

# Do 3D Stereoscopic Virtual Environments Improve the Effectiveness of Mental Rotation Training?

James Quintana, Kevin Stein, Youngung Shon , and Sara McMains\*

*\*corresponding author*

*Department of Mechanical Engineering  
University of California, Berkeley, CA 94720-1740*

**ABSTRACT** - *This research investigated the effectiveness of two types of mental rotation training: one using exercises performed in a 3D stereoscopic virtual environment, and one using exercises on a standard 2D computer display. The 41 subjects, volunteers from a freshman Engineering Design Graphics class, were randomly assigned to one of the two training groups. Performance was assessed using Vandenberg and Kuse's Mental Rotations Test (MRT-A) before and after the training. Data were analyzed for statistically significant differences between training groups and between genders. We found no statistically significant difference in improvement in subjects' mental rotation ability between the 2D mental rotation training and the 3D virtual reality training. In contrast to prior related work, we found that the difference between male and female subjects' performance remained statistically significant after the virtual reality training.*

## **I. Introduction**

Studies of undergraduate engineering students' performance on the rotation portion of the Purdue Spatial Visualization Test show significant performance differences between students of different sexes, and between students with different cultural backgrounds, e.g. students from the US and Europe

versus the United Arab Emirates (Leake, 2000). There is strong evidence that spatial reasoning skills, at least as measured by this particular test, can be significantly improved through practice, and that such training can improve the retention of students with low initial scores in an engineering program, particularly for female students (Sorby, 2000). In the current research, we studied whether the effectiveness of mental rotation training was improved by delivering it in a 3D virtual reality environment, compared to traditional computer-based training methods.

## **II. Related Work**

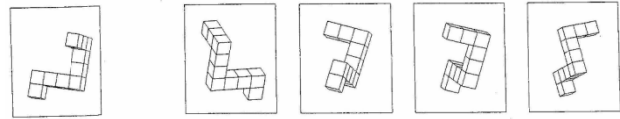
Our work was primarily inspired by Rizzo et al.'s study on the potential of a single session of virtual reality training to improve mental rotation ability. They found that although the training did not have a statistically significant effect on performance of the overall study population, there were interesting patterns for subgroups. Among subjects who scored poorly (less than or equal to 50% correct) on the initial mental rotation test, the improvement by subjects in the virtual reality group compared to those in the control group was statistically significant. They also found that while the difference between males' and females' performance before training was statistically significant, this difference was no longer statistically

significant after they both went through the virtual reality training.

### III. Methodology

**Subjects:** Subjects were recruited from a lower division Basic Engineering Design Graphics course at UC Berkeley. The research assistants made a presentation in lecture mentioning the possible benefit of the training for students who wished to improve their mental rotation abilities. Follow-ups with a sign-up sheet, flyers, and a message on the class website yielded 47 volunteers out of roughly 300 students enrolled in the class over two semesters. Volunteers were not compensated for their participation. Two subjects served for pilot runs, one subject did not complete the post-test, and three subjects were randomized into a third treatment group that was discontinued due to insufficient sample size, for a total of 41 subjects for the study.

**Assessment:** Performance was assessed using Peters et al.'s (1995) redrawn Vandenberg and Kuse (1978) Mental Rotations Test, MRT(A). In this test, subjects are presented drawings of target objects formed from about ten small cubes. For each target shown in the left column, they are asked to choose which two of four possible answers to its right are pictures of rotated versions of the same object (see Figure 1). Both choices must be correct to receive credit for the question. There are a total of 24 objects on the test, divided into two sections. Following Peters et al., we gave the subject three minutes to complete each 12-question section, separated by an optional four-minute break.



**Figure 1. Sample rotation question (Peters et al., 1995).**

After the subject completed the assigned training, they took a modified version of the test as the post-test. Following Rizzo et al., this post-test was constructed by switching the position of each sample object in the left column and one of the matching objects to its right. The same timing procedure was used for the post-test.

**Training:** The subjects were randomly assigned (by a roll of a die) into one of two treatment groups: the Virtual Reality 3D Rotation Group (VR Group) or the CD-ROM Exercise Group (CD Group). The first group had 18 subjects, 5 female and 13 male; the second group had 23 subjects, 7 female and 16 male.

The first training method used a 3D stereoscopic virtual reality environment. Subjects wore head-tracked Crystal Eyes VR stereoscopic shutter glasses and sat at a FakeSpace Immersadesk with a large stereo screen, so that objects displayed appeared to project out of the screen in 3D. A picture of a 3D object made of cubes, similar to the objects pictured in the Mental Rotations Test, would appear in one corner of the screen. A rotated version of the same 3D object, with orientation controlled by the subject, would appear in the other corner. The subject was asked to rotate the second object until it appeared to have the same orientation as the stationary object. Once they indicated that they thought they had a match, if they were correct, a new pair of objects would appear on the screen; otherwise, the question would not advance.



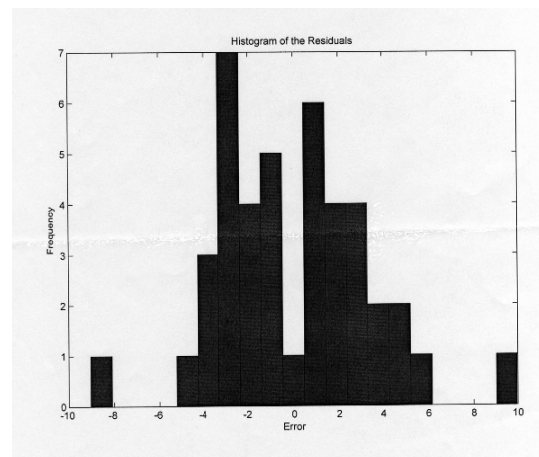
**Figure 2. The virtual reality desk and keyboard with trackball.**

Subjects controlled the objects with a trackball and keyboard system (Figure 2). The trackball provided rotation over two degrees of freedom, and a pair of keys controlled the third degree of freedom. The investigator explained the use of the trackball/keyboard system and the training exercises and the subject was given one minute to practice rotating a single object with this system. Subjects then completed a set of 90 rotation exercises. The subject was offered a short (up to 5 minutes) break halfway through the 90 rotations. This training was based on Rizzo's virtual reality training software, with three modifications: questions only advanced once subjects indicated that they were ready to check their answer, manipulation was with the trackball and keyboard rather than a Flock of Birds tracking device, and 90 instead of 140 exercises were completed to reduce subject fatigue. The most challenging questions were retained. Total completion time was roughly 40 minutes, exclusive of the break.

The second training method consisted of completing approximately 40 minutes of computer exercises practicing visualizing the results of rotating objects. These were visualization exercises rather than active rotation exercises, performed on a standard computer without virtual-reality capabilities. The

exercises were the "Rotation of objects about a single axis" and "Rotation of objects about two or more axes" modules of Sorby et al.'s CD-ROM (2003). Subjects read the instruction slides and took the interactive quizzes, finishing the two sections or working for 40 minutes, whichever came first. Subjects were offered a short (up to 5 minutes) break after 20 minutes.

**Analysis:** Multiple regression analysis (Freedman, 2005) was implemented using the Statistics Toolbox in MATLAB. Since our sample sizes are less than 30, we used Student's t-distribution in place of a normal distribution. For each test, we measure the probability  $p$  that the measured differences occurred by chance, based on the variability of the data. To verify that the residuals were normally distributed, the difference between the measured values and those predicted by the regression were plotted. An example histogram for the residuals from the comparison between improvement in the VR group and the CD group is shown in Figure 3, consistent with a normal distribution.



**Figure 3. Histogram of residuals.**

#### IV. Results

Consistent with other researchers, we found that male scores were higher than female scores by a statistically significant amount. Because this tendency is well-established, a one-tailed test is appropriate for calculating probabilities. The effect was found both on the pre-test ( $p = 0.0011$ ) and post-test ( $p = 0.0002$ ). Across both training groups, males got an average of 15.4 questions correct on the pre-test, whereas females got an average of 10.8 questions correct. On the post-test, males got an average of 15.7 questions correct, whereas females got an average of 11.8 questions correct. Looking only at the group with VR training, these sex differences remained statistically significant after the training with over 99% probability ( $p = 0.0028$ ), in contrast to Rizzo et al.'s findings.

To test the null hypothesis that there was no difference in overall improvement between the VR Group and the CD Group, we use a two-tailed test, since the deviation could be in either direction. We found no statistical significance ( $p = 0.5283$ ).

The CD Group only improved by 0.65 questions on average, and subjects in the VR Group only improved by 0.21 questions on average. For comparison, the average scores for a group of undergraduates in the bachelor of science program at the University of Guelph were 11.5 for males and 9.9 for females taking the MRT(A) for the first time, and 21.0 for males and 15.6 for females when the same subjects retook the identical test a week later (Peters 1995).

One hypothesis that could explain this difference in improvement is that these students, who scored lower than the Berkeley students initially, had more room to improve. However, when we analyzed only the Berkeley students who had scored 12 or below (less than or equal to 50% correct), we again found no

statistically significant difference in improvement between the two training methods ( $p = 0.3081$ ).

Another hypothesis is that fatigue on the part of the subjects could have played a role. The research assistants observed that the subjects appeared less enthusiastic by the end of the experiment, as well as concerns raised by the subjects themselves. Quantitatively, we can see that although subjects' scores improved by 1.34 questions on part 1 of the post-test, they decreased by 0.74 questions on part 2 of the post-test. Thus fatigue might have affected our data. However, in terms of differences in the effect of the two types of training, there was not a statistically significant effect on the improvement on part 1 of the post-test between the two training groups, either when considered as a whole ( $p = 0.8728$ ) or if looking only at those who initially scored 12 or lower ( $p = 0.2373$ ).

Looking at each training group individually, the amount of improvement (over both parts of the test) was also not statistically significant for either training ( $p = 0.3002$  for the CD group,  $p = 0.4363$  for the VR group).

However, when looking at low scorers only, or part 1 scores only to control for fatigue, effects of some statistical significance were found. For the CD training group, scores improved by a statistically significant amount on part 1 ( $p = 0.0467$ ). For low scorers with CD training, statistical significance was borderline when considering total score ( $p = 0.0528$ ), but much stronger statistical significance was found for improvement of low scorers on part 1 scores ( $p = 0.0098$ ). For the VR training group, scores improved on part 1 with only borderline statistical significance ( $p = 0.0608$ ). For low scorers with VR training, the improvement was still not statistically significant for total scores ( $p = 0.1440$ ), and only borderline for

improvement of low scorers on part 1 scores ( $p = 0.0699$ ).

A summary of all of the p-values is provided in Table 1.

**Table 1. P-values for all hypotheses tested.**

<b>Female Scores vs. Male Scores</b>	
Pre-Test	0.0011
Post-Test	0.0002
<b>Female Scores vs. Male Scores, VR Group Only</b>	
Pre-Test	0.0374
Post-Test	0.0028
<b>Improvement in CD Group vs. in VR Group</b>	
Total Score	0.5283
Part 1 Score	0.8728
<b>Improvement in CD Group vs. in VR Group, Low Scores Only</b>	
Total Score	0.3081
Part 1 Score	0.2373
<b>CD Group Pre-Test vs. CD Group Post-Test</b>	
Total Score	0.3002
Part 1 Score	0.0467
<b>CD Group Pre-Test vs. CD Group Post-Test, Low Scores Only</b>	
Total Score	0.0528
Part 1 Score	0.0098
<b>VR Group Pre-Test vs. VR Group Post-Test</b>	
Total Score	0.4363
Part 1 Score	0.0608
<b>VR Group Pre-Test vs. VR Group Post-Test, Low Scores Only</b>	
Total Score	0.1440
Part 1 Score	0.0699

## V. Conclusions

Disappointingly from a pedagogical point of view, we found no evidence that 3D stereoscopic virtual reality environments are a magic bullet for training

mental rotation abilities. Nor did we find evidence that a single training session in a virtual reality environment reduced sex differences to a statistically insignificant level. The primary utility of the availability of the VR training would seem to be that the novelty factor helps motivate more students to participate in training, rather than an increase in the effectiveness of the training.

## VI. Acknowledgments

We gratefully acknowledge the generosity of Albert “Skip” Rizzo for sharing his group’s virtual reality training source code with us. We also thank Mikhail Traskin and John Chung for statistics consulting services.

## VII. References

- Freedman, D. A. (2005). *Multiple Regression, Statistical Models, Theory and Practice*. New York: Cambridge University Press.
- Leake, J. M. (2000). Visualization Testing in the United Arab Emirates. *EDGD 55th Annual Mid-Year Conference Proceedings*, pages 47–53. Engineering Design Graphics Division, American Society for Engineering Education.
- Peters, M., Laeng, B., Lathan, K., Jackson, M., Zaiyouna, R., and Richardson, C. (1995). A redrawn Vandenberg and Kuse Mental Rotations Test: Different versions and factors that affect performance. *Brain and Cognition*, 28, 39-58.
- Rizzo, A. A.; Buckwalter, G. J.; McGee, J. S.; Bowerly, T.; van der Zaag, C.; Neumann, U.; Thiebaut, M.; Kim, L.; Pair, J.; Chua, C. (2001). Virtual Environments for Assessing and Rehabilitating Cognitive/Functional Performance. *Presence: Teleoperators and Virtual Environments* 10 (4): 359-374 .
- Sorby, S. A. (2000). Improving the Spatial Skills of Engineering Students: Impact on Graphics Performance and Retention. *EDGD 55th Annual Mid-Year Conference Proceedings*, pages 67–73. Engineering Design Graphics Division, American Society for Engineering Education.

Sorby, S.; Wysocki, A.; Baartmans, B. (2003), software accompanying *Introduction to 3D Spatial Visualization: an active approach*, Thomson Delmar Learning.

Student [Gosset, W S] (1908). The Probable Error of a Mean. *Biometrika* 6 (1): 1–25.

Vandenberg, S. G., and Kuse, A. R., (1978). Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills* 47, 599-604.

### Appendix A

A summary of our raw data, with the number of questions completed correctly before and after the training, is provided below.

#### Virtual Reality Training Group

Sex	Pre-Test			Post-Test		
	Set 1	Set 2	Total	Set 1	Set 2	Total
M	7	7	14	10	7	17
F	4	2	6	7	2	9
F	7	4	11	6	3	9
M	11	5	16	10	8	18
M	5	3	8	7	7	14
M	9	5	14	8	6	14
M	8	5	13	9	5	14
M	9	9	18	12	5	17
M	9	8	17	9	5	14
F	8	6	14	7	7	14
M	5	4	9	7	4	11
M	10	11	21	12	7	19
F	6	5	11	7	4	11
F	8	4	12	7	4	11
M	10	12	22	11	6	17
M	9	10	19	11	7	18
M	8	8	16	12	6	18
M	11	10	21	12	9	21

#### Traditional Computer CD-ROM Training Group

Sex	Pre-Test			Post-Test		
	Set 1	Set 2	Total	Set 1	Set 2	Total
F	8	7	15	9	7	16
F	6	5	11	6	6	12
M	12	9	21	12	6	18
F	5	3	8	9	2	11
M	6	6	12	12	7	19
M	11	6	17	8	10	18
M	10	9	19	9	7	16
M	11	10	21	12	4	16
M	8	6	14	10	5	15
M	4	2	6	7	6	13
M	5	5	10	9	1	10
M	9	8	17	12	7	19
F	8	8	16	12	7	19
M	6	5	11	8	5	13
M	6	4	10	7	7	14
M	9	6	15	9	4	13
M	7	6	13	7	4	11
M	10	10	20	10	5	15
F	6	4	10	7	5	12
M	11	11	22	9	9	18
F	2	3	5	3	4	7
M	8	7	15	11	7	18
F	6	4	10	6	4	10