructor teach an even number of both stereo and perspective labs. There was only on instructor who had an uneven number of stereo and perspective labs because there were an uneven number of labs to begin with.

2. We spread the stereo and perspective labs out over the entire week to get rid of the day of week bias.
3. We maintained an even number of male to female instructors to get rid of gender bias. For the stereo labs, there were four labs taught by males and four labs taught by females. For the perspective labs, there were three labs taught by males and four labs taught by females.
4. We brought the equipment into the normal lab room in order to remove the location bias.

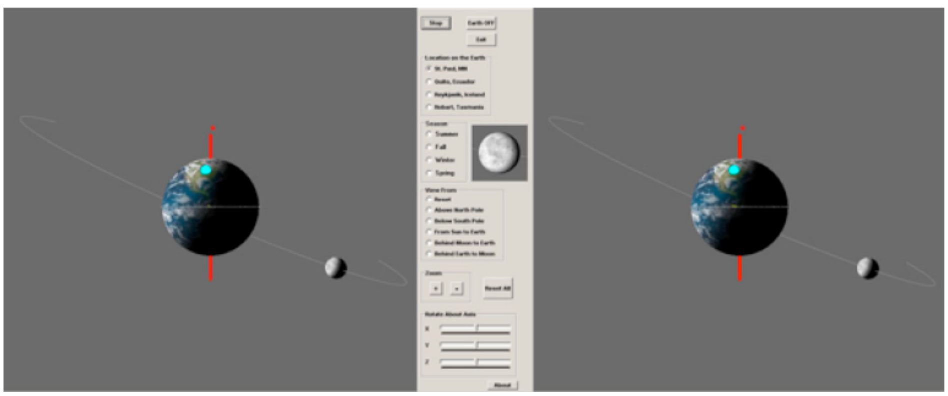


Figure 3.3: AstroWall Software. This is the AstroWall display in the two-projector view as seen from the dusk side in the Full Moon Phase. The white line represents the orbit of the Moon around the Earth. In this view the Sun is off screen to the left. The blue dot represents a person standing on the Earth. The red line is the Earth's rotation axis. The grey box on the right side of the left image controls the demonstration and shows the phases of the moon as seen from an observer on Earth. The two Earth-Moon images, which are slightly offset from each other, are superposed by the GeoWall. Viewers wearing polarized glasses see the image in stereo. In the non-stereo labs, the right image was covered up. *Courtesy of* *http://www.geowall.org/astrowall.html*

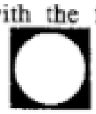
Because the exact same equipment was used in all labs (with the exception of the glasses), all students experienced the same lab procedure with almost the same script (the only difference was an explanation of how passive stereo works for the 3-D stereo labs). To reiterate, the main difference was that for one set of labs the images used were in stereo, while the other labs had only a 2-D perspective view of the system, though that perspective could be from vantages points spread out in 3-D space. We assessed student comprehension of phases of the moon using the Lunar Phase Concept Inventory (LPCI) [Lindell and Olsen, 2002]. A pretest with the LPCI was given in the lecture several weeks before the lab, and a posttest was given at the end of the lab.

3.3.1 Lunar Phases Concept Inventory and Misconceptions in Astronomy

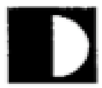
The Lunar Phases Concept Inventory (LPCI) [Lindell and Olsen, 2002] was created in order to assess students' comprehension of Lunar Phases. The LPCI is a multiple choice assessment with 20 questions devoted to orbital time period, phase time period, direction of rising and setting, main phases (New Moon, 1st quarter, Full Moon, 3rd quarter), eclipses, and phases if viewed from different locations on Earth. Many of the questions refer to the spatial relationship between the Sun, Earth and Moon, and included in the choices of answers are distracters (see Figure 3.4 for example of LPCI questions). There are an additional 9 multiple choice questions devoted to collecting student demographics which included the following: confidence for answers chosen, major, age, home community (i.e. urban suburban, rural, etc.), ethnic background, highest math level achieved, confidence in math, and confidence in science.

Lunar Phases Concept Inventory

17. You observe the moon with the following shape and phase tonight,



At some later time, you observe the moon again and it has the following shape and phase.



How long between your two observations?

a. 1 hour	b. 3 hours	c. 6 hours	d. 1/2 day	e. One day
f. 1 week	g. 2 Weeks	h. 3 Weeks	i. 1 month	j. More than 1 month

18. What caused the moon to appear different in your two observations in question 17?
   a. an object is now between the Moon and the Earth
   b. the Moon is now covered by the shadow of the Sun
   c. the Moon is now covered by the shadow of the Earth.
   d. the Moon's position relative to the Earth has changed.
   e. Both a and d.
   f. Both b and d
   g. Both c and d
   h. None of the above

Figure 3.4: Example of LPCI Question. Question 18 uses the different phases of the Moon from question 17. The multiple choices presented include the common misconception that the Earth is casting it's shadow, but it also includes some distracter answers such as the shadow from the Sun and having and object pass between the Moon and the Earth. In everyday experiences when an object passes between two objects it has the ability to cast a shadow on one of the objects so it is a logical conclusion for a novice student.

As discussed above, the explanation for Lunar Phases is a topic in introductory astronomy with which many students have difficulty, and Lindell and Olsen [2002] claim that one of the reasons for this is because students have a prior understanding of Lunar Phases caused by observations, prior instruction, and, in some cases, incorrect conclusions based on correct conceptual ideas. Chapter 3 showed one example of how students can have embedded conceptual understandings that are so strong that even when prompted they refuse to let go of their ideas. It is known that if a student possesses these prior understandings, unless instruction can work with or disprove the foundations that have already been created by the student, it becomes very difficulty for the student to fully comprehend the new material being presented [*How People Learn, 1999].*

Clear examples of common misconceptions are depicted in the documentary *A Private Universe* [Schnepps and Sadler, 1987]. *A Private Universe* begins at a Harvard graduation where graduates and faculty are asked to explain the seasons and phases of the Moon. A common response in the movie (and amongst students still today) is that summer is caused because the Earth is closer to the Sun. This idea comes from an incorrect application of a correct idea: objects that are closer to a heat source are warmer than objects that are farther from the same heat source. So students take this common understanding and apply it incorrectly to the system. The missing information is from a lack of spatial comprehension of the Sun-Earth-Moon system.

Likewise, when asked what causes the phases of the Moon, two very common beliefs, by the general population, are 1.) the Earth's shadow and 2.) clouds. Both of these responses can be derived from an incorrect application of a correct idea. During a sunny day, when clouds cover the Sun, it gets darker outside, so everyday experiences with cloud cover leads people to an incorrect understanding of the phases of the Moon. Furthermore, people have experience with shadows on a daily basis. People understand that a shadow is caused when an object is blocking light. Therefore it seems like a logical conclusion that if normal objects like trees can cast shadows by blocking sunlight, then the different phases of the Moon can be caused because Earth is casting its shadow on the Moon to cause the different phases, and in fact the notion that the Earth's shadow causes lunar phases is widespread. We argue that these types of misconceptions are rooted in the lack of comprehension of spatial relations of the Sun, Earth, and Moon system.

In a different section of *A Private Universe*, interviews were conducted by 9^{th} grade students from a nearby (close to Harvard) junior high school and Heather, one of the brighter students in the course, spends time discussing the phases of the moon. The interviewers conduct an initial interview with Heather and it is clear that there are certain aspects of what causes seasons and what causes lunar phases that she does not understand. After instruction, during a follow up interview, when asked why she was confused about lunar phases, she says following:

> *Here… in the other… in eighth grade class when you did the videotape? Uhm… well we learned the phases of the Moon, but we didn't learn where the Moon was at those times. 'Cause I don't remember learning that. And that makes it sort of hard, 'cause you know what the phases are, but you don't know where the Moon was. I mean the Moon could be over here, could be over here, could be over here, could be… you know practically anywhere on its orbit around…*

At first it seems that she now has a grasp of the concept, but when pressed again, it is realized that she is trying to incorporate her previous ideas of shadows into the new material she has learned. She understands that a lunar eclipse is caused by the Earth's shadow, but she is unsure about the location of the Moon when it is in the Full Moon phase and New Moon phase.

In trying to reason out her explanation, she grabs three spheres, which represent the Sun, the Earth, and the Moon, and moves them around while she is explaining the New Moon and Full Moon phases. It is during this time that she fully grasps the concept. In order to understand the spatial relationship between the three-body system, she is manipulating the locations of the physical spheres instead of manipulating a mental visualization. Utilizing the spheres reduces the cognitive load and allows her to focus her mental capabilities on the conceptual question of the relative locations of the objects in the Sun-Earth-Moon system during Full Moon and New Moon.

By forcing Heather to address her misconceptions of shadows, she was able to correctly internalize the concept. Again this goes back to the idea that new concepts compete with previous embedded ideas and unless you can truly make the students aware of their misconceptions they will continue to hold onto them and incorporate them, unknowingly, into the new conceptual framework from instruction. Heather has also expressed, in her own words, the difficulty that can be faced when either spatial information is lacking or when the spatial information is presented in a manner that is not comprehensible to the student as a basis for creating a mental model of the situation.

3.4 Lunar Phases Lab and Data

The AstroWall-based lab begins at the full Moon phase looking down on the Sun-Earth-Moon system from above the North Pole. The instructor described the different visuals present in the display and oriented the students to the East and West directions in order to have the students understand that the Moon rises in the East and sets in the West (same as the Sun). The AstroWall display was stepped through time until the Moon reached the 3rd quarter phase. At the 3rd quarter phase, the time evolution was stopped. At that point the instructor rotated the system to provide the students with different perspectives (from the dawn side, from the dusk side, view from the Sun, view from behind the Earth). After the perspective views were finished, the demo was returned to the view looking down on the Earth from above the North Pole. The time evolution of the visualization was started again and advanced until it reached the New Moon phase. It took one minute for the AstroWall software to transition from phase to phase (starting with full Moon, 3rd quarter, new Moon, 1st quarter, and ending with full Moon) and 3-5 minutes to demonstrate the different perspective views. Starting with the 3rd quarter phase and during the different perspective views, the students were asked a set of questions, to discuss amongst themselves, regarding the position of the Moon, rise and set times, percent of illuminated Moon seen by an observer on the Earth, etc. The same procedure was followed for the New Moon phase, 1st quarter phase, and then finished with the Full Moon phase. Directly following the AstroWall lab, the students were given the LPCI post assessment.

Our operating hypothesis was that the 3-D stereo image would produce a significantly larger gain in student comprehension, as measured by the LPCI, compared to the use of an image that shows only 2-D perspective. Similarly our null hypothesis was that there was no significant gain for either 3-D stereo lab or 2-D perspective lab. Other studies have examined the effect of a single lab or the effect of a single intervention on student understanding of a concept [e.g., Abbott et al., 2000], and actually one lab can make a measureable difference in student understanding of an idea.

The demographics of our sample are presented in the following tables. Due to the small sizes of each group, we did not do statistics on them by group; however, we wanted to share an overview of our sample.

Table 3.1: Gender and Major

	Males	**Females**	**STEM Major**	**Non-STEM Major**
2-D Perspective	34	41	2	73
3-D Stereo	51	44	4	91

Table 3.2: Age and Classification

	18	**19**	**20**	**21-25**	**>25**	**FRSH**	**SOPH**	**JR**	**SR**
2-D Perspective	24	25	5	12	4	40	25	4	6
3-D Stereo	21	32	17	22	3	30	41	13	11

Table 3.3: Previous Astronomy or Physics Courses

	Astronomy	**Physics**
2-D Perspective	14	18
3-D Stereo	19	14

Table 3.4: Means and Standard Deviation for Pre and Post LPCI Data

	N	**Pre-LPCI**	**Post-LPCI**
2-D Perspective	75	*M* = 10.48 *SD* = 4.07	*M* = 11.88 *SD* = 4.36
3-D Stereo	95	*M* = 9.45 *SD* = 3.68	*M* =10.36 *SD* = 4.12

In order to analyze our data, we used a two-tailed, related measures, t-test. Even though table 3.4 shows means and standard deviations for pre and post LPCI in order to do the statistics we use the difference scores (i.e. post scores – pre scores). For student comprehension of lunar phases in a 3-D stereo lab, as predicted, there was a statistically significant gain as measured by the LPCI (M_D = 0.905, SD = 3.346), t_D (95) = 2.637, p = 0.00978, r^2 = 0.069. We also measured a statistically significant gain for student comprehension on lunar phases in a 2-D perspective lab, as measured by the LPCI, (M_D = 1.4, SD = 3.665), t_D (75) = 3.308, p = 0.00978, r^2 = 0.129. Based on the r^2 value, the 3-D stereo lab did not have a larger effect on student comprehension of lunar phases than the 2-D perspective lab; therefore we could not reject that part of the null hypothesis. Thus, while the stereo AstroWall labs had a positive effect on student learning in a single lab experience, the 2-D labs also had a positive effect, but the two are not statistically different. In this case, we found that stereo visualization did not add to the educational experience.

3.5 Discussion

Physics and astronomy are subjects in which highly spatial information is crucial to understanding concepts. For students learning physics and astronomy, it has been shown that student comprehension can be improved by reducing the cognitive load produced by mental processing of spatial relationships by representing the system with 3-D views [e.g. Hansen et al. 2004; Barnett et al. 2005; Küçüközer 2008; Küçüközer et al., 2009]. In this manner students are not forced to try to imagine the system presented to them, manipulate the system in their minds, and then try to understand concepts based (in part) on spatial relationships. However, a perspective view provides important spatial cues, and, depending on the context, 2-D perspective information may be sufficient for students to deal with the material as suggested by Cockburn and McKenzie [2002]. Learning about lunar phases seems to be such a case. In this study both the 3-D stereo and 2-D perspective lunar phase labs had a statistically significant gain in student comprehension as measured by the LPCI, but the gains were not statistically different. We conclude that having a well-constructed spatial lab with a system that has 2-D perspective visual cues would be enough to have a positive effect on student comprehension of lunar phases. Therefore we argue that while 3-D stereo displays are effective in improving student comprehension, they are not necessary to create the same amount of gain in student comprehension. The AstroWall software allowed us to create an easily comparable system between the two labs, but the stereo capabilitities are not essential to produce a positive result. We suspect that other topics in introductory astronomy with a similar level of spatial complexity (e.g., seasons) also are likely to draw no significant benefit from the use of a stereo display in instruction as compared to perspective displays that are easier to create and use.

CHAPTER 4

CURRENT WORK

4.1 Introduction

We have a strong interest in determining how the cognitive load resulting from the mental manipulation of spatial information impacts success and comprehension in Physics. From Chapter 1, we know that there are other STEM fields that have made claims that high visual-spatial skills are necessary for success in STEM fields, but there has not been a lot of research on how spatial skills relate to student success in physics courses. There are specific studies that show there are correlations to visual-spatial skills related to topics like kinematics [Kozhevnikov et al., 2007] but not much more than that.

Siemakowski and MacKnight [1971] conducted a study to test whether science students had higher spatial abilities than non-spatial students. They found that science students had statistically significantly higher scores on spatial tests than non-science students and that physics students had the highest spatial ability of all science students. Their population was taken from the undergraduate population at SUNY-Buffalo. They were interested in student spatial abilities from the perspective of academic advising. For example, incoming art students, in some universities, are required to submit an art portfolio as part of their entrance exams, music students are tested on tone distinctions and music reading capabilities, but there are no such assessments (aside from general math and reading assessments that all students must take) that explore the strengths of science students.

Their first study was to exam the differences between non-science students (liberal arts) and science (introductory level) students and, as stated above, science students scored higher on their spatial assessment (Survey of Object Visualization). Once it was established that science students had statistically significantly higher scores than non-science students, they wanted to compare the different sciences in terms of spatial abilities.

Their first study was a comparison of prominent faculty members in Physics, Biology, Psychology, and Anthropology. From highest to lowest scores, the subjects are as followed: Theoretical Physicists, Experimental Physicists, Psychologists, Biologists, and then Anthropologists. When they looked at student populations, the order goes as follows (from highest scores to lowest scores): Physics, Geoscience, Biology, and Chemistry. Siemakowski and MacKnight [1971] recognize that chemistry students who are studying molecular diagrams would probably score higher but did not look at those students. From this study, both experts (faculty members) and novices (students) in Physics had the highest scores for spatial abilities.

Pallrand and Seeber [1984] examined 3 different visual-spatial abilities of first semester introductory calculus-based physics students from a community college (in a large urban state). They had 3 groups of students who they were examining. One group of students (the experimental group) were given a spatial intervention along with regular instruction. The second group of students (the placebo group) were given lectures on the history of physics by the same individuals who administered the spatial intervention. The third group of students (the control group) were given no extra instruction. They also included liberal arts students for a test-retest control group. For a measure of success in the course, all students were given the same departmental exam regardless of the class they were in.

Students from the experimental group, who were given the spatial intervention, scored higher on the final exam than those students who did not have the spatial intervention. An interesting finding was looking at students who dropped the course. They compared initial spatial ability scores from those students who dropped out of the course to those students who remained in the course and they found that those students who dropped out of the course had statistically significantly lower scores than those who finished the course even though they had similar mathematics scores. These results corroborate results that Sorby [1999] found for the increased retention rates of students with low spatial skills enrolling in the visual spatial course at Michigan Tech.

Because of our interest in cognitive load, particularly that produced by mental manipulation of spatial visual objects, we have chosen to revisit the Siemakowski and MacKnight [1971] and Pallrand and Seeber [1984] studies. We are interested in how physics students compare to other STEM students and how rotational spatial abilities are related to student success in physics.

4.2 Spatial Ability Comparison of STEM Majors

At UT Arlington there is a new program called Arlington Undergraduate Research-based Achievement in STEM (AURAS) designed to reduce attrition rates in introductory courses. AURAS is funded by an NSF STEP grant and is a collaborative project between the College of Engineering, Department of Physics, Department of Chemistry, and Department of Mathematics. In preparation for the submission of the proposal, the AURAS team conducted a self-study to determine at what point UT Arlington STEM majors were suffering the greatest attrition. Most of the STEM losses occurred in the first 2 years, which is not surprising [e.g., Carey, 2005], and were highly correlated with poor student performance in either their first mathematics class or their first chemistry class. AURAS created an Emerging Scholars Program (ESP) for Pre-Calculus, Calculus, and Chemistry to address this issue. The ESP model itself is based on the work of Uri Treisman at UC Berkeley [Treisman, 1992], which the author of this dissertation experienced first-hand as an undergraduate at UC Berkeley. It is known that all STEM undergraduate students have to take calculus and most of them also take chemistry, so it is hoped that improving retention in these courses, through ESP, trickles down to the other STEM departments. The work presented in this section, supported by AURAS, focuses on how spatial ability in Engineering, Mathematics, Physics, and Chemistry are correlated with success in the introductory courses.

In order to determine if spatial ability has an effect on success in STEM courses, we need to first identify the differences, if any, between the different fields therefore we are revisiting Siemakowski and MacKnight's [1971] study.

4.2.1 Methodology

The target population for this study included undergraduate students taking an introductory STEM course. The accessible population included undergraduate enrolled in a public university in North Texas. The sample population for this study was drawn from undergraduate students enrolled at the University of Texas at Arlington (UT Arlington), a large (>29,000 students), comprehensive doctoral/research based university in Arlington, Texas (centrally located between Dallas and Fort Worth). The target populations, accessible populations, and sample populations will be the same for all data presented in this chapter and therefore will not be repeated for each experiment presented in this chapter. All data presented in this chapter was also collected ethically via the rules put forth by the Institutional Review Board (IRB) present at UT Arlington and only data from students who signed the informed consents is presented here in conjunction with regulations at UT Arlington.

The method used for selecting the sample was semi-convenient. During the 2008, 2009, and 2010 school years, we collected data from Calculus 1, Introductory 1105 Engineering courses, calculus-based Introductory Mechanics and Electricity and Magnetism Physics courses, and Introductory Chemistry. The fields present in the introductory engineering courses were Mechanical and Aerospace Engineering (MAE), Computers Science and Engineering (CSE), Electrical Engineering (EE), Civil Engineering (CE), and Industrial and Manufacturing Systems Engineering (IE). This represents all eight departments that offer baccalaureate degrees because some of the departments combing their students in these courses.

In each course, we gave the MRT in the beginning of the semester as a Pre-test and then we gave the MRT again at the end of the semester as a Post-test. Calculus post-tests, however, had some different circumstances (that will not be discussed here), therefore calculus is not included in the comparison with the other STEM courses. There is, however, a section devoted to some of the data that we collected with Calculus. With the engineering courses there were no statistical differences between pre-tests so we were able to combine them into one data set. There was also no statistical difference between the post-tests so again we combined the scores into one data set. We did the same analysis for the physics courses and we were able to combine the data into one data set as we did with the engineering data set. The two Chemistry courses, however, had statistically different scores so we could not combine them into one set therefore they will be labeled Chemistry 1 and Chemistry 2 as to avoid any identifying markers. It must also be noted that the students in the Chemistry 2 group were subjected to a stereotype threat.

A stereotype threat is when a subject "is in a situation where [he/she] faces judgment based on societal stereotypes about one's group" [Spencer et al., 1999]. For example, a common stereotype is that women cannot perform as well as men in math. Spencer and colleagues [1999] showed that when giving a math exam, women exposed to this stereotype performed lower than women who were not exposed to this stereotype. Therefore simply exposing a group to a negative stereotype can significantly impact scores of that group.

In the Chemistry 2 course, while gaining the students attention, the instructor introduced the assessment instructors and announced that it would be an assessment of spatial ability. The instructor went on to inform the class that males tend to perform better than females so "the girls should try their best". Therefore this subject pool is tainted and the results cannot be considered an accurate representation of the population. This could also be one of the reasons for why the scores showed a statistical difference.

4.2.2 Data

Our operating hypothesis was that there is a difference between gains between the different STEM fields as measured by the Mental Rotation Test. Similarly, our null hypothesis was that there is no difference in the gains between the different STEM fields as measured by the Mental Rotation Test. Table 4.1 shows the sample size for each course for both pre and post assessments. There was some mortality (loss of subjects) due to students who dropped the course, students who no longer wanted to participate, and student absences the day we administered the post-tests. These are the reasons for the different numbers between pre and post assessments. The means and standard deviations, for each STEM course, for pre and post MRT scores are listed in Table 4.2.

One thing to note about Table 4.1 is that the sample size for both pre and post data for chemistry are small. In fact they do not meet the minimum sample size of 30, for a random sample, [Gravetter and Wallnau, 2007] in order to generalize results to larger populations. However, we present the data in order to give a complete report of the data collected.

Table 4.1: Sample Size for Pre-Post MRT STEM Comparison.

N	Pre-Test	Post-Test	Total
Engineering	104	65	169
Physics	50	39	89
Chemistry 1	24	18	42
Chemistry 2	22	12	34
Total	200	134	334

Table 4.2: Means and Standard Deviation for Pre Post MRT

Mean/Standard Deviation	Pre	Post
Engineering	*M* = 9.66 *SD* = 5.10	*M* = 11.34 *SD* = 5.21
Physics	*M* = 10.20 *SD* = 4.54	*M* = 13.54 *SD* = 4.27
Chemistry 1	*M* = 8.92 *SD* = 5.05	*M* = 11.5 *SD* = 5.28
Chemistry 2	*M* = 12.23 *SD* = 4.72	*M* = 15.08 *SD* = 4.38
Total	*M* = 9.99 *SD* = 4.96	*M* = 12.36 *SD* = 4.93

4.2.3 Results

We were interested in reevaluating Siemakowski and MacKnight's [1971] study. Their study used a different spatial ability assessment and, in addition technology, especially video games, which have an impact on spatial ability [Spence and Feng, 2010], have advanced significantly since the 1970's. In order to assess if there exists a difference between different STEM courses, in terms of spatial ability, we conducted a two-way Analysis of Variance (ANOVA). The two-way ANOVA found that the course had a significant effect ($F(3, 326) = 4.825$, $p = 0.003, \eta^2 = 0.043$); this means that there was a statistical difference between the different STEM courses as measured by the MRT. The two-way ANOVA found that the time (pre or post) had a significant effect ($F(1, 326) = 15.492$, $p < 0.001$, $,\eta^2 = 0.045$); this means that there was a statistical difference between the total pre and post MRT test. The two-way ANOVA did not find a significant interaction (course*time) effect ($F(3, 326) = 0.591$, $p = 0.621$, $,\eta^2 = 0.005$). Figure 4.1 shows a plot for a more visual description of this data.

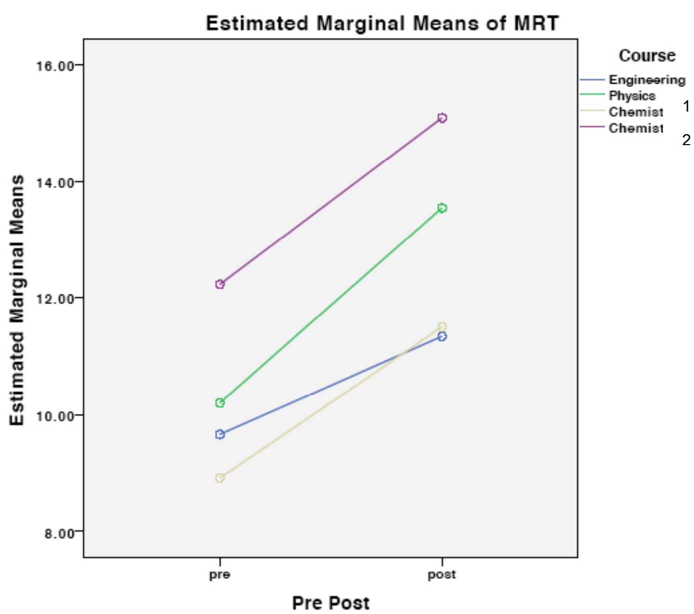


Figure 4.1: Graphical Representation of the Pre-Post MRT Data. This shows the means relative to each other. Please remember that the chemistry sample sizes are too small to accurately representation the populations and they subjects have also been subjected to stereotype threats.

If we look at just Engineering and Physics, we can see that there is a difference for post-test as measured on the MRT. To show that there was no statistical difference in pre-MRT scores, we conducted an independent two-tailed t-test; $t\,(152) = -0.633$, $p = 0.156$, *n.s*. The sample size, means, and standard deviations are presented in Table 4.1 and Table 4.2. Figure 4.2 shows just the engineering and physics data.

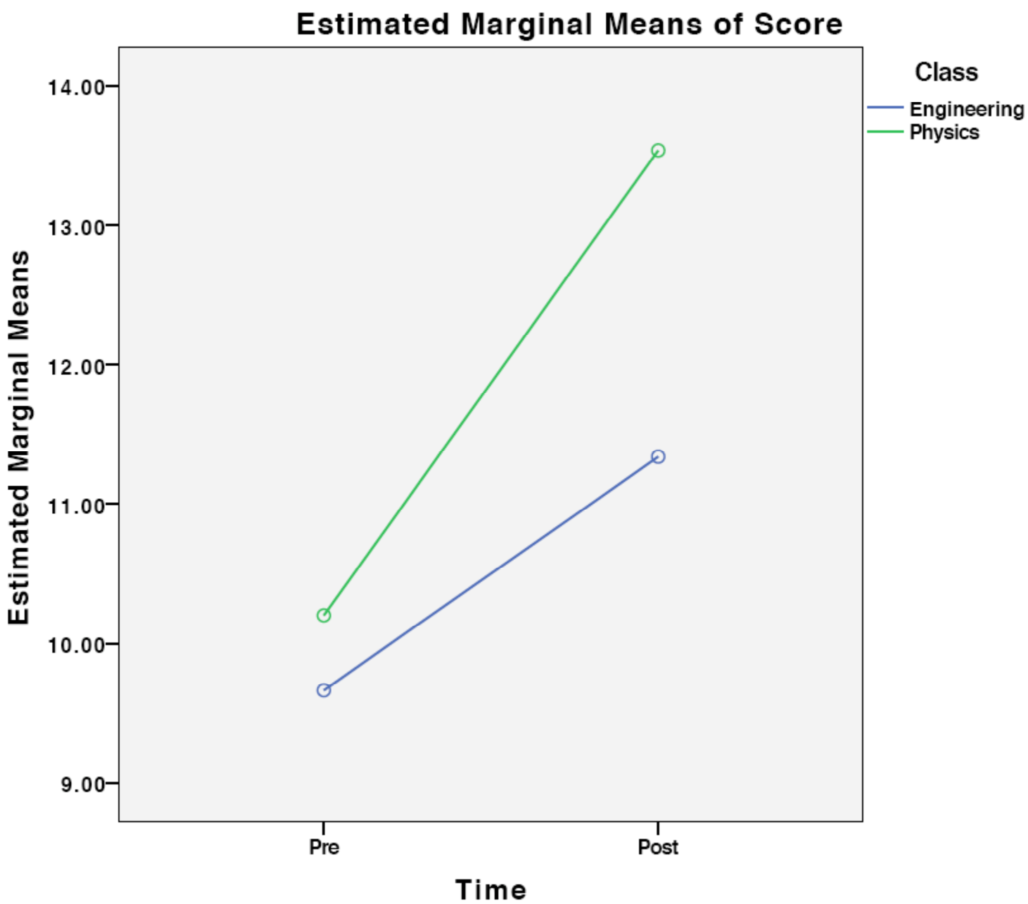


Figure 4.2: Graphical Representation of the Pre-Post MRT Data for Engineering and Physics. This shows the means relative to each other without chemistry's influence. It is clear that physics has a stronger influence on spatial ability, as measured by the Mental Rotation Test, than engineering.

When looking at a two-way ANOVA for just Physics and Engineering, we found that the course had a significant effect ($F(1, 258) = 4.400$, $p = 0.037, \eta^2 = 0.017$); this means that there was a statistical difference between the Physics and Engineering courses as measured by the MRT.

The two-way ANOVA found that the time (pre or post) had a significant effect ($F(1, 258) = 14.769$, $p < 0.001$, ,$\eta^2 = 0.055$); this means that there was a statistical difference between the total pre and post MRT test. The two-way ANOVA did not find a significant interaction (course*time) effect ($F(1, 258) = 1.626$, $p = 0.203$, ,$\eta^2 = 0.006$).

4.2.4 Discussion

From the data presented in Sections 4.2.2 and 4.2.3, we can see that the specific STEM courses influence students' spatial ability. Unfortunately chemistry had some uncontrollable circumstances that prevented us from having a large enough sample size to do an accurate comparison. We were also unable to obtain pre and post MRT data in mathematics. Therefore we were not able to reproduce Siemakowski and MacKnight's [1971] study in its entirety. We did however confirm Pallrand and Seeber's [1981] study that suggested that the simple act of taking a physics course has an effect on students' spatial ability.

Because Engineering and Physics had no statistical difference in their pre scores on the MRT, the fact that they had a statistical difference on their post scores confirms that simply taking an introductory physics courses has a statistical effect on students' spatial ability and it is a stronger effect than taking an introductory engineering course.

4.3 Test ReTest Effect of MRT

In Pallrand and Seeber's [1981] study, in order to test if there was a test retest effect, they used liberal arts students. There is a fundamental difference between liberal arts students and STEM students and it has already been shown that on average STEM students have higher spatial ability than non-STEM students. Therefore we argue that a test retest effect should be measured with the same population of students.

4.3.1 Methodology

At the end of the 2009 semester we administered the MRT in introductory Physics and Engineering courses. We invited the students who signed the informed consent to return 14 weeks later in the summer to come back for a follow up assessment. We only chose students who were not taking a STEM course over the summer. The justification is that if students are not enrolled in STEM course during the summer they would not be influenced, spatially by those courses. These samples were drawn from equivalent courses as the data from Section 4.2 was taken from. Therefore the variability should be the same and because we already had post test data from these courses, it would be an equivalent comparison. Unfortunately only 11 students returned to participate.

At the end of the spring 2010 semester we repeated the MRT assessment in introductory engineering and physics courses. We again invited students to return 14 weeks later in the summer for a follow up study. During this set of data collection, 28 students returned for the follow up. Therefore the total sample size for test-retest assessment is N = 39. The data for the test retest assessment is presented in the next section.

4.3.2 Data and Results

We were able to combine the 2009 and 2010 data because there was no statistical difference between their Pre scores on the MRT as measured by a two-tailed independent t-test; t (37) = -.242, p = .25, *n.s.* As a point of clarification, pre-tests in this data set were taken at the end of the semester whereas pre-tests in Section 4.2 were collected at the beginning of the semester. Table 4.3 shows the means and standard deviations for Pre-Post MRT data and Figure 4.3 shows a graphical representation of their means.

Table 4.3: Means and Standard Deviation for Engineering, Physics, and Test ReTest

Mean/Standard Deviation	**Pre**	**Post**
Engineering	*M* = 9.66 *SD* = 5.10	*M* = 11.34 *SD* = 5.21
Physics	*M* = 10.20 *SD* = 4.54	*M* = 13.54 *SD* = 4.27
Test ReTest	*M* = 12.05 *SD* = 5.18	*M* = 15.08 *SD* = 4.26

Though we were able to meet the minimum number of 30 for the Test Retest data, it did not consist of a random sample, therefore the sample size was too small to make an accurate comparison. The students who returned in the summer were above average for reasons we can only speculate about. Some of them had expressed that they had taken the MRT several times in the past, outside of the course where we solicited their participation. Some expressed their excitement about assessments like the MRT and had been practicing throughout the summer.

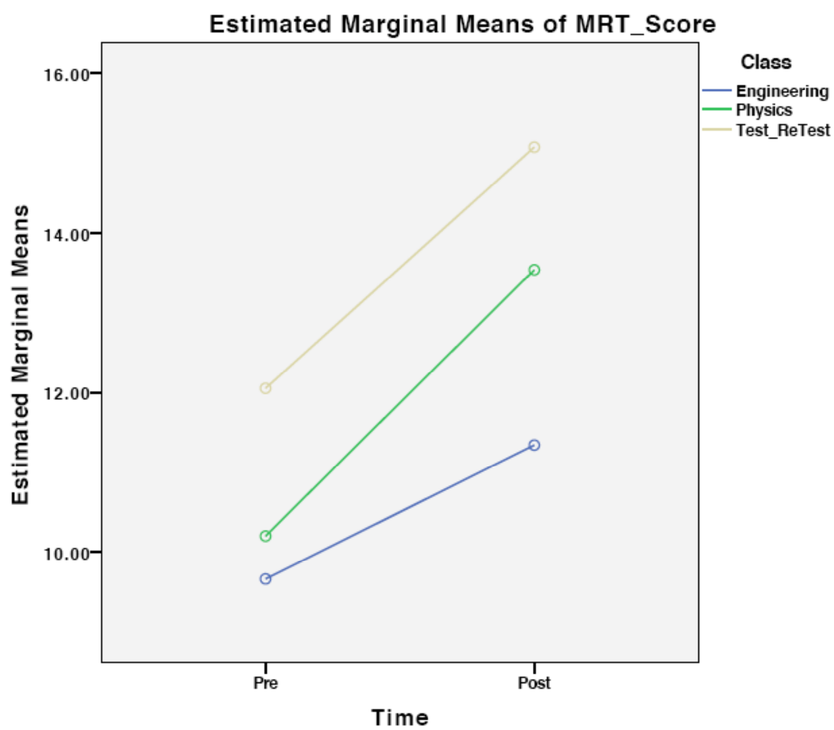


Figure 4. 3: Graphical Representation of Mean Differences for Engineering, Physics, and Test ReTest. It is clear that the Test Retest data is statistically higher than both Engineering and Physics.

4.3.3 Discussion

Even though we were not able to get an accurate representation from the test retest data, we can still make claims about their effects. We can look at the engineering and physics data and see that there is a difference. If both sets of data had no statistical differences to begin with and physics ended with statistically higher scores, then we know that these results are not due to a test retest effect alone.

4.4 Correlation Between Spatial Ability and Mathematics Comprehension

One of the research questions of AURAS is to exam the correlation between spatial ability and success in introductory mathematics courses. The self-study done by the AURAS leadership team determined that the first math course and the first chemistry course taken by STEM majors, at UT Arlington, were significant predictors of future retention. Students who did poorly in one or both of those courses were much more likely to drop out of the STEM major track. Factors that can influence performance in that first math course or chemistry course are thus very important and directly related to the AURAS goal of improving STEM student retention and graduation rates. Therefore as part of the research component of AURAS, we are investigating the possible connection between spatial skills and performance in these gateway courses and here we report the findings for the correlation between spatial ability and performance in Calculus 1 courses.

4.4.1 Methodology

During the Fall 2010 semester, we were allowed to administer the MRT in the beginning of the semester and we were allowed to collect data from two sections of Calculus 1. These sections were designated to be AURAS Emerging Scholars Program (ESP) sections. Calculus 1 students enrolled in ESP sections are required to attend normal lecture and lab sections, but are additionally required to attend 2 two-hour ESP workshops each week were problem based instruction is encouraged. We were allowed to collect post-MRT data but we will not be presenting that data here. At the end of the semester, we were also provided with final grades, which we used as a measure of success, in order to test if there was a correlation between pre-MRT scores and student success.

4.4.2 Data

For this data, our null hypothesis (H_0) is that there is no correlation between student pre-MRT scores and success in Calculus 1. Our alternate hypothesis is that there is a correlation between pre-MRT and student success in Calculus 1. We were able to collect data from two different sections for Calculus 1. In order to combine the data into one set we conducted a two-tailed independent *t-test*. The means and standard deviations for each sample are presented in Table 4.4.

Table 4. 4: Mean and Standard Deviation for Pre-MRT and Final Grades for Calculus 1.

	N	**Mean**	**Standard Deviation**
Pre MRT Calculus 1_1	49	10.59	5.40
Pre MRT Calculus 1_2	42	8.62	4.83
Final Grades$_{Calc1}$	48	1.94	1.44
Final Grades$_{Calc2}$	42	1.83	1.34

The t-test revealed that there is no statistical difference ($t(89)$ = 1.804, p = 0.071, n.s.) between these two populations thereby allowing us to combine the data into one sample. There was also no statistical difference between the final grades of these two populations ($t(88)$ = .268, p = 0.606, n.s.) so, again, we could combine all of the data into one sample.

In order to test if the MRT is a predictor of success in Calculus 1 we did a correlation study. Table 4.5 shows the Means and Standard deviation for the data as one sample size. Figure 4.4 shows the correlation table between for Pre MRT scores and Final Grades in Calculus 1.

Table 4.5: Mean and Standard Deviation for Total Calculus Sample

	N	**Mean**	**Standard Deviation**
Pre MRT	91	9.68	5.21
Final Grades*	90	1.89	1.39

**There is one less number because one student was not included in the final grade roster*

Correlations

		Pre_MRT	Final_Grade_Point
Pre_MRT	Pearson Correlation	1	.338**
	Sig. (2-tailed)		.001
	N	91	90
Final_Grade_Point	Pearson Correlation	.338**	1
	Sig. (2-tailed)	.001	
	N	90	90

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 4.4: Correlation Table for Pre MRT and Final Grades in Calculus 1

4.4.3 Results

Looking at the Figure 4.4 we can see that the Pre MRT scores are positively correlated with Final Grades in Calculus 1, Pearson's $r(90) = .338$, $p = 0.001$. This means that a student's incoming spatial ability, as measured by the MRT, has a medium effect size for predicting their Final Grade in Calculus 1. This has implications for retention of STEM students. If the preliminary results, indicating a correlation between spatial skills and Calculus performance, represents a causal relationship, that would suggest that students who score low on the MRT could improve their performance in Calculus by improving their spatial skills.

4.5 Correlation Between Spatial Ability and Engineering

We conducted a similar experiment as Section 4.4 but with Engineering students. The strongest influence for this research is Sheryl Sorby's work [e.g. Sorby and Baartmans, 1996; Sorby, 1999; Sorby, 2005; Sorby, 2005] at Michigan Tech. Her work suggests that improving students' spatial skills will improve retention of students, particularly women, in Engineering course and other STEM fields (including Physics), and can act as a predictor for success in these types of courses. We wanted to reproduce these studies here at UT Arlington with our Engineering students.

4.5.1 Methodology

Our sample population was taken from same engineering population described in Section 4.2.1 i.e. Mechanical and Aerospace Engineering (MAE), Computers Science and Engineering (CSE), Electrical Engineering (EE), Civil Engineering (CE), and Industrial and Manufacturing Systems Engineering (IE). Again, this represents five of the eight departments that offer baccalaureate degrees. Our sample size was a little bit larger because we were able to add another CSE course to the data. This data was collected during the 2010 spring and fall semesters.

We gave students in the different introductory engineering courses the MRT (pre data) at the beginning of the semester. At the end of the semester we were only able to collect post data from a subset of the courses. Therefore we will present the data for the two populations individually; one with all of the pre data and one with the subset of data as comparisons.

4.5.2 Data for Total Engineering

We did a correlation study with Pre MRT data for all engineering data. We then collected final grades for each student who signed the informed consent and participated in the study. We were not able to get a break down of each students' grades in terms of percentages, but were given grades as final letter grade. We converted the letter grades with the following point values: A = 4, B = 3, C = 2, D = 1, F = 0 (incompletes and withdrawals were also counted as 0) . The sample means and standard deviations are presented in Table 4.6 and correlations are shown in Figure 4.5.

Table 4. 6: Means and Standard Deviation for Total Engineering Data

	N	**Mean**	**Standard Deviation**
Pre MRT Scores	141	9.55	4.96
Final Grades	141	2.68	1.35

Correlations

		Final_Grade	Pre_Mrt
Final_Grade	Pearson Correlation	1	.040
	Sig. (2-tailed)		.635
	N	141	141
Pre_Mrt	Pearson Correlation	.040	1
	Sig. (2-tailed)	.635	
	N	141	141

Figure 4.5: Correlations for Total Engineering Pre MRT and Final Grades

From the data presented in Figure 4.5 we can see there is no correlation between Pre-MRT scores and Final Grades in Engineering (Pearson's $r(141) = 0.040$, $p = 0.635$, *n.s.*).

4.5.3 Data for Subset of Engineering

Because we were not able to collect post data for several engineering courses we were not able to do a comparison with G scores or difference scores on the entire set of data.

Therefore, we used a subset of engineering, which only includes CSE, MAE, and IE. The means and standard deviations are presented in Table 4.7.

Table 4.7: Means and Standard Deviation for Subset of Engineering Data

	N	Mean	Standard Deviation
Pre MRT Scores	73	9.73	1.30
Post MRT Scores	62	12.94	5.18
Difference Scores	62	2.50	2.89
G Scores	62	0.12	0.14
Final Grades	73	3.01	1.30

The first thing to take note of is the mean for the final grades in table 4.7 is higher than the mean for total sample size in table 4.6. The correlations are presented in Figure 4.6.

Correlations

		Final_Grade	Pre_Mrt	Post_MRT	Diff_MRT	G_MRT
Final_Grade	Pearson Correlation	1	.276*	.159	.030	.020
	Sig. (2-tailed)		.018	.217	.816	.875
	N	73	73	62	62	62
Pre_Mrt	Pearson Correlation	.276*	1	.846**	-.293*	-.142
	Sig. (2-tailed)	.018		.000	.021	.270
	N	73	73	62	62	62
Post_MRT	Pearson Correlation	.159	.846**	1	.262*	.394**
	Sig. (2-tailed)	.217	.000		.040	.002
	N	62	62	62	62	62
Diff_MRT	Pearson Correlation	.030	-.293*	.262*	1	.963**
	Sig. (2-tailed)	.816	.021	.040		.000
	N	62	62	62	62	62
G_MRT	Pearson Correlation	.020	-.142	.394**	.963**	1
	Sig. (2-tailed)	.875	.270	.002	.000	
	N	62	62	62	62	62

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 4.6: Correlations Between MRT and Subset Engineering Final Grades

If we ignore the correlations for any aspect of MRT with itself and look at the correlations between the different scores of MRT and Final Grades, we can see that the only correlation present is the Pre-MRT scores.

4.5.4 Results

There is no correlation from the total engineering data and scores on the MRT, but there is a correlation for the subset engineering Pre MRT data and Final Grades. This means that for the subset data, the Pre MRT can act as a predictor for success in some introductory engineering course. These results agree with Sorby's work but it is not a complete representation of the engineering population, therefore we are hesitant to suggest that these results are generalizable to a larger population.

4.6 Correlations Between Spatial Ability and Physics Comprehension

We are really interested in the idea of a student's spatial ability as a predictor for success in physics courses. Previous research suggests that spatial ability has an impact on retention rates in other STEM fields and even acts as a predictor for success in STEM fields. We wanted to test whether these previous results remained when applied strictly to physics.

4.6.1 Methodology

During the fall 2008 semester were able to collect pre and post MRT data from a Physics 1443 class. Physics 1443 is calculus-based mechanics. Pre MRT data was collected at the beginning of the semester and post-MRT data was collected at the end of the semester. We were also able to collect pre and post data using the Force Concept Inventory [Hestens and Wells, 1992] (FCI). The FCI is a 30-question concept inventory intended for student evaluation in Kinematics and Mechanics. Concepts are not always accurately assessed with course exams so concept inventories are often used to assess conceptual comprehension of a topic. In this set of data we only conducted statistical analysis on matching pre and post data. By using just matching pre and post data, we can reduce the amount of variability in the sample.

The FCI has a time limit of 30 minutes to answer the 30 questions in the assessment. Instead of purchasing sets of scantrons or having the students purchase their own scantrons, we created our own (see Appendix B). Answers left blank, at the end of the 30-minute time period, were considered wrong. Because of the time commitment for both the Informed Consent, the MRT and the FCI, we were only able to collect data from one course. One of our biggest constraints in collecting data presented in this chapter was getting professors to agree to give us time in their courses. It was easier to convince professors at the beginning of the semester than at the end of the semester because often, professors are trying to catch up and finish material in order to prepare students for the next class in the sequence.

Additionally we were able to collect data from 4 Physics 1444 class from the spring 2009 and 2010 semesters. Physics 1444 is calculus-based introductory Electricity and Magnetism course. During random times during the semester we were able to go into each Physics 1444 course and administer the MRT. We only collect one set of MRT data and at the end of the semester, final grades were given to us. All but one course gave us final grades in terms of percentages. The fourth course gave us grades as final letter grades. We converted each grade into the following point values: A = 4, B = 3, C = 2, D = 1, F = 0 (incompletes and withdrawals were counted as 0). Because we have two sets of data, they will be presented separately.

4.6.2 Data for Physics 1443

We collected data from 50 students from Physics 1443 and assessed the data in several different ways. We used seven different variables to do a correlation study. The variables are as follows: Pre MRT, Post MRT, Difference Scores for MRT, G scores for MRT, Pre FCI, Post FCI, Difference Scores for FCI, G scores for FCI, and Final Grades in the course. Difference score are looking at the difference between Post and Pre total scores for each assessment (Post Score – Pre Score). We assume students will improve in their Post scores from their Pre scores, but it doesn't matter which score you take first as long as you remain consistent. The G score looks at the ratio of how much a student gained to the total possible gain. The equation used to calculate a G score is as follows:

$$G = \frac{Post - \Pr e}{Total - \Pr e}$$

The correlation matrix for this study is presented in Figure 4.7. The means and standard deviation are presented in Table 4.8.

Table 4.8: Means and Standard Deviation for MRT, FCI, and Final Grades for Physics 1443

	Mean	**Standard Deviation**
Pre MRT	12.94	4.32
Post MRT	15.38	3.96
Difference in MRT	2.44	2.39
G Score for MRT	0.37	0.37
Pre FCI	12.38	6.03
Post FCI	16.70	5.94
Difference in FCI	4.32	3.61
G Score for FCI	0.24	0.28
Final Grade	82.28	11.13

Correlations

		Pre_MRT	Post_MRT	Diff_MRT	Final_Grade	Pre_FCI	Post_FCI	Diff_FCI	G_FCI	G_MRT
Pre_MRT	Pearson Correlation	1	.836**	-.420**	.132	.060	.103	.068	-.033	.130
	Sig. (2-tailed)		.000	.002	.362	.677	.477	.639	.820	.369
	N	50	50	50	50	50	50	50	50	50
Post_MRT	Pearson Correlation	.836**	1	.146	.105	.148	.145	-.009	-.056	.546**
	Sig. (2-tailed)	.000		.312	.467	.306	.316	.952	.700	.000
	N	50	50	50	50	50	50	50	50	50
Diff_MRT	Pearson Correlation	-.420**	.146	1	-.063	.135	.054	-.137	-.033	.670**
	Sig. (2-tailed)	.002	.312		.662	.349	.709	.342	.820	.000
	N	50	50	50	50	50	50	50	50	50
Final_Grade	Pearson Correlation	.132	.105	-.063	1	.262	.444**	.293*	.377**	.050
	Sig. (2-tailed)	.362	.467	.662		.066	.001	.039	.007	.732
	N	50	50	50	50	50	50	50	50	50
Pre_FCI	Pearson Correlation	.060	.148	.135	.262	1	.818**	-.326*	-.057	.198
	Sig. (2-tailed)	.677	.306	.349	.066		.000	.021	.695	.168
	N	50	50	50	50	50	50	50	50	50
Post_FCI	Pearson Correlation	.103	.145	.054	.444**	.818**	1	.277	.439**	.173
	Sig. (2-tailed)	.477	.316	.709	.001	.000		.052	.001	.229
	N	50	50	50	50	50	50	50	50	50
Diff_FCI	Pearson Correlation	.068	-.009	-.137	.293*	-.326*	.277	1	.816**	-.046
	Sig. (2-tailed)	.639	.952	.342	.039	.021	.052		.000	.750
	N	50	50	50	50	50	50	50	50	50
G_FCI	Pearson Correlation	-.033	-.056	-.033	.377**	-.057	.439**	.816**	1	.056
	Sig. (2-tailed)	.820	.700	.820	.007	.695	.001	.000		.697
	N	50	50	50	50	50	50	50	50	50
G_MRT	Pearson Correlation	.130	.546**	.670**	.050	.198	.173	-.046	.056	1
	Sig. (2-tailed)	.369	.000	.000	.732	.168	.229	.750	.697	
	N	50	50	50	50	50	50	50	50	50

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Figure 4.7: Correlation for MRT, FCI, and Final Grades for Physics1443.

4.6.3 Data for Physics 1444

We did a correlation study between MRT and Final Grades for Physics 1444. Because we have grades given to us as both percentages and letter grades (converted to point values) we did a correlation with both of them. It can be argued that percentages are more sensitive and therefore might have a different correlation than just a final letter grade. The means and standard deviations are presented in Table 4.9. Correlations are presented in Figure 4.8.

Table 4.9: Means and Standard Deviations for Physics 1444

	N	Mean	Standard Deviation
Final Grade Percentage	54	81.12	10.70
Final Grade Numeric	65	2.75	1.02
Pre MRT	65	12.49	4.88

Correlations

		Final_Grade_Percent	Final_Grade_Numeric	Pre_MRT_Scores
Final_Grade_Percent	Pearson Correlation	1	.950**	.243
	Sig. (2-tailed)		.000	.077
	N	54	54	54
Final_Grade_Numeric	Pearson Correlation	.950**	1	.239
	Sig. (2-tailed)	.000		.055
	N	54	65	65
Pre_MRT_Scores	Pearson Correlation	.243	.239	1
	Sig. (2-tailed)	.077	.055	
	N	54	65	65

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 4.8: Correlations for Physics 1444 Final Grades and MRT

4.6.4 Results

The correlation matrix is symmetric for both Physics 1443 and Physics 1444, therefore we only need to look above or below the diagonal. If we look at the data for Physics 1443, ignoring the correlations for each assessment with itself, for the confidence level of α = 0.05, you can see that the only variables that have a correlation are Final Grades and FCI. From previous studies conducted on the Force Concept Inventory, we know that the FCI has a correlation with final grades in mechanics type courses, so students at UT Arlington confirm the validity of the FCI.

If we look at the data for Physics 1444 we can also see there is no correlation between MRT and Final Grades, although the correlation between MRT and Final Numeric Grades come close to the α = .05 level. It also must be pointed out, however, that there is a slight discrepancy between the average percentage grade and average numeric grade. The average for the percentage grades is in the B range whereas the average for the numeric grades is in the C range.

4.7 General Conclusions

We have collected data from Introductory Engineering, Chemistry, Mathematics, and Physics courses. We were looking to reproduce Siemakowski and MacKnight's [1971] study, however we were not able to obtain an accurate sample size for Chemistry and we were not able to obtain an accurate sample size for a test retest assessment using the MRT. In a comparison of Engineering and Physics courses, samples had statistically similar pre-MRT scores, but statistically differed in the post-MRT scores. Physics students had higher post-MRT scores, which leads us to claim that students enrolled in introductory physics courses improve their spatial ability more than students enrolled in introductory engineering courses.

Our introductory physics courses are considered service courses because the dominant populations enrolled in these courses are not physics majors, but are engineering majors. Knowing this we have verified Pallrand and Sieber's [1981] study that suggests simply taking an introductory physics course improves a student's spatial ability more than simply taking an introductory engineering course. Students in these courses are similar populations but the data shows us that physics has a higher impact on spatial ability, as measured by the Mental Rotation Test. This leads us to believe that physics courses have more spatial tasks and spatial related problems than introductory engineering courses at UT Arlington.

We were also trying to extend aspects of Sheryl Sorby's work to the UT Arlington student population to determine if similar interventions as conducted at Michigan Tech might improve STEM retention at UT Arlington. She makes the claim that there is a correlation between spatial ability and success in engineering courses. Engineering students who were able to improve their spatial ability also had a correlation with success in other STEM courses, in particular physics. We collected data for introductory engineering, physics, and final grades for each student who participated. We were only able to find a small correlation between a sub-sample of engineering pre-MRT scores and final grades. We were unable to find correlations for MRT and physics final grades or scores on the Force Concept Inventory, which can also be considered a measure of success in mechanics based physics. These findings leave us with questions about the differences between our samples and other published samples. One obvious difference is that we are using the Mental Rotation Test as a measure of spatial ability where as other groups use the Purdue Spatial Visualization Test: Rotation (PSVT: R) as a measure of spatial ability. The MRT and the PSVT: R have been shown to have a correlation in other research so it is interesting to show that we are not finding a similar correlation between the MRT and physics or engineering grades as had been reported elsewhere.

When looking at the correlation between spatial ability and success in calculus 1, we found a medium effect. These preliminary positive correlations could represent a causal relationship between spatial ability and calculus 1. Students with low spatial skills are not as successful in their introductory Math course (as measured by final grades) as students with high spatial skills. While we cannot say that a causal relationship exists, seeing that the highest attrition at UT Arlington is occurring because of performance in a student's first math and/or chemistry course, the AURAS project will investigate the impact of spatial interventions among low-scoring (MRT) students in calculus.

CHAPTER 5

GENERAL DISCUSSION AND FUTURE WORK

5.1 General Discussion

This document is in the realm of Physics Education Research. In particular, we have laid the foundation for how visual and spatial cognitive abilities are connected to how students learn physics, space science, and related fields. We are in an era where technology is advancing and leading STEM educators to rely more and more on computer simulations and visual diagrams to teach introductory topics. Even lectures are leaving the chalkboard realm and moving towards PowerPoint based lectures where instructors can include links to Youtube videos that demonstrate examples being discussed in the lecture. Textbooks are becoming downloadable with action type examples that students can use instead of traditional paper-based problems that rely on different frames to get the temporal aspect across. In such an environment, a better understanding of how visual and spatial information is related to student learning is very important.

In Chapter 1 we laid the foundation for understanding the role of processing visual/spatial information in producing cognitive load, and we introduced the differences between mental frameworks of novices and experts. Experts possess the ability to reason through problems fairly quickly because they have experience, practice, and stored information that novices do not possess. Novices respond to and rely on surface features of problems, and often use, preconceived ideas, which they may not be aware of, when reasoning through and solving problems. It is also known that experts fail to remember what it is like to be a novice and not understand information.

Our first published study (Chapter 2) demonstrated how novice students interpreted visual information in the case of the use of color as a third dimension in 2-D topographical representations. Most STEM researchers use color to represent additional information on visual representations but fail to realize that common practices in the field are learned processes and not natural occurring phenomena. For example, it is common for people to represent red as hot temperatures and blue as cold temperatures but in astronomy and other space sciences blue stars have the highest temperatures and red stars are generally referred to as cool stars. Unless a students has been exposed to frequencies (or wavelengths) in the Electromagnetic Spectrum, this color scale, where blue is hot and red is cold, would be unknown.

In this study, the major finding was that students were relying on other visual cues instead of using color as primary information. In a student interview it was clearly stated that the mental image of the topography was created first and the color was applied secondary. Even when prompted to maintain a consistent color scheme, the students desire to rely on their own initial intuitions and perceptual model of the diagram dominated their cognitive abilities and prevented them from obtaining a correct interpretation. This demonstrated how difficult it is to overcome and replace either a preconceived idea or a perceptually salient aspect of a diagram, which could lead to incorrect mental models.

Chapter 3 explored another common misconception of incorrect mental models. The cause of Lunar phases is well documented to be a difficult topic to teach in introductory astronomy courses. The cause of Lunar phases is also a case where students often have preconceived ideas that conflict with instruction. We argue that part of the misconceptions come from a lack of spatial understanding about the Sun-Earth-Moon system. In addition, trying to create a mental model of a system, that is not well understood to begin with, increases the cognitive load and prevents the student from forming correct conceptual ideas.

Our study focused on reducing the cognitive load by presenting the system in two different visual diagrams; one in a 3-dimensional stereo view and the other in a 2-dimensional perspective view. We found that both views, 3-dimensional stereo and 2-dimensional perspective, had an increase in comprehension, as measure by the Lunar Phases Concept

Inventory, but there was no statistical difference between the two views. We speculate that one of the reasons there is no statistical difference between the two views, is that the system is not complicated enough to require a 3-dimensional stereo image. The perspective views of the system provided enough spatial information for whatever learning was going to occur. It can also argued that with the increase of detail of perspective views in technologies such as video games, this generation of students are gradually increasing their spatial abilities and, hence, are able to process the perspective information easier. We did not, however, assess students' spatial abilities with this study to verify these claims.

Our last group of studies examined the correlation between spatial abilities of STEM students and success in STEM courses. For theses studies, the spatial ability of a student was measured using the Mental Rotation Test (MRT) and student success was measured by student final grades. Chapter 1 described other STEM fields, which have explored the relationship between spatial abilities and student comprehension. Research in engineering education has been our biggest guide in the sense that they have found a significant correlation between retention and success rates and students' spatial ability.

The Arlington Undergraduate Research-based Achievement in STEM (AURAS) program is interested in improving retention rates of STEM students. A self-study conducted by the AURAS team showed that the highest attrition rates were in the first math course and or chemistry course taken at UT Arlington.

We collected data from introductory chemistry, calculus 1, introductory engineering, and introductory calculus-based physics. We were not able to collect enough data to conduct accurate statistics on our chemistry courses. We did not find a correlation between student spatial abilities and success in engineering for the total data sample. However, in smaller subset of engineering data we did find a small correlation between student spatial abilities and success in introductory engineering courses. Although it must be noted that the average grades for the subset engineering data was higher than the average grades for the total data set, so it leaves us questioning how generalizable theses findings are.

We did not find a correlation between student spatial abilities and success in physics. We did, however, find that in a comparison of spatial abilities between engineering and physics students, physics students had statistically higher spatial scores than engineering students at the end of the semester, even though their initial scores showed no statistical difference. This means that the simple act of taking an introductory physics course improves students' spatial ability more than taking an introductory engineering course.

With calculus, we did find a correlation between students' spatial abilities and success, which generates excitement because our highest attrition rates come from students' first math course. Every STEM student has to take calculus and often physics students are concurrently enrolled in calculus while taking their first physics course. If a students' spatial ability is acting as predictor for success in calculus, we can implement interventions in AURAS' Emerging Scholars Program (ESP) calculus sections to improve students with low scoring spatial skills with the hopes that it will improve success. By improving retention of our calculus students, it might be possible to improve retention of our physics students.

5.2 Future Work

Our results point to a number of future research possibilities examining the role of spatial intelligence in physics, space science, mathematics, and engineering education. In particular, the preliminary result that studying physics improve your spatial skills is something that could be followed up in more detail. Moreover, causal mechanisms should be investigated. What is it about physics instruction that produces the effect?

We would like to finish collecting enough data in order to fully reproduce the Siemakowski and MacKnight [1971] study. This means that we need more data from mathematics and chemistry in order to accurately reproduce the comparison of STEM fields and rotation ability. We also want to look at temporal evolution of STEM students to see how students' spatial ability progresses through their undergraduate career. We were in the process of obtaining data from upper division (junior and senior level) Physics, Chemistry, and Engineering courses and were able to collect data from two upper division physics courses. However there were a number of students who were enrolled in both courses so we were not able to meet the minimum sample size. We obtained a sample size of 44 students from one upper division chemistry course but did not have the data from lower division chemistry to compare against. We were working with the Dean of the College of Engineering to find a few upper division engineering courses for this semester, but the loss of classes due to the snow days deterred many of the faculty from giving us the time.

With the correlation between the pre-MRT scores and calculus 1, we have a very interesting result, namely a real correlations with grades. Is this relationship causal? If it is, what is it about spatial intelligence that allows students to do better in calculus? Is it that better spatial skills give students an advantage in understanding graphical representations, and, for example, using such visual concepts like the tangent to a line at a given point, lead to a better understanding of derivatives? Given that calculus is used in introductory physics, why is there no grade correlation there? Could it be that calculus concepts are not used to a level that would result in a noticeable effect on grades? All of these questions are amenable to research.

Finally, the results here suggest that there could be a real effect on student calculus grades if we attend to spatial intelligence. Therefore AURAS will want to look into the possibility of implementing spatial training interventions in the ESP sections. These same students, many times, are concurrently enrolled in calculus-based mechanics. If spatial skills are acting as a predictor of success in calculus 1 and, additionally, calculus 1 has a high attrition rate, then improving the spatial skills of low scoring students could have an impact on retention in physics as well, even if there is no clear direct signal in the physics data of an impact on grades due to spatial intelligence as measured by the MRT.

APPENDIX A

SURVEYS FROM CHAPTER 2

Survey 1 Part A

Name:____
Major:____

Age: ____

Gender (please circle one): Male Female

Email: ____

Please use Figure 1 to answer the following questions

1. Do you know what a topographic map is (please circle one)? Yes No
2. If yes, how often have you used topographic maps in the past?

   ____

   ____

3. Place a 1 at the <u>highest spot</u> in Figure 1. Figure 1 represents a large-scale geographic feature.
4. Place a 2 at the <u>lowest spot</u> in Figure 1.
5. How many <u>peaks</u> are there in Figure 1?

   ____

6. What might you call such a landform if you saw it in real life?

   ____

7. Why?____

   ____

Survey 1 part B

Name: ____

Please use Figure 2 to answer the following questions

Do not look back at answers from Survey 1A when answering questions from Survey 1B

1. Place a 1 at the highest spot in Figure 2. Figure 2 represents a large-scale geographic feature.
2. Place a 2 at the lowest spot in Figure 2.
3. How many peaks are there in Figure 2?

   ____

4. What might you call such a landform if you saw it in real life?

   ____

5. Why?____

   ____

6. Which of the two figures did you find easiest to understand (please circle one)?

   Figure 1 Figure 2

7. Why?____

   ____

   ____

8. Please circle which set you had (it is listed on the bottom of both figures).

   Set A Set B

Survey 2

Name ____

Year in School (please circle one)   Freshman   Sophomore   Junior   Senior

(Note: Color is not represented the same for this survey as it was in Survey 1)

1. In terms of temperature, what does the color red represent?

   ____

2. Why?____

   ____

3. In terms of temperature, what does the color blue represent?

   ____

4. Why?____

   ____

Where did you learn this?

____

____

____

APPENDIX B

SCANTRON USED FOR FORCE CONCEPT INVENTORY

Force Concept Inventory Answer Sheet

Name (print):______________________ Date:____________

Professor:________________________

One answer per question Erase Completely to Change (pencil)

Completely fill in the box as shown.

[a] [b] ■ [d] [e]

1	[a]	[b]	[c]	[d]	[e]
2	[a]	[b]	[c]	[d]	[e]
3	[a]	[b]	[c]	[d]	[e]
4	[a]	[b]	[c]	[d]	[e]
5	[a]	[b]	[c]	[d]	[e]
6	[a]	[b]	[c]	[d]	[e]
7	[a]	[b]	[c]	[d]	[e]
8	[a]	[b]	[c]	[d]	[e]
9	[a]	[b]	[c]	[d]	[e]
10	[a]	[b]	[c]	[d]	[e]
11	[a]	[b]	[c]	[d]	[e]
12	[a]	[b]	[c]	[d]	[e]
13	[a]	[b]	[c]	[d]	[e]
14	[a]	[b]	[c]	[d]	[e]
15	[a]	[b]	[c]	[d]	[e]
16	[a]	[b]	[c]	[d]	[e]
17	[a]	[b]	[c]	[d]	[e]
18	[a]	[b]	[c]	[d]	[e]
19	[a]	[b]	[c]	[d]	[e]
20	[a]	[b]	[c]	[d]	[e]
21	[a]	[b]	[c]	[d]	[e]
22	[a]	[b]	[c]	[d]	[e]
23	[a]	[b]	[c]	[d]	[e]
24	[a]	[b]	[c]	[d]	[e]
25	[a]	[b]	[c]	[d]	[e]
26	[a]	[b]	[c]	[d]	[e]
27	[a]	[b]	[c]	[d]	[e]
28	[a]	[b]	[c]	[d]	[e]
29	[a]	[b]	[c]	[d]	[e]
30	[a]	[b]	[c]	[d]	[e]

REFERENCES

1. Abbott, D. S., Saul, J. M., Parker, G. W., Beichner, R. J., 2000, "Can One Lab Make a Difference?", *American Journal of Physics,* 68 (S1) S60–S61, doi:10.1119/1.19521.
2. Ambrose, B., S., Heron, P., Vokos, S., and McDermott, L. C., 1999, "Student Understanding of Light as an Electromagnetic Wave: Relating the Formalism to Physical Phenomena", *American Journal Of Physics, v. 67*, pp. 891.
3. Anderson, K. C. and Leinhardt, G., 2002, "Maps as Representations: Expert Novice Comparison of Projection Understanding", *Cognition and Instruction, v. 20(3),* pp. 283.
4. Baddeley, A., 2003, "Working Memory: Looking Back and Looking Forward", *Nature Reviews Neuroscience, v. 4,* pp. 829.
5. Bailey, J. M., and Slater, T. M., 2004, "A Review of Astronomy Education Research", *The Astronomy Education Review, v. 2(2),* pp. 20.
6. Barnet, M., Ymagata-Lynch, L., Keating, T., Barab, S. A., and Hay, K. E., 2005, "Using Virtual Reality Computer Models To Support Student Understanding of Astronomical Concepts", *Journal of Computers in Mathematics and Science Teaching, v.* 24(4), pp. 333.
7. Barrell, G. V. and Cooper, P. J., 1986, "Cognitive Processes in Orienteering: The Interpretation of Contours and Responses to the Map as a Whole", *Scientific Journal of Orienteering, v. 2(1),* pp. 25.
8. Beichner, R. J., 1994, "Testing Student Interpretation of Kinematics Graphs", *American Journal of Physics, v. 68(8),* pp. 750.
9. Beichner, R., 1996, "The Impact of Video Data Analysis on Kinematics Graph Interpretation Skills", *American Journal of Physics, v. 64,* pp. 1272.
10. Biederman, I. and Ju, G., 1988, "Surface Versus Edge-Based Determinants of Visual

Recognition", *Cognitive Psychology, v. 20,* pp. 38.

11. Bell, R. L., and Trundle, K. C., 2007, "The Use of a Computer Simulation to Promote Scientific Conceptions of Moon Phases", *Journal of Research in Science Teaching, v. 45(3),* pp. 346.
12. Bors, D. A. and Vigneau, F., 2011, "Sex Differences on the Mental Rotation Test: An Analysis of Item Types", *Learning and Individual Differences, v. 21,* pp. 129.
13. Botzer, G. and Reiner, M., 2005, "Imagery in Physics Learning – From Physicists' Practice to Naïve Students' Understanding", in *Visualization in Science Education,* eds. J. K. Gilbert, pp. 147, Springer, the Netherlands.
14. Casey, M. B., Nuttall, R., Pezaris, E., and Benbow, C. P., 1995, "The Influence of Spatial Ability on Gender Differences in Mathematics College Entrance Test Scores Across Diverse Samples", *Developmental Psychology, v. 31(4),* pp. 697.
15. Chase, W. G. and Simon, H. A., 1978a, "The Minds Eye in Chess", in *Visual Information Processing,* ed. Chase, W. G., pp. 215.
16. Chase, W. G. and Simon, H. A., 1978b, "Perceptions in Chess", *Cognitive Psychology, v. 4,* pp. 55.
17. Cockburn A., and McKenzie, B., 2002, "Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments", *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves,* Minneapolis, Minnesota, USA, doi:10.1145/503376.503413, April 20-25.
18. Corbin, J. M. and Strauss, A., 1990, "Grounded Theory Research: Procedures, Canons, and Evaluative Criteria", *Qualitative Sociology, v. 13(1),* pp. 3.
19. DeBuvitz, W., 1990, "The Importance of Scale Drawing or: Let's Not Blow Things Out of Proportion!", *Physics Teacher*, *v. 28(9),* pp.604.
20. De Groot, A. D., 1978, *Thought and Choice in Chess (2nd Ed.)*, Mouton Publishers, The Hague, Netherlands.

21. Eason, G. T., 2000, *The Effects of Higher-Order Questioning Strategies on Nonscience Majors' Achievement in an Introductory Environmental Science Course and Their Attitudes Toward the Environment*, Q180.A15 E37, Ph.D. Dissertation, Florida Institute of Technology, Melbourne, Florida.

22. Engle, R. W., 2002, "Working Memory Capacity as Executive Attention", *Current Directions in Psychological Science, v. 11(1),* pp. 19.

23. Gerson, H. B. P., Sorby, S. A., Wysocki, A. F., and Baarmans, B. J., 2001, "The Development and Assessment of Multimedia Software for Improving 3-D Spatial Visualization Skills", *Computer Applications in Engineering Education, v. 9(2),* pp. 105.

24. Gravetter, F. J. and Wallnau, L.B., 2007, *Statistical Behavioral Sciences, Seventh Edition,* Thomas High Education, Richmond, California.

25. Guay, R. B., 1977, Purdue Spatial Visualization Test: Rotation. West Lafayette, IN Purdue Research Foundation.

26. Hake, R. R., 1998, "Interactive-Engagement Versus Traditional Methods: A Six-Thousand-Student Survey of Mechanics Test Data for Introductory Physics Courses", *American Journal of Physics, v. 66(1),* pp. 64.

27. Halloun, I., & Hestenes, D., 1985, "The Initial Knowledge State of College Physics Students," *American Journal of Physics*, *v. 53(11),* pp. 1043.

28. Hansen, J., Barnett, M., Makinster, J., and Keating, T., 2004, "The Impact of Three-Dimensional Computational Modeling on Student Understanding of Astronomy Concepts: A Qualitative Analysis", *International Journal of Science Education, v. 26(14),* pp. 1555, doi:10.1080/09500690420001673766.

29. Hestenes, D., and Wells, M., 1992a, "A Mechanics Baseline Test," *The Physics Teacher*, *v. 30*, pp. 159.

30. Hestenes, D., Wells, M., and Swackhamer, G.,1992, "Force Concept Inventory," *The Physics Teacher*, *v. 30(3),* pp. 141.

31. *How People Learn: Brain, Mind, Experience, and School,* 1999, National Academy Press, Washington DC.

32. Hubona, G. S., Wheeler, P. N., Shirah, G. W., and Brandt, M., 1999, "The Relative Contributions of Stereo, Lighting, and Background Scenes in Promoting 3D Depth Visualization", *ACM Transactions on Computer-Human Interaction (TOCHI), v. 6(3),* pp. 214, doi:10.1145/329693.329695, September.

33. Kastens, K. A., Kaplan, D. and Christie-Blick, K., 2001, "Development and Evaluation of a Technology-Supported Map-Skills Curriculum, Where are We?", *Journal of Geoscience Education*, v. *49(3),* pp. 249.

34. König, C. J., Bühner, M., and Mürling, G., 2005, "Working Memory, Fluid Intelligence, and Attention are Predictors of Multitasking Performance, but Polychronicity and Extraversion are not," *Human Performance, v. 18(3),* pp. 243.

35. Kosslyn, S. M., 1995, "Mental Imagery", in *Invitation to Cognitive Science (Vol. 2): Visual Cognition,* eds. S. M. Kosslyn and D. N. Osherson, pp. 267, MIT Press, Cambridge, MA.

36. Kosslyn, S. M., 2005, "Mental Images and the Brain", *Cognitive Neuropsychology, v. 22(3/4),* pp. 333.

37. Kozhevnikov, M., Hegarty, M., and Mayer, R., 2002, "Spatial Abilities in Problem Solving in Kinematics", in *Diagrammatic Representation and Reasoning,* eds. M. Anderson, M., Meyer, B., and Olivier P., pp. 155, Springer, London.

38. Kozhevnikov, M., Motes, M. A., and Hegarty, M., 2007, "Spatial Visualization in Physics Problem Solving", *Cognitive Science, v. 31,* pp. 549.

39. Küçüközer, H., 2008, "The Effect of 3D Computer Modeling on Conceptual Change About Seasons and Phases of the Moon", *Physics Education, v. 43(6),* pp. 632.

40. Küçüközer, H., Korkusuz, M. E., Küçüközer, H. A., and Yurumezoglu, K., 2009, "The Effect of 3D Computer Modeling and Observation-Based Instruction on the Conceptual

Change Regarding Basic Concepts of Astronomy in Elementary School Students", *Astronomy Education Review,* 010104, 10.3847/AER2009006.

41. Kuhn, T. S., 1996, *The Structure of Scientific Revolutions, Third Edition,* University of Chicago, Chicago, Illinois.

42. Land, B. and LoPerfido, A., 1993, "A Case for Scientific Visualization in Undergraduate and Graduate Classrooms", *Journal of Computing in Higher Education, v. 4(2),* pp. 3.

43. Larkin, J., McDermott, J., Simon, D. P., and Simon, H. A., 1980, "Expert and Novice Performance in Solving Physics Problems", *Science, v. 208,* pp. 1335.

44. Liben, L. S, 2006, "Education for Spatial Thinking", in *Handbook of child psychology: V 4. Child Psychology in Practice*, 6th ed., series eds. Damon W. and Lerner, R., and eds. Renninger, K. A., and Sigel, I. E., pp. 197, Hoboken, New Jersey: Wiley.

45. Linn, M. C. and Petersen, A. C., 1985, "Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis", *Child Development, v. 56,* pp. 1479.

46. Lindell, R. S., and Olsen, J. P. 2002, "Developing the Lunar Phases Concept Inventory," *Proceedings of the 2002 Physics Education Research Conference*, ed. by S. Franklin, J. Marx and K. Cummings. PERC Publishing, NY.

47. Lopez, R. E., and Hamed, K. 2004, "Student Interpretations of 2-D and 3-D Renderings of the Substorm Current Wedge," *Journal of Atmospheric and Solar-Terrestrial Physics, v.* 66, pp. 1509.

48. Lopez, R. E., Koskinen, H. E. J., Pulkkinen, T. I., Bösinger, T., Potemra, T. A., and McEntire, R. W., 1993, "Simultaneous Observation of the Poleward Expansion of Substorm Electrojet Activity and the Tailward Expansion of Current Sheet Disruption in the Near-Earth Magnetotail", *Journal Geophysical Research*, v. *98*, pp. 9285.

49. Mayer, R. E., 1997, "Multimedia Learning: Are We Asking the Right Questions?", *Educational Psychologist, v. 31*(1), pp. 1.

50. Mazur, E., 1997, *Peer Instruction: A User's Manual,* Prentice Hall, Upper Saddle River,

NJ.

51. McDermott, L. C. and the Physics Education Group at the University of Washington, 1996, *Physics by Inquiry Volume II*, New York: John Wiley & Sons, Inc.

52. McDermott, L.C., and Redish, E.F., 1999, "Resource Letter: PER-1: Physics Education Research," *American Journal of Physics*, *v. 67(9),* pp. 755.

53. McDermott, L. C., 2001, "Oersted Medal Lecture 2001: 'Physics Education Research—The Key to Student Learning'", *American Journal of Physics, v. 69(11),* pp. 1127.

54. Meltzer, D. E., 2005, "Relation Between Students' Problem-Solving Performance and Representational Format", *American Journal of Physics, v. 73(5),* pp. 463.

55. Mestre, J. P., 1994, "Cognitive Aspects of Learning Science", Chapter 3 in *Teacher Enhancement for Elementary and Secondary Science and Mathematics: Status, Issues, and Problems*, eds. Fitzsimmons, S. J. and Kerpelman, L. C., Abt Associates, Inc., Cambridge, Massachusetts.

56. Miyake, A., Friedman, N. P., Rettinger, P. A., Shah, P., and Hegarty, M., 2001, "How are Visuospatial Spatial Working Memory, Executive Functioning, and Spatial Abilities Related? A Latent-Variable Analysis", *Journal of Experimental Psychology: General, v. 130(4),* pp. 621.

57. Morin, P. J., Leigh, J., van Keken, P., and Johnson, A., 2003, "The State of the GeoWall", *Eos Trans. AGU, v. 84(46),* Fall Meet. Suppl., Abstract ED31A-01.

58. Orion, N., Ben-Chain, D., and Kali, Y., 1997, "Relationship Between Earth-Science Education and Spatial Visualization", *Journal of Geoscience Education, v. 45,* pp. 129.

59. Pallrand, G. J. and Seeber, F., 1984, "Spatial Ability and Achievement in Introductory Physics", *Journal of Research in Science Teaching, v. 21(5),* pp. 507.

60. Rapp, D. N., 2005, "Mental Models: Theoretical Issues for Visualizations in Science Education", in *Visualization in Science Education*, ed. Gilbert, J. K., pp. 43, Springer,

Netherlands.

61. Rapp, D. N., Culpepper, S. A., Kirby, K., and Morin, P., 2007, “Fostering Students' Comprehension of Topographic Maps”, *Journal of Geoscience Education, v. 55 (1),* pp. 5.

62. Redish, E. F., 1999, “Millikan Lecture1998: Building a Science of Teaching Physics,” *American Journal of Physics, v. 67(7),* pp. 562.

63. Richtmyer, F. K., 1933, “Physics is Physics,” *American Physics Teacher (which became American Journal of Physics), v. 1(1),* pp. 1.

64. Salthouse, T. A., Babcock, R. L., Skovroned, E., Mitchell, D. R. D., and Palmon, R., 1990, Age and Experience Effects in Spatial Visualization”, *Developmental Psychology, v. 26(1),* pp. 128.

65. Schnepps, H., and Sadler, P., *A Private Universe* (video), 1987, Harvard-Smithsonian Center for Astrophysics, Science Media Group.

66. Shepard, R. N. and Metzler, J., 1971, “Mental Rotation of Three-Dimensional Objects”, *Science, v. 171(3972),* pp. 701.

67. Siemankowski, G. and MacKnight, F., 1971, “Spatial Cognition: Success Prognosticator in College Science Courses”, *Journal of College Science Teaching, v. 59,* pp. 1.

68. Sorby, S. A., 1999, “Developing 3-D Spatial Visualization Skills”, *Engineering Design Graphics Journal, v. 63(2),* pp. 21.

69. Sorby, S. A., 2005, “Assessment of a ‘New and Improved’ Course for the Development of 3-D Spatial Skills”, *Engineering Design Graphics Journal, v. 69(3),* pp.6

70. Sorby, S. A., 2009, “Educational research in Developing 3-D Spatial Skills for Engineering Students”, *International Journal of Science Education, v. 31(3),* pp. 459.

71. Sorby, S. A. and Baartmans, B. J., 1996, “A Course for the Development of 3-D Spatial Visualization Skills”, *Engineering Graphics Journal, v. 60(1),* pp. 13.

72. Spence, I. and Feng, J., 2010, “Video Games and Spatial Cognition”, *Review of*

General Psychology, v. 14(2), pp. 92.

73. Spencer, S. J., Steele, C. M., and Quinn, D. M., 1999, "Stereotype Threat and Women's Math Performance", *Journal of Experimental Social Psychology, v. 35*, pp. 4.
74. Stieff, M., Bateman, R., and Uttal, D., 2005, "Teaching and Learning with Three-Dimensional Representations", in *Visualization in Science Education,* ed. Gilbert, J. K., pp. 93, Springer, Netherlands.
75. Stieff, M., 2007, "Mental Rotation and Diagrammatic Reasoning in Science", *Learning and Instruction,* DOI: 10.1016/j.learinstruc.2007.01.012.
76. Stieff, M., 2010, "When is a Molecule Three Dimensional? A Task-Specific Role for Imagistic Reasoning in Advanced Chemistry", *Science Education,* DOI: 10.1002/sce.20427.
77. Strauss, A. and Corbin, J., 1990, *Basics of Qualitative Research: Grounded Theory Procedures and Techniques,* Sage Publications, Newbury Park, CA.
78. Sweller, J., 1994, "Cognitive Load Theory, Learning Difficulty, and Instructional Design", *Learning and Instruction, v. 4,* pp. 295.
79. Taylor, H., Renshaw, C., and Choi, E., 2004, "The Effects of Multiple Formats on Understanding Complex Visual Displays", *Journal of Geoscience Education, v. 52(1),* pp. 115.
80. Thornton, R. K., and Sokoloff, D. R., 1998, "Assessing Student Learning of Newton's Laws: The Force and Motion Conceptual Evaluation and the Evaluation of Active Learning Laboratory and Lecture Curricula," *American Journal of Physics*, *v. 66(4),* pp. 338.
81. Trumper, R., 2000, "University Students' Conception of Basic Astronomy Concepts", *Physics Education*, *v. 35*, pp. 9.
82. Trundle, K. C., Atwood, R. K., & Christopher, J. E., 2007, "A Longitudinal Study of Conceptual Change: Preservice Elementary Teachers' Conceptions of Moon Phases",

Journal of Research in Science Teaching, v.44 (2), pp. 303.

83. Trundle, K. C., Atwood, R. K., & Christopher, J. E., 2002, “Preservice Elementary Teachers’ Conceptions of Moon Phases Before and After Instruction”, *Journal of Research in Science Teaching, v.39* (7), pp. 633.
84. Tufte, E. R., 1997, *Visual Explanations: Images and Quantities, Evidence and Narrative*, Graphics Press, Cheshire, Connecticut.
85. Turner, N. E., Lopez, R. E., Hamed, K. M., 2003, “Effectiveness of GeoWall Technology in Conceptualizing the Earth-Moon System,” *Eos Trans. AGU, v. 84(46),* Fall Meet. Suppl., Abstract ED31A-06.
86. Tversky, B., 1993, “Cognitive Maps, Cognitive Collages, and Spatial Mental Models”, in *Spatial Information Theory: A Theoretical Basis for GIS, Proceedings COSIT ’93*, eds. Frank, A.U. and Campari, I., Lecture Notes in Computer Science, 716, pp.14, Springer: Berlin.
87. Volbracht, S., Domik, G., Shahrbabaki, K., and Fels, G., 1997, “How Effective are 3D Display Modes?”, *Proceedings CHI'97*, pp. 540, March.
88. Vandenberg, S. G. and Kuse, A. R., 1978, “Mental Rotation, A Group Test of Three-Dimensional Spatial Visualization”, *Perceptual and Motor Skills, v. 47,* pp. 599.
89. Ware, C., and Franck, G., 1996, “Evaluating Stereo and Motion Cues for Visualizing Information Nets in Three Dimensions”, *ACM Transactions on Graphics, v. 15(2),* pp 121, April.
90. Winn, W., Stahr, F., Sarason, C., Fruland, R., Oppenheimer, P., and Lee, Y., 2006, “Learning Oceanography from a Computer Simulation Compared with Direct Experience at Sea”, *Journal of Research in Science Teaching*, *v. 43(1),* pp. 25.
91. Cid, X. C., Lopez, R. E., and Lazarus, S. M., 2009, “Issues Regarding Student Interpretation of Color as a Third Dimension on Graphical Representations”, *Journal of Geoscience Education, v. 57(5),* pp. 372.

92. Cid, X. C. and Lopez, R. E., 2010, “The Impact of Stereo Display on Student Understanding of Phases of the Moon”, *Astronomy Education Review, v. 9(1),* 010105, 10.3847/AER2009044.

93. Zacks, J. and Tversky, B., 1999, “Bars and Lines: A Study of Graphic Communication”, *Memory and Cognition*, v. *27(6)*, pp. 1073.

BIOGRAPHICAL INFORMATION

Ximena Clara Cuicatl Cid is of Mexican American and Native American decent. She was born in Sacramento, CA to Armando Cid and Josie Talamantez. She has an older brother, Armando Jr, an older sister, Zenaida, a twin sister, Amar Azucena, and a younger brother, Miguel. She earned her Bachelor's degree in Astrophysics from the University of California at Berkeley in 2005 and her Master's degree in Physics from the University of Texas at Arlington in 2010 as progress towards her Ph.D.

While she was working towards her bachelor's degree she was conducting research in Space Physics at the Space Sciences Laboratory, under the guidance of Janet Luhmann, located in the hills behind UC Berkeley. She also had the opportunity to teach with the Professional Development Program (PDP), the first scholars program in the country. Working with PDP sparked her interest with how people learn science. For graduate school she sought out Dr. Ramon Lopez because he conducted research in both Space Physics and Physics Education Research (PER). She quickly learned that her passion was in PER and started to conduct research solely in PER.

Ximena has been honored with several awards in research from numerous institutions, including the Society for the Advancement of Chicanos and Native Americans in Science (SACNAS), the Texas Section of the American Physical Society (APS), and the University of Texas at Arlington. She has served as a Graduate Student Liaison for the Physics Department and a representative for the department to the dean of the College of Science.

Her future aspirations include joining the faculty at a research institution. In the mean time she would like to focus on research for a few years in the form of a Postdoc. Her current greatest excitement comes from being a new Tia.