Practicing in Virtual Reality Improves Mental Rotation Ability: Lower Scorers Benefit More

Elaine Jiang*
Department of Computer Science
Brown University

David H. Laidlaw†
Department of Computer Science
Brown University

Abstract

This paper presents results from an evaluation of using virtual reality techniques to measure improvement of mental rotation ability. We compared the performance improvements of university students in Mental Rotation Tasks (MRTs) in two different settings: a control setting in which participants practice mental rotation in a desktop environment, and an experimental setting in which participants practice mental rotation tasks in a virtual reality environment. Pre- and post-learning session MRT performance measures show that practicing mental rotation tasks improves the ability to do those tasks ($p < .001$) and that a virtual reality environment can more effectively aid learning ($p = .017$). We also find that low spatial ability participants benefit significantly more when learning in a virtual reality environment compared to learning in a desktop computer environment ($p = .015$). Survey feedback from participants suggests that VR learning sessions were helpful because the MRTs were displayed in 3D, possibly alleviating extraneous cognitive load. These results support the proposition that VR can be utilized as an aid for spatial task learning, particularly among those with lower spatial abilities.

Index Terms: Virtual reality—Spatial ability—Mental rotation—Visualization techniques

1 Introduction

Spatial ability, the ability to solve problems of navigation and visualize objects from different angles, is one of the main components of human intelligence [32]. High spatial ability is important for success in STEM fields, as well as in occupations such as sports, medicine, and aviation [34]. This ability can be divided into three further sub-domains: 1) spatial relations, the ability to manipulate objects in space quickly and accurately; 2) spatial visualization, the ability to analyze complex spatial information that involves multiple steps; and 3) spatial orientation, the ability to perceive spatial relationships with respect to the position of one's body [7].

One direction that many recent studies have taken is the use of modern technology to facilitate spatial ability research [8]. Among the most popular technologies employed is virtual reality (VR), computer-generated environments that extend physical reality by immersing users completely [8]. VR has been shown to aid spatial learning, providing an environment that potentially alleviates extraneous cognitive load [18].

This paper advances our understanding of which technologies are more effective in facilitating mental rotations learning. Various studies have examined the effects of immersive VR environments and desktop environments in aiding spatial orientation and spatial visualization learning [17]. However, few have studied the effectiveness of these technologies in improving mental rotation ability [16]. This study extends previous work by examining two additional factors: learning mode and spatial ability. We assign participants to learn in either a desktop or a VR learning mode. Within each learning mode, we classify each participant as having high or low spatial ability. We investigate whether these factors affect the learning of mental rotations.

More specifically, this work contributes to the fields of virtual reality and spatial learning in the following ways:

1. We confirm that practicing MRTs on a desktop computer or in VR can improve mental rotation ability, as suggested in related studies.

2. We show that VR can facilitate the learning of MRTs at least as well as traditional media such as desktop computers.

3. We observe that low spatial ability participants benefit more when learning MRTs in a VR environment compared to a desktop environment.

While this study recruited university students as participants, our results can lead to further research with more vulnerable groups. Insights from this study can help determine whether virtual reality has the potential to help individuals from various demographics enhance their spatial abilities.

In this paper, we first outline the background for this study and discuss related work. We then present the experimental procedure, design choices, and materials used. We discuss our results and demonstrate how they relate to the contributions stated above. Finally, we highlight the implications of our work and how they relate to the fields of virtual reality and spatial learning.

2 Background and Related Work

Here we discuss the background and relevant work for this study. We are inspired by research indicating that modern technology can both aid spatial learning and improve spatial ability. We begin by describing cognitive load theory (CLT), a theory that claims that learning is more effective when extraneous cognitive load is reduced [2], and we explain how VR can help reduce cognitive load. We then introduce MRT, a core component in the study, and the surrounding research about MRTs in VR.

2.1 Cognitive Load Theory

Cognitive load theory describes how our working memory can only handle a certain amount of activity at one time [5]. When the amount of cognitive load exceeds the capacity of working memory, learning is no longer effective [33].

The three types of cognitive load are intrinsic load, extraneous load, and germane load [33]. Intrinsic load is the notion that all tasks have an associated difficulty that can not be changed. However, instructors can make a task less difficult by breaking them down into sub-tasks [33]. Extraneous load represents cognitive load that occurs as information is presented [11]. Good instructors will aim to reduce this type of load by presenting the information in a way that can be easily processed. Finally, germane load refers to the cognitive load that creates schema to process information learned [33]. CLT states that intrinsic load, extraneous load, and germane

*e-mail: ejiang1@cs.brown.edu
†e-mail: dhl@cs.brown.edu

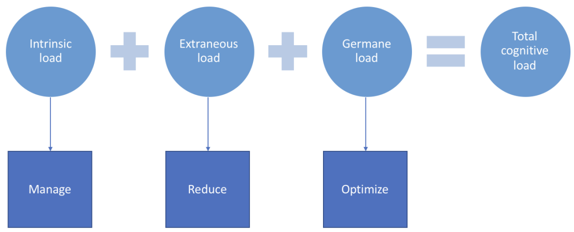

FIGURE 1: This diagram illustrates the cognitive load theory. Here, we see that intrinsic load, extraneous load, and germane load collectively make up total cognitive load [33]. The total cognitive load is limited by the capacity of our working memory [2]. Overall, we want to manage intrinsic load, reduce extraneous load, and optimize germane load for the most effective learning [5].

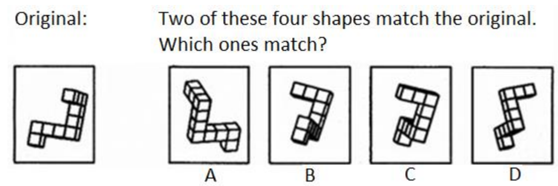

FIGURE 2: In this mental rotation task, the stimulus on the far left is the target figure. Of the remaining four stimuli (A, B, C, D), two are geometrically identical to the target figure. Completing an MRT involves three distinct cognitive stages: 1) making mental images of the stimulus from all angles; 2) mentally rotating the stimulus to compare it to the target figure; and 3) deciding whether two stimuli are geometrically identical [15].

load collectively make up cognitive load. In order to make learning more effective, instructors are encouraged to manage intrinsic load, reduce extraneous load, and maximize germane load, as seen in Figure 1.

Evidence suggests that presenting information in VR can reduce extraneous load. An experiment conducted to measure EEG-based cognitive load found that watching a 2D instructional video on a computer produced a significantly higher cognitive load index compared to watching a 3D instructional video [6]. Our study seeks to take advantage of this property in VR to facilitate spatial learning.

2.2 Spatial Ability and Learning Environments

There are two opposing hypotheses surrounding how spatial ability affects performance in 3D learning environments. The ability-as-compensator hypothesis suggests that low spatial ability individuals benefit more when learning from models compared to their high spatial ability counterparts due to less cognitive load when processing the models [20]. The ability-as-enhancer hypothesis posits the reverse—that those with high spatial abilities have enough cognitive load remaining to process the models [20].

Currently, there is research that supports both hypotheses. On one hand, researchers have found that high spatial ability individuals were able to learn more effectively when presented with 3D visualizations [12, 14, 20]. On the other hand, low spatial ability individuals were able to construct and process schema with the help of computerized 3D models [6, 18].

In this study, we seek to verify whether learning in a 3D learning environment is more beneficial than learning in a 2D desktop environment for low spatial ability individuals.

2.3 Mental Rotation Task (MRT)

The core skill assessed in this study is mental rotation ability. Mental rotation is the ability to rotate representations of 2D and 3D objects within one's imagination (see Figure 2).

2.4 Spatial Learning in Virtual Reality

Practicing spatial exercises in VR can help improve spatial performance [16, 22, 27]. One study found that VR effectively trained astronauts to orient themselves in simulated sub-gravity environments [24]. While this study focuses on learning spatial orientation, our study focuses on learning of mental rotations.

In a study exploring the effectiveness of learning anatomy in a VR environment, researchers discovered that low spatial ability learners demonstrated a greater performance achievement in a VR-based learning environment than high spatial ability learners [18]. These results support the hypothesis that low spatial ability learners benefit more from the VR-based learning environment because they struggle to mentally reconstruct their own visualizations [14]. While Lee & Wong (2014) [18] examined the difference in performance between low and high spatial ability individuals on a spatial visualization and memory task, our work aims to verify this hypothesis within the sub-domain of spatial relations.

Our work thus utilizes these concepts but focuses on using VR to aid the practice and improvement of mental rotation ability. Our purpose is to leverage spatial information by taking advantage of virtual reality's immersiveness and physical navigation to address problems like cognitive load.

2.5 MRTs in Virtual Reality

Previous work has also been done to investigate the effects of rotating 3D objects in virtual reality.

A study revealed that while differences in MRT performance were seen between the sexes after practicing MRTs on paper, no significant differences were observed in a virtual environment [30]. Other studies have demonstrated that practicing mental rotations in VR can aid in the rehabilitation of visuospatial abilities in those with acquired brain injury [29], and enhance spatial rotation skills in hard-of-hearing children [25].

Like the studies above, we employ mental rotation as the core skill of the experiment, but we investigate two different factors. We evaluate the effectiveness of practicing mental rotations in a desktop environment versus in a VR environment. In addition, our study determines whether individual spatial ability differences affect the learning experience in VR environments.

3 Research Goals

The main goal of this research project is to evaluate whether a VR environment is more effective in improving mental rotation ability than non-VR environments. A secondary goal is to determine whether there is a difference in improvement between low spatial ability participants and high spatial ability participants. More specifically, we aim to: (1) compare participants' performance between baseline and post-training conditions; (2) compare the effectiveness of desktop learning versus VR learning; (3) compare low spatial ability participants' performance in VR to low spatial ability participants' performance in the desktop environment.

Following these goals, we form our hypotheses:

1. H_P: Practicing mental rotation tasks improves the ability to do those tasks.

2. H_{VR}: A virtual reality training environment is more effective than a conventional learning environment (e.g., a desktop environment).

3. H_A: Low spatial ability participants will benefit more in a VR learning environment than high spatial ability participants.

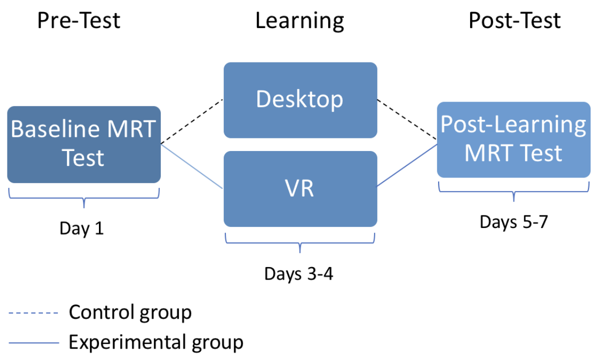

FIGURE 3: This diagram outlines the experimental process. All participants take a baseline MRT assessment before being split into either the control group, in which they will practice MRTs using a desktop computer, or the experimental group, in which they will practice MRTs in VR. Finally, all participants take the post-learning MRT assessment.

4 METHODOLOGY

We now give an overview of the study design and then describe the test used to assess mental rotation ability, followed by the details on the experimental procedure and materials used.

4.1 Experimental Design

We had three experimental settings:

1. A baseline assessment that tested participants' mental rotation ability
2. A learning session in which participants were asked to practice MRTs on a desktop computer
3. A learning session in which participants were asked to practice MRTs in a VR environment

Overall, each participant took the baseline MRT assessment, completed the learning session, and took a post-learning MRT assessment. Participants completed the learning session two to three days after the initial baseline assessment, and were assigned to either the desktop training environment or the VR environment. The post-learning assessment was administered two to three days after each participant's learning session.

4.2 Design Discussion

4.2.1 Baseline Assessment

A baseline MRT assessment was administered at the beginning of the experiment to all participants. Although it exposed participants to mental rotation tasks before the actual training, the baseline assessment scores were used to ensure a similar number of high spatial ability participants and low spatial ability participants within each control and experimental group.

We chose to administer Vandenberg and Kuse's MRT assessment as it is widely used in spatial ability research to assess mental rotation ability [35].

4.2.2 Post-Learning Assessment

A post-learning MRT assessment was given to participants after the completion of their learning session. The assessment is of identical structure to the baseline MRT assessment. Each participant took the post-learning MRT assessment two to three days after completing the learning session, allowing a few days of rest. In a pilot study, we found that administering the post-learning assessment immediately after participants complete the learning session often lead to worse scores on the assessment than in the baseline. From participants' anecdotal feedback, we observed that administering the test immediately after the learning session can lead to feelings of fatigue and mental exhaustion. Our experimental design enforces two to three days of rest between the learning session and the post-learning assessment. However, we acknowledge that the extended time between practicing and the post-learning assessment can also "flush out" short-term learning effects [3].

4.2.3 MRT Learning Program

In both VR and desktop environments, the MRT learning program contains 50 tasks. As shown in Figure 4, each task contains a target stimulus and four other stimuli, two of which are geometrically identical to the target figure. Participants complete a task by selecting which two figures they believe are identical to the target figure.

Participants are limited to one minute per task. We restricted the amount of time per task in order to provide an upper bound on the total time spent on training.

4.2.4 VR and Desktop Environments

The VR learning program is nearly identical to the desktop learning program. One key difference is that the VR program provides stereoscopy, whereas the desktop program does not. Another difference is that the VR environment uses ray-casting for selection while the desktop program uses mouse selection.

4.3 Procedure

4.3.1 Baseline Assessment

To assess participants' mental rotation ability, we used an MRT assessment created by Vandenberg and Kuse [35]. This is a paper and pencil assessment that comprises of 24 MRT questions. We used the first 12 for the baseline assessment. As in the learning program, each question in the assessment presents five stimuli. There is one target stimulus, and four other stimuli, two of which are geometrically identical to the target stimuli. The assessment works as follows: the assessment is out of 24 points, and each question is worth two points. 1 point is given for each correctly identified stimulus. In addition, a 1 point penalty is assigned for each wrong answer in order to discourage guessing. Participants are given a time limit of four minutes to complete the assessment.

4.3.2 Demographics Survey

All participants took a demographics survey before taking the initial baseline MRT assessment. The survey asked for the participant's age, sex, area of study, whether they wore glasses or contacts, and how much prior experience they had with VR. They had four options to describe their prior experience with VR: 1) No experience; 2) Tried it before; 3) Lots of experience; and 4) Expert.

4.3.3 MRT Learning Session

Two to three days after taking the baseline assessment, each participant took part in an MRT learning session. The learning session comprised of 50 MRT questions, as detailed in the design section above. The maximum amount of time each participant could spend practicing tasks was 50 minutes. There was no minimum amount of time each participant had to spend on the tasks. Regardless of whether the learning session was done in VR or on a desktop computer, all participants completed the sessions while sitting.

The difficulty of an MRT can be derived by computing the angle of rotation from a stimuli to its target stimuli [31]. The greater the angle of rotation, the more difficult the task [31]. In the MRT learning program, tasks become more difficult as they progress. Specifically, the first 25 tasks feature stimuli rotated between 90 to 120 degrees along one axes; the next 10 tasks feature stimuli rotated between 90 to 150 degrees along two axes; the final 15 tasks feature stimuli rotated between 90 to 180 degrees along three axes.

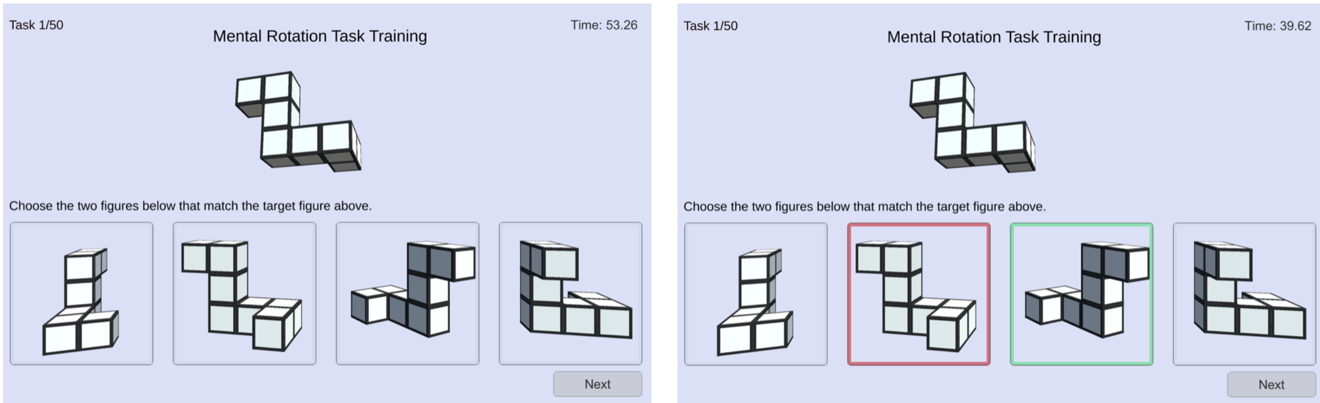

FIGURE 4: The MRT learning program: the program can be taken using a desktop computer or using VR. The program comprises of 50 MRTs, with a time limit of one minute per task. On the left, an MRT is displayed. On the right, two figures are selected as figures that potentially match the target figure. The red outline conveys that the participant's answer is incorrect. The green outline conveys that the participant's answer is correct. The immediate feedback system aims to help individuals learn faster.

As participants select stimuli that match the target stimuli, they receive immediate feedback for their answers (see Figure 4). If they are correct, the figures are outlined in green. Otherwise, they are outlined in red.

4.3.4 Post-Learning Assessment

Two to three days after completing the learning session, each participant took the final post-learning assessment. The four-minute post-learning assessment is identical to the baseline assessment, using the latter twelve Vandenberg and Kuse MRT questions.

4.4 Materials

4.4.1 Apparatus

We used an HTC Vive Pro [13] head-mounted display (HMD) set up in a room approximately three meters by three meters. The room was quiet during the experiment. Occasionally, there was one other person working quietly in the room who was not a participant nor the experimenter. All participants received the same HMD, and the locations of the base stations remained the same throughout the experiment.

4.4.2 Computers

Two graphics computers of the same model (Asus PA248 1920X1200; 24.1 inches; 50 - 76 Hz frame rate) were used during the experiment: one for participants in the experimental group, and one for participants in the control group.

4.4.3 Mental Rotation Tasks

The MRTs were retrieved from a study that sought to replicate and digitize the tasks from the Vandenberg and Kuse assessments [26]. The training program was implemented in Unity (v.2018.2.7f).

4.4.4 Participants

We recruited 22 participants (5 female) in total for the experiment. Participants were undergraduate and graduate students aged 18-26. We placed 11 participants in the control group and 11 in the experimental group. The control group practiced MRTs in a desktop environment, while the experimental group practiced them in a VR environment. We further split each group into a low spatial ability subgroup and a high spatial ability subgroup based on the initial MRT assessment.

TABLE 1: Average improvement of individuals from baseline and post-learning MRT scores

Statistic	**Value**
Average Improvement	4.78
Std	2.96
t-value	6.84
p	$< .001$

5 Analyses and Results

This section presents our analyses and interpretation of the experimental results.

We begin by clarifying the data obtained and the measures used in the analyses. Then we present the analyses of performance between baseline and post-training assessments using paired t-tests.

We dropped 2 participants after the completion of the experiment, as they scored perfectly on both MRT assessments (24 out of 24 points). We thus ended up with 10 participants from the control group and 10 participants from the experimental group (5 female).

The qualitative data we collected were from the demographics survey given before the training session, as mentioned in the methods.

The quantitative data we collected were from the mental rotation task assessments. There are two scores for each participant: the baseline score and the post-learning score.

5.1 Comparing Baseline and Post-Training MRT Scores (H_P)

As mentioned before, each participant completed two MRT assessments over the duration of the experiment: one at the beginning (baseline), and one after the training session (post-training). Here, we investigate whether practicing mental rotation tasks can improve the ability to do those tasks.

5.1.1 Method

To compare the baseline MRT score and post-training MRT scores, we calculate performance improvements within individuals to account for the differences between demographics (e.g., sex, experience with VR, etc.). We use a one sample t-test in R to test H_p, and we assume that the data is normally distributed.

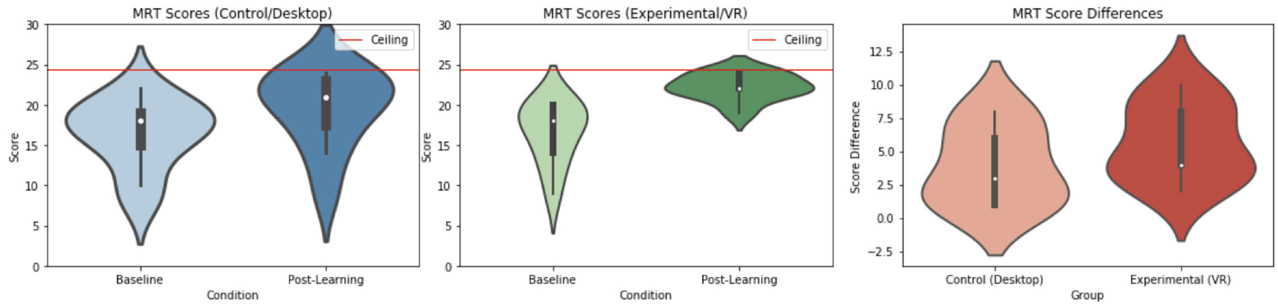

FIGURE 5: Results of MRT scores for the control (desktop) and experimental (VR) groups. The graph on the left illustrates the distribution of MRT scores in the control group, with $\mu_{baseline} = 16.56$ and $\sigma_{baseline} = 3.81$, while $\mu_{post} = 22.22$ and $\sigma_{post} = 1.64$. In the experimental group (middle)), $\mu_{baseline} = 16.36$ and $\sigma_{baseline} = 4.25$, while $\mu_{post} = 18.81$ and $\sigma_{post} = 5.54$. We see a much smaller variance in the post-learning distribution in the experimental group than that of in the control group. Finally, the graph on the right depicts the MRT score differences (post-learning score - baseline score) for each group.

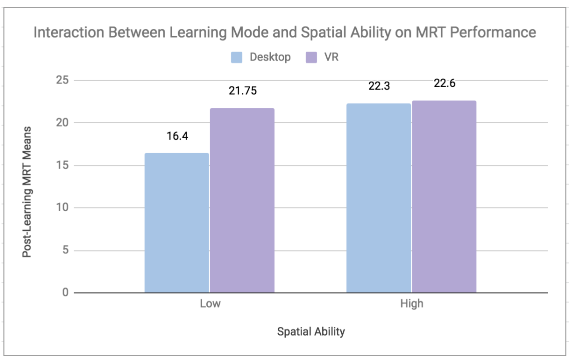

FIGURE 6: We compare the post-learning MRT score means within both learning modes (desktop vs. VR) and spatial ability (low vs. high). We find that low spatial ability participants benefit more in VR, scoring an average of 21.75 points out of 24 in the VR condition. This is 5.35 points higher than the mean of the low spatial ability group in the desktop condition. For high spatial ability participants, the learning mode did not have a significant effect on their post-learning MRT scores.

5.1.2 Results

We compute the improvement between the baseline score and post-learning score in each participant. The data shows that the average improvement is 4.78 points (out of 24 total points). The variance, however, is quite high (2.96 points). We find that participant improvement is highly significant ($p < .001$), and thus we confirm H_P.

5.2 Comparing Baseline and VR Post-Training MRT Scores (H_{VR})

The goal of this section is to see whether a virtual reality environment can be more effective than a desktop computer environment in improving mental rotation ability. In other words, we test H_{VR}. This hypothesis relies on H_P, which we have confirmed in the previous section. To test H_{VR}, we create a multi-variable regression model and fit it to our data, taking into account the covariates that may have affected participants' scores (e.g. experience with VR, sex, area of study, etc.).

5.2.1 Method

We first applied a two sample t-test using R, but we did not find a significant correlation between learning mode and performance improvement ($p = .064$). Because the results have many potential covariates, our next step was to fit the data to a regression model.

TABLE 2: Covariates and the Results of Multivariable Regression Analysis on MRT Assessment Improvement

Variable	Coef.	Std. Error	p
Condition (VR)	.506	.209	**.017**
Spatial Ability	-.737	.313	**.001**
Familiarity with VR	-.357	.298	.098
Learning Experience	.381	.275	.095
(Intercept)	1.543e-16	0	1.000

Our multi-variable regression model takes the form:

$$y_i = \beta_c x_{i,c} + \beta_1 x_{i,1} + \ldots + \beta_5 x_{i,5} + \varepsilon_i$$

where y_i is the performance improvement within individual i (post-learning MRT score subtracted by baseline MRT score), β is the covariate coefficient, and ε_i is the error term for that particular observation. β_l represents the learning mode covariate (e.g., VR versus desktop), and thus is the covariate we focus on in this section.

Here, spatial ability refers to performance on the baseline assessment. Individuals who scored less than the median (17 out of 24) were categorized as having low spatial ability. Otherwise, they were placed in the high spatial ability group.

We will also clarify how we quantified some of the variables. Familiarity with VR had four possible values: 0, indicating no previous experience in VR; 1, indicating having tried VR before; 2, indicating lots of experience; and 3, indicating expert. Learning experience was evaluated on a scale of 0-2: 0, indicating a negative experience (e.g., had physical discomfort during the learning session); 1, indicating a neutral experience; and 2, indicating a positive experience (e.g., found the session engaging and helpful).

Other potential covariates such as sex, age, major, and familiarity with MRTs were eliminated because they did not produce significant p-values in our regression model. We use the LMM model in R to fit to our data of 20 observations.

5.2.2 Results

The regression results for each covariate are presented in Table 2. The model generally fits well for two factors, learning mode ($\beta_l = .506$) and spatial ability ($\beta_1 = -.737$). It shows that learning in a virtual reality environment is highly correlated with performance improvement ($p = .017$). We therefore confirm H_{VR}, that is, that a VR environment can be more effective than a desktop computer environment in improving mental rotation ability. We also note that spatial ability significantly affects performance improvement ($p = .001$); in particular, lower spatial ability participants improve more. However, we must consider the ceiling effects that higher spatial ability participants may encounter. In addition, other demographic factors such as age, familiarity with VR, field of study, sex, learning time, learning experience, and familiarity with MRTs do not appear to correlate with performance improvement. We come back to these findings in the discussion.

5.3 Correlation Between Performance Achievement and Spatial Ability (H_A)

In the previous section, we observed that lower spatial ability participants demonstrated a greater difference in performance achievement than higher spatial ability participants. Here we investigate whether lower spatial ability participants benefited more from one learning mode over another by comparing their post-learning MRT scores.

5.3.1 Method

To find the correlation between performance achievement and spatial ability, we first divide each group (control and experimental) into high spatial ability and low spatial ability participants. We use the baseline MRT scores of participants as the indicator for spatial ability. After defining the two spatial ability groups, we compute the mean of post-training MRT scores of the two groups within their respective condition (control or experimental). We use a Welch two sample t-test in R to test H_A.

5.3.2 Results

We find the interaction between learning mode and spatial ability on MRT performance significant for the low spatial ability group ($p = .015$) but not for high spatial ability group ($p = .32$). Thus, we confirm H_A. In particular, we observe that participants in the low spatial ability group who learned in VR performed better ($\mu = 22.3$) than those in the low spatial ability group who learned in the desktop environment ($\mu = 16.4$), as seen in Figure 6. We also note that participants in the high spatial ability group who learned in VR performed only slightly better ($\mu = 22.6$) than those in the high spatial ability group who learned in the desktop environment ($\mu = 21.75$). This small improvement is due to the ceiling effect, which we come back to in the discussion.

6 Discussion

Here we discuss the results and observations from the experiment, and note the factors that may have produced unexpected results.

6.1 Small Sample Size

We had 20 total participants, after removing two who scored perfect on both the baseline and post-learning MRT assessment. The small sample size may limit the generalizability of the data. In addition, all of our participants were college-age students, which may further limit generalizability.

6.2 Ceiling Effect

We encountered the ceiling effect during this experiment. High-scoring participants have little room to improve ($\mu_{high} = 22.45$); on average these participants can at most improve 1.55 points. Our spatial ability covariate in the regression analysis for H_{VR} does not take into account this effect.

Thus, we focused on investigating whether learning mode affected low spatial ability participants. While we cannot confirm that high spatial ability participants benefit more in one learning mode over another, we can draw conclusions for the low scorers.

6.3 Individual Differences

Although our sample size was small, we noticed many differences between individuals during our experiment. For instance, some individuals finished the MRT assessment well within the 4 minute limit, while others struggled to finish even two-thirds of the assessment in that time. In addition, participants had varying levels of familiarity towards MRTs, but this did not seem to have a significant effect on performance improvement ($p = .29$).

During the learning sessions, some individuals complained that they felt fatigued, saying that they *"have a headache,"* or *"have trouble concentrating"* and getting through the 50 practice tasks. Others were engaged and enthusiastic, asking lots of questions during their learning session. According to our regression analyses, learning experience appeared to have a marginal effect on the improvement of mental rotation ability ($p = .095$).

The speed at which participants completed the learning session varied between individuals as well. While a few participants took the full 50 minutes to complete all of the tasks, many took about half the time, and a few others took only 10 minutes. However, learning time did not appear to significantly affect performance improvement ($p = .13$).

These differences illustrate that even among our small sample size, there is high variability among demographic factors, spatial ability, and engagement in the task. Overall, we did not observe significant correlations between these differences and performance improvement on the MRT assessment. Future studies might consider using NASA-TLX or a similar survey to gauge these differences between individuals [23].

6.4 Variance of Difficulty in MRT Questions

The Vandenburg and Kuse MRT assessment contains a total of 24 questions [35]. We split the assessment into two separate assessments by taking the first 12 for the baseline assessment and the remaining for the post-learning assessment. However, we later discovered that the difficulty level of questions was not randomly distributed as we had previously assumed. One study found that the first set of twelve questions has an average difficulty level of 63.6, measured by average percent correct on each of the questions. It found that the remaining 12 questions has an average difficulty level of 32.1 [4]. Thus, the latter 12 questions are more difficult than the first 12.

This variance of difficulty between the two assessments may have skewed our results. In particular, it is likely that if the two assessments were more similar in difficulty level, the performance improvements observed may have been more significant.

6.5 Biases

One bias of our study is the disproportionate number of males to females in our study. Because our participant population comprised of 75% males, this could have skewed our results. Furthermore, literature suggests that men outperform women in spatial tasks [28]. More specifically, men outperform women in MRTs [19]. While we did not see a strong correlation between sex and performance improvement in mental rotations ($p = .63$), we acknowledge that a sample size of 5 females is not representative of the general population.

Response bias, the tendency for someone to respond inaccurately to questions, may also be present in this experiment [10]. For example, those who are interested in virtual reality or wish to try VR for the first time are more inclined to sign up for the study. Even though all participants were told before the study that they would be given the opportunity to play games in VR after the study, we observed that some participants were disappointed when they were assigned to the control group, and thus could not use the virtual reality devices.

In addition, a large portion (approximately 85%) of our participant population majored in STEM (Science, Technology, Engineering, Mathematics). Research has shown that STEM majors are more likely to have higher spatial ability than students in non-STEM majors [1]. In this study, we did not observe STEM majors outperforming non-STEM majors at a significant level ($p = .95$).

Finally, in the experimental condition, the experimenter had to interact more with participants (e.g., introduce virtual reality and help them put on the virtual reality devices); in the control conditions, all sessions and tasks were performed in a familiar desktop environment with less experimenter interaction.

6.6 Learning Time and Durability

Although we have discussed a variety of factors and biases that could have affected the results of the experiment, one important factor is the time spent on learning mental rotations. In the experiment, each participant spent at most 50 minutes ($\mu_{time} = 23$) practicing MRTs. The reason we allowed a minute per task during learning (as compared to 4 minutes per 12 questions during the assessment) was to give participants time to receive and process feedback during the learning session. Although we incorporated a feedback system to accelerate the process of learning [9], this is a short amount of time to learn a skill, especially if the skill is novel to some participants.

Participant feedback supports the potential future design decision to have multiple learning sessions. Some felt that they were getting *"warmed-up"* in the learning session and did not get a chance to fully develop strategies to solve MRTs. Others suggested creating shorter but more frequent learning sessions. A participant noted that they felt *"tired after doing 50 tasks straight but completing 25 each day would probably have been more helpful."* Future extensions of this work might hence consider a more longitudinal approach to learning or improving upon a spatial skill.

We did not test for the durability of mental rotation ability, and this may be a factor of interest in future work. Participants in this study were given one post-learning assessment a few days after their learning sessions. One potential improvement to this design is to administer multiple assessments over a month after participants undergo the learning program.

6.7 Spatial Learning, VR, and Future Design

Our work suggests that learning mental rotations in VR is more effective than learning mental rotations in a desktop environment. This may also support the notion that processing 3D, stereoscopic visualizations in VR uses less extraneous cognitive load than the amount required to process visualizations from a 2D desktop display. Therefore, virtual reality may provide an environment that requires less mental effort when learning or improving upon a skill. Individuals with lower spatial abilities appear to benefit more, supporting the ability-as-compensator hypothesis that those with lower spatial abilities have advantages in learning when visualizations are externally presented to them [20].

Although not explicitly tested in this experiment, evidence suggests that improving a spatial skill can lead to transfer effects in other spatial tasks that are not directly related to the learned spatial skill. Literature has shown that practicing mental rotation tasks can improve the ability to complete other spatial rotation tasks (i.e., in object rotation and perspective-taking tasks), and that these benefits can persist even after a month [21].

Overall, improving spatial ability comes with various advantages. High spatial ability is correlated with success in STEM fields. As demonstrated in this study, having a higher spatial ability may lead to more effective learning regardless of the learning environment.

In this study, we have presented a successful method to help individuals improve mental rotation ability through the aid of VR. We hope that this will inspire future work in the intersection of learning and VR. We believe that our findings can also contribute to research into the rehabilitation of spatial skills in more vulnerable groups. Subsequent visualization studies may be able to reduce the variance in participants' spatial abilities through an initial MRT training session. Training all participants on MRT tasks beforehand can potentially eliminate the need to exclude low-scoring participants in studies that require a certain level of spatial ability. As demonstrated in this study, even short MRT learning sessions can significantly improve one's mental rotation ability. Our work also supports the proposition that processing stimuli in 3D stereoscopic environments such as VR is more effective than processing stimuli in 2D displays. Finally, future studies may want to consider different spatial tasks and determine whether the learning effects are the same across a broader set of skills.

7 Conclusion

This paper presents an evaluation of using VR and desktop techniques to assist in learning mental rotations. We conducted an experiment to compare the MRT performance improvements of university students when learning in a desktop environment and when learning in a VR environment. Our quantitative results of pre- and post-learning session MRT assessment scores demonstrate a significant advantage of learning in a VR environment over learning in a desktop environment. Although we encountered the ceiling effect for high MRT scorers, we find that low spatial ability participants benefit more when learning in a virtual reality environment compared to learning in a desktop computer environment, thereby supporting the ability-as-compensator hypothesis. We observe that factors including sex, majoring in STEM, and familiarity with VR did not significantly affect improvement in mental rotation ability. Finally, we discovered open questions such as the durability and transfer effects of improved spatial skills after practicing. Our work suggests that 3D, stereoscopic visualization techniques can help individuals learn more efficiently than 2D visualization techniques. It demonstrates that VR can effectively aid spatial task learning, particularly among those with lower spatial abilities.

Acknowledgments

First and foremost, I'd like to thank my thesis advisor, David Laidlaw, for his guidance and encouragement throughout this past year. I truly could not have completed this project without him. I recall actually writing about wanting to do research with David in my application essays, so this has really been a dream come true!

I'd also like to thank James Tompkin for his help in this project and for his advice regarding graduate school and beyond. In addition, a big thank you to Fumeng Yang for her valuable feedback and expertise in statistics.

Finally, I'd like to thank my parents and my sister Rena for always being there for me and for always motivating me to challenge myself. A huge shout-out to my roomie and best friend Jenny for providing me with so much emotional support, as well as for helping me pilot the user study.

Thank you all for making my experience at Brown so rewarding and unforgettable.

References

[1] L. Andersen. Visual-spatial ability: Important in stem, ignored in gifted education. *Roeper Review*, 36:114–121, 04 2014. doi: 10.1080/02783193.2014.884198

[2] S. A. W. Andersen, P. T. Mikkelsen, L. Konge, P. Cayé-Thomasen, and M. S. Sørensen. Cognitive load in distributed and massed practice in virtual reality mastoidectomy simulation. *The Laryngoscope*, 126 2:E74–9, 2016.

[3] A. Baddeley. Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience*, 4(3):281–288, 1992. doi: 10.1162/jocn.1992.4.3.281

[4] A. Caissie, F. Vigneau, and D. Bors. What does the mental rotation test measure? an analysis of item difficulty and item characteristics. *The Open Psychology Journal*, 2:94–102, 12 2009. doi: 10.2174/1874350100902010094

[5] P. Chandler and J. Sweller. Cognitive load theory ans the format of instruction. *Cognition and Instruction*, 8(4):293–332, 1991. doi: doi:10.1207/s1532690xci0804_2

[6] A. Dan and M. Reiner. Eeg-based cognitive load of processing events in 3d virtual worlds is lower than processing events in 2d displays. *International Journal of Psychophysiology*, 122:75 – 84, 2017. Neural Patterns of Learning, Cognitive Enhancement and Affect. doi: 10.1016/j.ijpsycho.2016.08.013

[7] T. Donnon, J. G. DesCteaux, and C. Violato. Impact of cognitive imaging and sex differences on the development of laparoscopic suturing skills. *Canada Journal of Surgery*, 48(5):387–93, 2005.

[8] A. Dünser, K. Steinbügl, H. Kaufmann, and J. Glück. Virtual and augmented reality as spatial ability training tools. In *Proceedings of the 7th ACM SIGCHI New Zealand Chapter's International Conference on Computer-human Interaction: Design Centered HCI*, CHINZ '06, pp. 125–132. ACM, New York, NY, USA, 2006. doi: 10.1145/1152760.1152776

[9] M. L. Epstein, A. S. Lazarus, T. B. Calvano, K. A. Matthews, R. A. Hendel, B. B. Epstein, and G. M. Brosvic. Immediate feedback assessment technique promots learning and corrects inaccurate first responses. *The Psychological Record*, 52:187–201, 2002.

[10] A. Furnham. Response bias, social desirability and dissimulation. *Personality and Individual Differences*, 7(3):385–400, 1986. doi: 10.1016/0191-8869(86)90014-0

[11] P. Ginns. Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6):511–525, 2006. doi: doi:10.1016/j.learninstruc.2006.10.001

[12] T. N. Hoffler and D. Leutner. The role of spatial ability in learning from instructional animations — evidence for an ability-as-compensator hypothesis. *Computers in Human Behavior*, 27(1):209 – 216, 2011. Current Research Topics in Cognitive Load Theory. doi: 10.1016/j.chb.2010.07.042

[13] HTC Vive. *Vive Pro*, 2019.

[14] T. Huk. Who benefits from learning with 3d models? the case of spatial ability. *Journal of Computer Assisted Learning*, 22:392–404, 2006.

[15] A. M. Johnson. Speed of mental rotation as a function of problem-solving strategies. *Perceptual and Motor Skills*, 71(3):803–806, 1990. doi: doi:10.2466/pms.1990.71.3.803

[16] H. Kaufmann and D. Schmalstieg. Mathematics and geometry education with collaborative augmented reality. *Computers & Graphics*, 27(3):339 – 345, 2003. doi: 10.1016/S0097-8493(03)00028-1

[17] M. Kozhevnikov. The role of immersive 3d environments in three-dimensional mental rotation. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 52:2132–2136, 09 2008. doi: 10.1177/154193120805202710

[18] A. Lee and K. W. Wong. Learning with desktop virtual reality: Low spatial ability learners are more positively affected. *Computers & Education*, 2014. doi: doi: 10.1016/j.compedu.2014.07.010

[19] Y. Maeda and S. Y. Yoon. A meta-analysis on gender differences in mental rotation ability measured by the purdue spatial visualization tests: Visualization of rotations. *Educational Psychology Review*, 25:69–94, 03 2013. doi: 10.1111/j.1460-2466.1992.tb00815.x

[20] R. Mayer and V. Sims. For whom is a picture worth a thousand words?: Extensions of a dual-coding theory of multimedia learning. *Journal of Educational Psychology Glenberg & Langston*, 86:389–401, 09 1994. doi: 10.1037/0022-0663.86.3.389

[21] C. Meneghetti, E. Borella, and F. Pazzaglia. Mental rotation training: transfer and maintenance effects on spatial abilities. *Psychological Research*, 80, 2000. doi: doi:10.1007/s00426-014-0644-7

[22] J. L. Mohler. Improving spatial ability with virtual reality: A review of research & applications. *WebNet Journal: Internet Technologies, Applications & Issues*, 3(1):28–35, 2001.

[23] NASA. *Nasa Task Load Index (TLX) v. 1.0 Manual*, 1986.

[24] C. M. Oman, W. L. Shebilske, J. T. Richards, T. C. Tubre, A. C. Beall, and A. Natapoff. Three dimensional spatial memory and learning in real and virtual environments. *Spatial cognition and computation*, 2(4):355–372, 2000.

[25] D. Passig and S. Eden. Virtual reality as a tool for improving spatial rotation among deaf and hard-of-hearing children. *CyberPsychology and Behavior*, 4(6):681–686, 2001. doi: doi.org/10.1089/109493101753376623

[26] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson. A redrawn vandenberg and kuse mental rotations test - different versions and factors that affect performance. *Brain and Cognition*, 28(1):39 – 58, 1995. doi: 10.1006/brcg.1995.1032

[27] J. Regian, W. Shebilske, and J. M. Monk. Virtual reality: An instructional medium for visual-spatial tasks. *Journal of Communication*, 42:136 – 149, 02 2006. doi: 10.1111/j.1460-2466.1992.tb00815.x

[28] D. Reilly, D. L. Neumann, and G. Andrews. Gender differences in spatial ability: Implications for stem education and approaches to reducing the gender gap for parents and educators. *Visual-Spatial Ability: Transforming Research into Practice*, pp. 195–224, 2017. doi: doi: 10.1007/978-3-319-44385-0_10

[29] A. Rizzo, J. Buckwalter, U. Neumann, C. Kesselman, M. Thiebaux, P. Larson, and A. V. Rooyen. The virtual reality mental rotation spatial skills project. *CyberPsychology and Behavior*, 1(2):113–119, 1998. doi: doi.org/10.1089/cpb.1998.1.113

[30] A. Rizzo, J. Buckwalter, U. Neumann, C. Kesselman, M. Thiebaux, P. Larson, and A. V. Rooyen. Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia*, 42(4):555–562, 2004. doi: doi.org/10.1016/j.neuropsychologia.2003.08.014

[31] R. N. Shepard and J. Metzler. Mental rotation of three-dimensional objects. *Science*, 171(3972):701–703, 1971. doi: 10.2307/1731476

[32] R. J. Sternberg. Human intelligence. *Encyclopdia Britannica*, 2017.

[33] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12:257–285, 1988. doi: doi:10.1207/s15516709cog1202_4

[34] D. H. Uttal, D. I. Miller, and N. S. Newcombe. Exploring and enhancing spatial thinking: Links to achievement in science, technology, engineering, and mathematics? *Current Directions in Psychological Science*, 22(5):367–373, 2013. doi: 10.1177/0963721413484756

[35] S. G. Vandenberg and A. R. Kuse. Mental rotations: A group test of three-dimensional spatial visualization. *Perceptual and motor skills*, 47:599–604, 11 1978. doi: 10.2466/pms.1978.47.2.599