Design and Validation of a Virtual Reality Mental Rotation Test

Kristin A. Bartlett[*]

University of Kentucky, kristibartlett@uky.edu

Almudena Palacios-Ibáñez

Universitat Politècnica de València, alpaib@doctor.upv.es

Jorge Dorribo Camba

Purdue University, jdorribo@purdue.edu

Mental rotation, a common measure of spatial ability, has traditionally been assessed through paper-based instruments like the Mental Rotation Test (MRT) or the Purdue Spatial Visualization Test: Rotations (PSVT:R). The fact that these instruments present 3D shapes in a 2D format devoid of natural cues like shading and perspective likely limits their ability to accurately assess the fundamental skill of mentally rotating 3D shapes. In this paper, we describe the Virtual Reality Mental Rotation Assessment (VRMRA), a virtual reality-based mental rotation assessment derived from the Revised PSVT:R and MRT. The VRMRA reimagines traditional mental rotation assessments in a room-scale virtual environment and uses hand-tracking and elements of gamification in attempts to create an intuitive, engaging experience for test-takers. To validate the instrument, we compared response patterns in the VRMRA with patterns observed on the MRT and Revised PSVT:R. For the PSVT:R-type questions, items requiring a rotation around two axes were significantly harder than items requiring rotations around a single axis in the VRMRA, which is not the case in the Revised PSVT:R. For the MRT-type questions in the VRMRA, a moderate negative correlation was found between the degree of rotation in the X direction and item difficulty. While the problem of occlusion was reduced, features of the shapes and distractors accounted for 50.6% of the variance in item difficulty. Results suggest that the VRMRA is likely a more accurate tool to assess mental rotation ability in comparison to traditional instruments which present the stimuli through 2D media. Our findings also point to potential problems with the fundamental designs of the Revised PSVT:R and MRT question formats.

CCS CONCEPTS • Human-centered computing • Human computer interaction (HCI) • Interaction paradigms • Virtual Reality

Additional Keywords and Phrases: Spatial Ability, Mental Rotation, Virtual Reality, Visual Perception, MRT, PSVT:R

[*] Corresponding author

1 INTRODUCTION

Spatial ability is one's ability to visualize and mentally manipulate relationships between objects and space. Researchers often use psychometric instruments to assess spatial ability [19], most of which are multiple choice tests that use black-and-white line drawings of shapes [15]. These psychometric tests are considered to assess "fundamental" spatial abilities [3]. Spatial ability assessment was once used to predict success in mechanical vocations [26], and some present-day researchers advocate for training spatial abilities in order to help students succeed in STEM [59].

Mental rotation tests are one of the most widely used psychometric spatial ability assessments. Two popular instruments include the Mental Rotation Test (MRT, [66]) and the Purdue Spatial Visualization Test: Rotations (PSVT:R, [20]). The MRT is commonly used in the fields of psychology and social studies, while the PSVT:R is more commonly used in STEM educational research [37]. Figure 1 shows an example question from the PSVT:R. The PSVT:R is presented in an analogous format. The test-taker must determine what rotation is applied between the first two shapes presented. The test-taker must then apply the same rotation to the shape in question and choose the correct answer from a bank of five choices. The PSVT:R contains 30 questions in total.

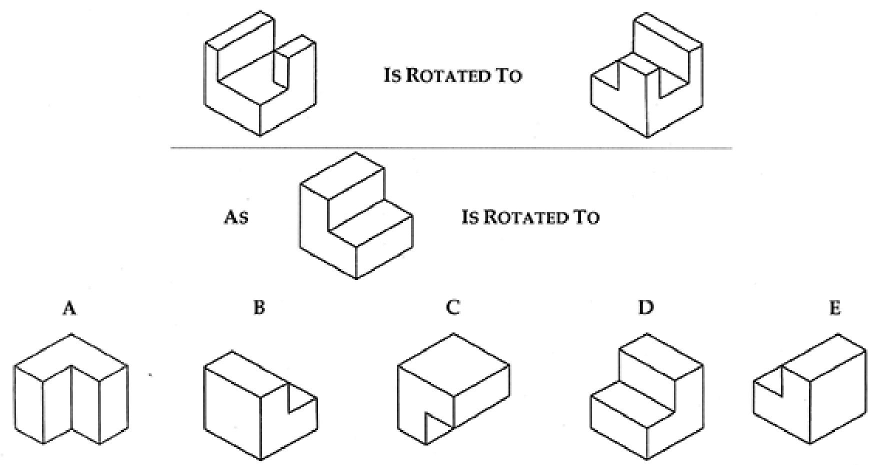

Figure 1. Example problem from the revised PSVT:R [70]*, correct answer is D.*

Figure 2 contains an example question from the MRT. In the MRT, the test-taker must determine which two shapes on the right are rotated views of the model shape on the left. The other two answer choices are distractors, which are either mirror images of the shape in question or entirely different shapes.

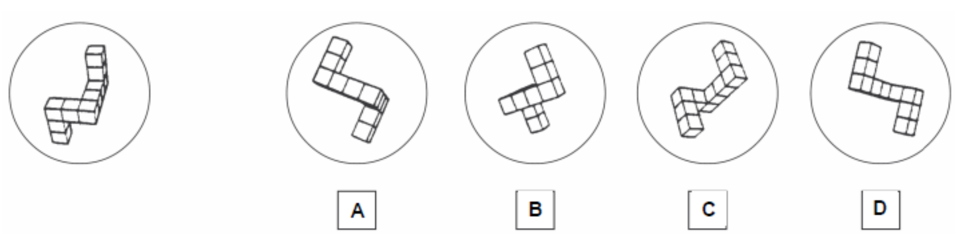

Figure 2. Example problem from the MRT [66]*, correct answers are B and D.*

The fact that these instruments have remained popular over the course of many decades does not in itself indicate that the MRT and PSVT:R are good measurements of people's mental rotation abilities. Many researchers have discussed various threats to validity in both of these popular instruments. One major issue is that item difficulty does not depend on the degree of the rotation needed to solve the item, which would be expected if a mental rotation process was the construct being tested by the instrument. We will discuss this principle in more detail in the literature review section. New instruments are needed in order to improve our understanding of fundamental spatial abilities and the accuracy of the assessments, and virtual reality (VR) technology is particularly well-suited [5,54].

While virtual and augmented reality has been widely used in spatial skills *training* applications, there are very few examples of VR-based spatial ability *tests*. However, VR presents an opportunity to address the fundamental problems with assessing 3-dimensional abilities in 2 dimensions. When it comes to interpreting a 3D shape from a 2D image, any 2D image is inherently ambiguous, because the viewer has no way of knowing the shape's depth or how the shape looks on the obscured side [51]. In this paper, we will explain why VR may be a more effective medium to deliver an accurate assessment of mental rotation ability in comparison to the 2D media-based assessments typically used. We present a prototype of a virtual-reality-based mental rotation test derived from the PSVT:R and MRT, the Virtual Reality Mental Rotation Assessment (VRMRA), and describe the results from the initial validation testing of the instrument. The purpose of our study is two-fold. We endeavored to create a new mental rotation assessment instrument which would one, bring new insights about the fundamental designs of the MRT and PSVT:R, and two, potentially serve as an more accurate mental rotation assessment for use by future researchers.

2 LITERATURE REVIEW

2.1 Visual problems with existing mental rotation tests

Mental rotation was originally introduced in a 1971 experiment by Shepard & Metzler [58], in which participants compared drawings of two figures rotated differently and judged whether the figures were the same shape or mirror images of the same shape. The figures were shown on an oscilloscope. An example item from their experiment is shown in Figure 3. In their experiment, which used a group of nine highly practiced participants, they demonstrated a linear relationship between the time it took participants to solve the problems and the degree of rotation between the two identical objects. Thus, based on this linear relationship, Shepard & Metzler concluded that a mental rotation process was being used to solve the problems. While this direct relationship between item difficulty and degree of rotation needed to solve the item was the foundation behind the concept of a "mental rotation" process, subsequent instruments that built on Shepard & Metzler's work did not uphold this same principle.

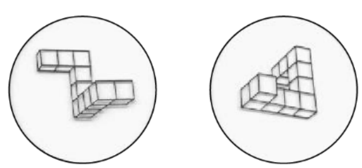

Figure 3. Item from Shepard & Metzler's experiment on mental rotation.

While Shepard & Metzler's work was based on demonstrating a direct relationship between item difficulty (measured as time to solve the items) and the degree of shape rotation, this same principle does not carry forward to the MRT and PSVT:R. Instead, in the MRT, figure perception, identification, and comparison account for the variance of item difficulty [9]. Other studies also demonstrated that individual differences in performance on the MRT are not likely to be related to the ability to rotated objects mentally and are related to other factors besides mental rotation [7,63]. Similarly, the item difficulty on the Revised PSVT:R is not related only to the degree and complexity of rotation of the shapes, and may be related to other factors like the features of the shapes [38,69]. The fact that item difficulty in these popular tests of "mental rotation" is not demonstrably related to mental rotation limits researchers' ability to accurately study the construct.

One reason for the instruments' shortcomings in mapping item difficulty to degree of rotation may be related to the figures used as stimuli in the instruments. It is reasonable to conclude that the imagery styles selected for mental rotation tests were driven by the technology available at the time. The tests use black and white line drawings of shapes shown in axonometric projections, a style of projection geometry that does not naturally replicate human vision and can be difficult to interpret. The MRT and PSVT:R were originally designed as paper-and pencil tests, reproducible through black and white photocopies. After a while, the existing copies of the MRT had degraded due to repeated photocopying, so a redrawn version was produced, this time with the help of a computer-assisted drawing program [50]. A revised, computer-drawn version of the PSVT:R was also produced because the original test contained drawing mistakes in many of the figures, such as missing features or extra features [70,71]. Thus, technology was used to improve the accuracy and clarity of the presentation of the stimuli as the tests were revised.

However, the tests continue to contain visual problems with the 2D presentation of the 3D stimuli. Difficulties include inherent ambiguity of isometric views, possibility of perceptual multistability, and occlusion [6]. These visual problems constitute construct irrelevant variance[6], which is a major threat to instrument validity [42]. Factors related to figure perception impact the difficulty level of problems on the MRT [9]. Seven of the 24 shapes in the MRT are occluded, meaning part of the object is hidden from view, which can affect people's ability to correctly interpret the shape [67]. Similarly, factors besides complexity of rotation impact item difficulty on the PSVT:R [69]. Comprehending the figures in the instrument may be a separate step [62], as they can be mistaken for 2D patterns rather than 3D shapes [4,8,71]. An experiment found that at least 30 answer choice shapes in the PSVT:R, when viewed outside of the context of the test, were not naturally seen as 3D shapes by over 50% of participants [4].

In everyday life, cues like shading, motion, and texture play an important role in contributing depth information to help us interpret 3D objects [52]. Rather than with black and white line drawings which lack these cues, mental rotation is likely best assessed with models which use perspective and shading cues [6,61]. Real world objects are not represented with black and white outlines - differences in color and value create "edges" that show us where one object ends and another begins. Multiple researchers have explored adding realism to stimuli in the PSVT:R and MRT through the use of more advanced computer rendering technology, and test performance improved as a result [6]. For example, computer rendering was used to depict the PSVT:R shapes in perspective views and with shading and texture, which led to higher scores for all students tested [72]. Analog technologies have been used as well. For example, Fisher et al. made versions of the MRT using photographs and physical blocks [17]. Scores improved on both versions, with the physical block version yielding the best performance [17]. The fact the photograph and block versions of the MRT were easier than the original version, when the degree of rotation required to solve the questions did not change, indicates that something besides the degree of rotation contributes to item difficulty in the MRT. This is likely an example of construct-irrelevant variance, which is a major threat to instrument validity [6,42]. Of these varying attempts to make test stimuli more realistic, any methods that continue to present the figures using 2D media (such as on paper or a still image on a computer screen) still fail to reduce the problem of occlusion, which may be one of the contributors to item difficulty in 2D test formats.

2.2 Applications of 3D media to spatial ability training and assessment

One way to address the problem of occlusion, or the fact that part of the shape remains hidden from view in a 2D image, is by using a virtual 3D image. Recognizing that VR is a useful technology for representing space, many researchers have used VR technology for spatial ability training applications [34]. Approaches have included training mental rotation skills by allowing participants to manually rotate the shapes in the virtual environment [1,2,10] or adding other visual cues in virtual environment to train rotation skills [13]. Other approaches have included allowing students to view their own 3D models in VR [18], training general spatial ability [28], or training spatial perspective taking ability [12]. Others have used AR to train spatial ability [31,39]. One study concluded that levels of immersion may be better suited to different levels of complexity of spatial learning [53]. However, these studies that used VR for mental rotation training continued to use the traditional MRT or PSVT:R to assess the efficacy of their training [1,43]. If the spatial tests used to measure the effectiveness of the intervention are not accurate, how can one truly know if the training intervention was effective?

Researchers who applied 3D visualization technologies to spatial assessments have frequently seen improved performance on the assessments [5], again suggesting the possibility of construct-irrelevant variance. One such modification is to show video animations or renderings of the shapes in motion. Showing participants video animation of rotations of the shapes in the Santa Barbara Solids Test led to improved performance compared to the original test which showed static 2D figures of the 3D shapes [56]. Performance on the Mental Cutting Test (MCT) also improved when participants could rotate the shapes 180 degrees in either direction about Y-axis [65]. A gamified version of a mental cutting test was created using augmented reality in an Android application [64]. Other researchers used VR combined with some tangible interfaces to assess perspective-taking ability [11,12].

Assessment of mental rotation ability in VR remains an underexplored area, with few implementations [1]. While multiple research groups have proposed designing VR-based spatial tests (ex. [24,30]), we were only able to find one group who built and validated such an instrument during the past decade [21,22]. (We note that another group used VR to assess a large-scale spatial ability of spatial orientation [49], which is typically considered to be in a different category of spatial ability than mental rotation.)

Guzsvinecz et al. used VR to create a spatial test with components following the style of the MRT, MCT, and PSVT:R [21,22]. The test was delivered using Gear VR which runs on some Samsung smartphones and has a touchpad for interaction. Participants could rotate their heads and see objects from a slightly different direction but could not adjust their viewpoint because the Gear VR could only account for rotations. Participant performance was compared with the same test delivered in a non-VR desktop version. The researchers found that the probability of getting answers correct was significantly higher in VR version than in desktop version of their test [23]. The perspective camera was also found to be advantageous in the VR spatial ability assessment in comparison with an isometric camera [23].

Some earlier studies of VR-based mental rotation tasks focused on the study of gender differences in mental rotation. Men tend to perform better than women on the traditional version of the MRT [35,68]. Larson et al. created a mental rotation task in a virtual environment. While they found gender differences on the paper version of the MRT, women and men performed equally in their virtual mental rotation task [33]. They suggested that gender differences may be related to the need to derive 3D representations from 2D drawings in the paper-based MRT [33]. Larson and colleagues' work was done in 1999, and their virtual environment presented the stimuli as "hologram-like" 3D objects floating above a projection screen. The participants rotated a physical controller to solve the questions. The researchers measured the amount of time to complete the rotation and the efficiency with which it was completed to compare performance between participants [33]. Parsons et al. also found no gender differences using the same system as Larson et al., the "Virtual Reality Spatial Rotation (VRSR)" system [48]. Neubauer et al. also did not find gender differences in their own virtual-environment-based mental rotation test [45]. Neubauer et al.'s test presented objects in 3D in a similar manner to what would be seen in a 3D cinema, but did not allow participants to move around in space, so it was not a "full-blown" presentation of virtual reality [45].

2.3 Opportunities and limitations of VR for mental rotation assessment

In this paper, we argue that none of these previous VR-based spatial assessments have leveraged the full range of possible improvements to spatial tests that VR technology offers. We suspect that some of these decisions may have been driven by hardware. Guzsvinecz et al. used the Gear VR, which was released in 2015 and ran on a Samsung Galaxy S6 Edge+ smartphone which is also from 2015. Due to hardware limitations, in their spatial test the camera could be rotated but could not move positionally [21]. Rather than presenting the shapes in a natural-looking context, Guzsvinecz and colleagues presented the shapes on a solid background. Instead of using life-like cues to represent the appearance of the shapes, the shapes still have black outlines on their edges [21]. The previous studies [33,45,48] used even older technologies and did not use a VR headset. A newer, non-smartphone-based VR headset device that has more rendering power would allow for a more realistic and immersive VR experience.

Multiple studies have shown that using physical blocks to re-create the MRT led to improved performance on the test [16,17,41,55]. The degree of "mental rotation" required to solve the items remained unchanged, as the physical block versions of the test recreated the same questions as the original paper test, yet the test became easier when the stimuli was presented differently, again reinforcing that factors besides mental rotation impact item difficulty in the original MRT. We suspect that this improvement arose from participants' ability to take in more realistic, comprehensive visual information about the shapes, including being able to look at the shapes from multiple viewpoints instead of a fixed viewpoint. A VR application could also be designed to present shapes in a more realistic manner and in a more lifelike environment, improving the immersive VR experience and allowing for more natural perception of shape. A room scale VR experience could allow participants to view shapes from different sides by moving the position of their body, which would eliminate the problem of occlusion. A VR test could also render the shapes in a natural looking manner by using shading and shadows instead of outlines and placing the shapes in a natural-looking environment.

Virtual reality also offers a possibility of better engagement of research participants through novelty, intuitive interactions, and gamification. Compared to a screen-based test, looking at shapes in VR where the scene adjusts based on participants' body movements would allow for a more natural interface compared to a computer mouse controlling the rotation of a shape's animation. Features like hand tracking, which is available on newer devices like the Meta Quest 2, could also be leveraged for a more immersive experience. For example, some researchers have used the hand tracking feature along with passive haptics for product design evaluation [47], as this approach enhances the virtual experience, potentially resulting in a more accurate and similar evaluation of the virtual prototype as compared to the real product. In another study, a hand-adaptive UI was found to outperform an eye-centered UI when used in a VR environment. In the hand-adaptive UI, the UI followed the position of the hand rather than the position of the eye [36]. Free hand interactions can be more intuitive and result in better usability and immersion in VR applications [36].

In terms of limitations, cybersickness is one possible problem when using VR to assess spatial skills. A meta-analysis concluded that there are unequal rates of VR sickness across different populations based on gender, real-world experience, technological experience, neurological disorder, or relevant phobias [29]. Sudden changes in direction and velocity of camera made people feel ill compared to smooth motion. Some researchers have concluded that level of cybersickness was related to type of VR environment [40], and that people believe they will feel less sick if they are in control of the motion [32]. Any application designed to test spatial skills in VR should be designed in such a way as to minimize cybersickness.

Another limitation of using VR to assess spatial skills is the issue of clarity based on object distance and interpupillary distance (IPD). People naturally use different amounts of ocular convergence to look at near or far objects, but virtual reality systems must produce an image that's rendered in essentially a "compromise" between near and far distance, though this could be resolved in the future through a design for a VR headset that automatically adjusted the image depending on where the user was looking [44]. Many VR headsets do have an adjustable IPD, which physically adjusts the distance between the lenses to accommodate the differences in people's IPDs. However, some have pointed out that many headset designs are designed to fit a greater percentage of males than females due to the fact that females tend to have smaller IPDs and the headsets are tailored more for larger IPDs. For example, the HTC Vive Pro would not be expected to fit the smallest 40% of females, and would not fit the smallest 18% of males and largest 1% of males [60]. Thus, the selection of a headset with an adjustable IPD and a range that fits as many people as possible is critical for accurately assessing spatial skills.

While the availability of VR hardware remains a limitation to widespread spatial skills assessment using VR, VR may be more accessible and appropriate than test setups using physical blocks. A VR application could offer better test repeatability because physical blocks may not be positioned in the exact same place every time. The VR application may also be less burdensome for researchers as it would be "plug-and-play" rather than requiring shapes to be rearranged in front of a participant for each test question. Furthermore, some newer headsets such as the Meta Quest 2 have become relatively inexpensive for a research instrument.

3 METHODS

3.1 Motivation and Initial Pilot Testing

Before deciding on VR as the ideal medium in which to deliver a spatial ability assessment, we pilot tested a computer-rendered 2D version of a very difficult question on the PSVT:R to see if this led to improvement in performance. Problem 30 is the most difficult problem on the PSVT:R, with around 33% of students answering correctly [38]. We used computer rendering to make a more realistic version of this question, where figures are presented in perspective views with color and shading. The original version is shown in Figure 4 and the revised version in Figure 5.

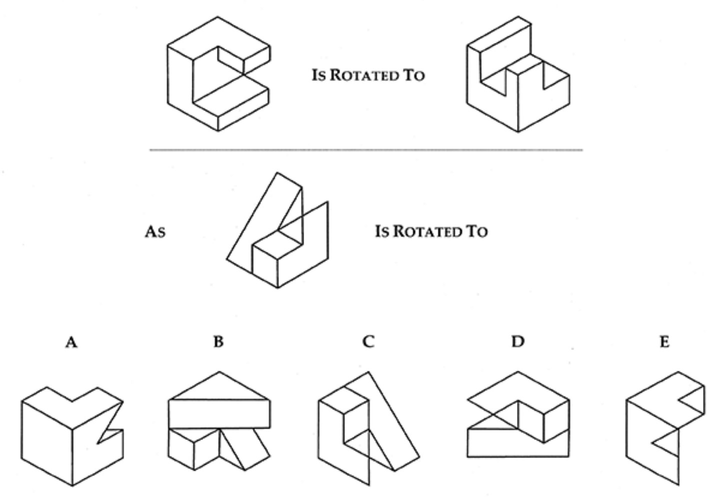

Figure 4. Question 30 from the revised PSVT:R [70]. *Correct answer is E.*

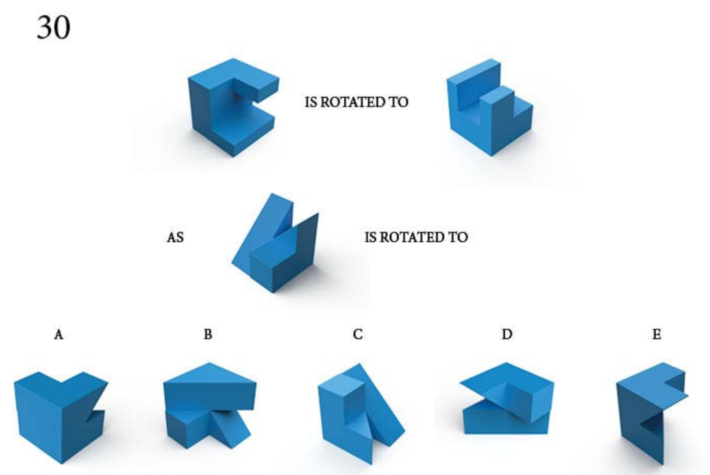

Figure 5. A revised version of question 30 from the PSVT:R, using perspective views, color, and shading to attempt to make the figure more realistic. The correct answer is E.

The computer rendered shapes should reduce many points of ambiguity in the isometric black and white line drawings in the original version of the PSVT:R. Our brains naturally interpret gradual changes of hue, saturation, and brightness and changes in illumination, and changes of hue, saturation, and brightness as changes in surfaces [27]. The computer-rendered figures allow us to use these natural cues to understand what shapes we are looking at. Answer choice E in Figure 4 is one of the shapes that was least likely to be interpreted as a 3D shape in the PSVT:R [4], and this is the correct answer, which may be the reason that so few people get the answer correct. This shape appears clearer in Figure 5 than in Figure 4. However, in the small pilot test with computer graphics students, the revised computer-rendered version of problem 30 did not appear to make the question easier, as most students continued to answer the question incorrectly. Some students reported that they still could not picture sides of the shapes that were hidden from view. Because of this, we decided to explore using VR to address the problem of occlusion in mental rotation instruments. Ultimately, our goal in designing this application was to investigate whether a VR-based test could improve upon the MRT and the PSVT:R by creating similar exercises in which item difficulty is more directly related to the degree of rotation needed to solve the questions, as this would imply that the new test is a better measure of mental rotation skills.

3.2 VRMRA Application Design

The VRMRA is a VR-based spatial assessment designed to run on the Meta Quest 2. We selected the Meta Quest 2 due to its inexpensive price point ($300 USD) which should make it fairly accessible to researchers in comparison to other VR headsets, and due to the availability of interaction features like hand tracking. The application was created in Unity version 2020.3.11f1 using C# scripts and Oculus Integration version 0.37.0. The virtual environment used the standard shader for materials and unidirectional real time light. Three-dimensional models were created in SolidWorks 2021 and reformatted in Blender version 3.0.1 before importing them into Unity. For our application, we selected 12 questions from the MRT and 12 questions from the PSVT:R. We recreated these questions in VR using solid models of the same shapes and placed the shapes in the same positions they are depicted in the original tests. A version of a question from the PSVT:R is shown in Figure 6, and from the MRT is shown in Figure 7.

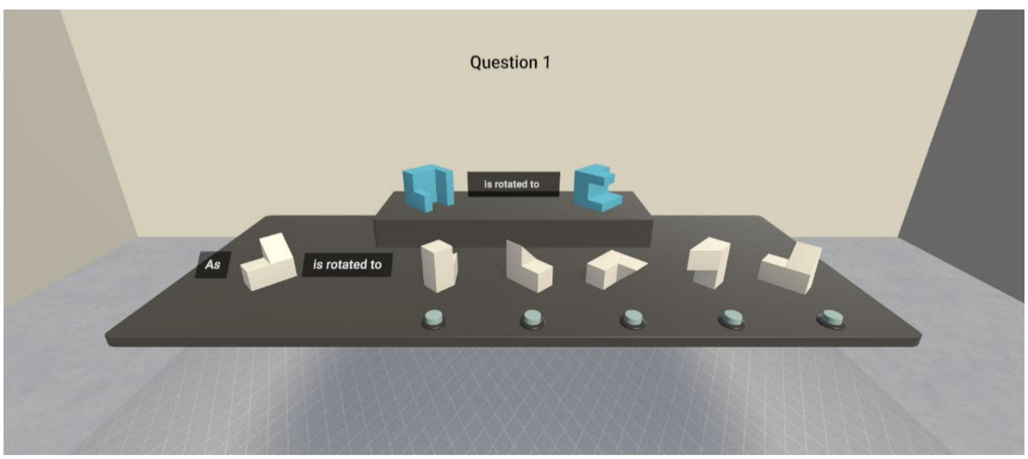

Figure 6. A PSVT:R question shown in our VR based test. The correct answer is the first answer option.

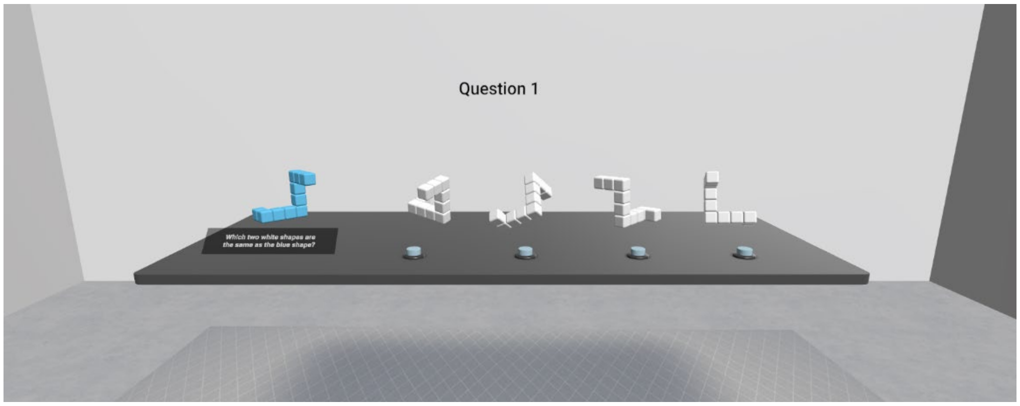

Figure 7. An MRT question shown in our VR based test. The two correct answers are the first and third options.

We followed the intention of the physical block versions of the MRT by placing the shapes atop a virtual table as if they were real objects in a real room. We used shading and shadow to render the shapes in a natural-looking manner. (The shadows are not visible in the zoomed-out views shown in Figures 6 and 7; see Figure 8 for an example of the shadows.) We put blank walls in the room to minimize distraction during the test. We placed buttons on the table in front of their corresponding shapes, so if the test-taker wanted to select a shape as the correct answer, they just needed to press the button in front of that shape. We used virtual buttons from the Oculus Interaction SDK which could be pressed with the user's real hands through hand tracking. The hand tracking feature of the Meta Quest 2 shows virtual hands in the position of the viewer's real hands. Figure 8 shows a virtual hand about to press a button to select an answer.

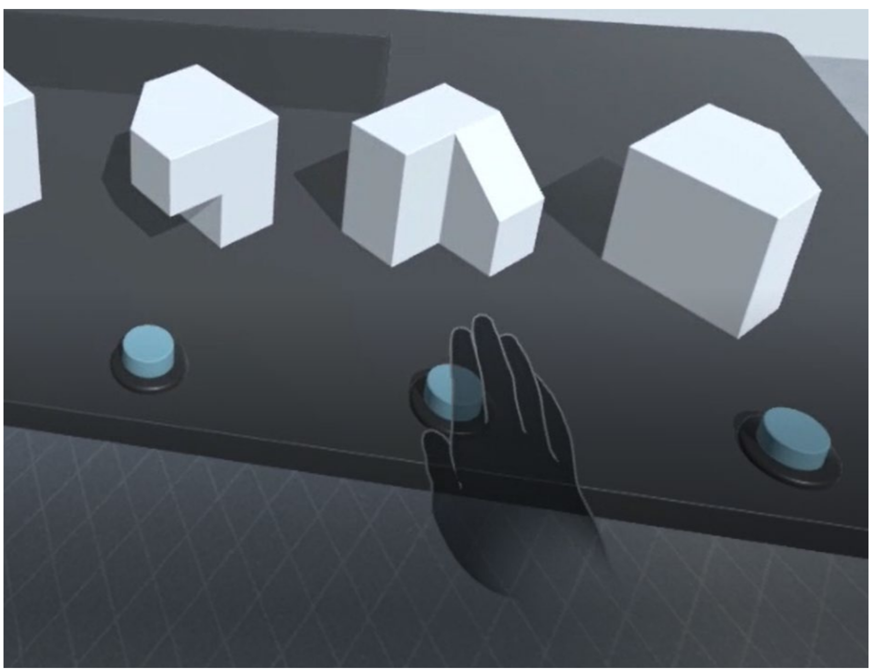

Figure 8. A virtual hand selecting an answer in the test.

We added some elements of gamification to the test through feedback provided to the user after they submit their answers. After the user selects an answer, a positive or negative sounding audio plays to indicate whether the answer was correct or not. Then, the correct answer shape turns green. An animation then plays that shows the answer shapes rotating into the correct position. This feedback demonstrates to the test-taker why the correct answer is correct, so that they can potentially learn from and understand any mistakes, in case they misunderstood the question format and directions at first. Figure 9 illustrates a frame of the animation. The black button shows the answer that the user had submitted. The green shape on the far right was the correct answer. The colored shapes in Figure 9 have all completed their rotations demonstrating why the shape on the far right was the correct answer.

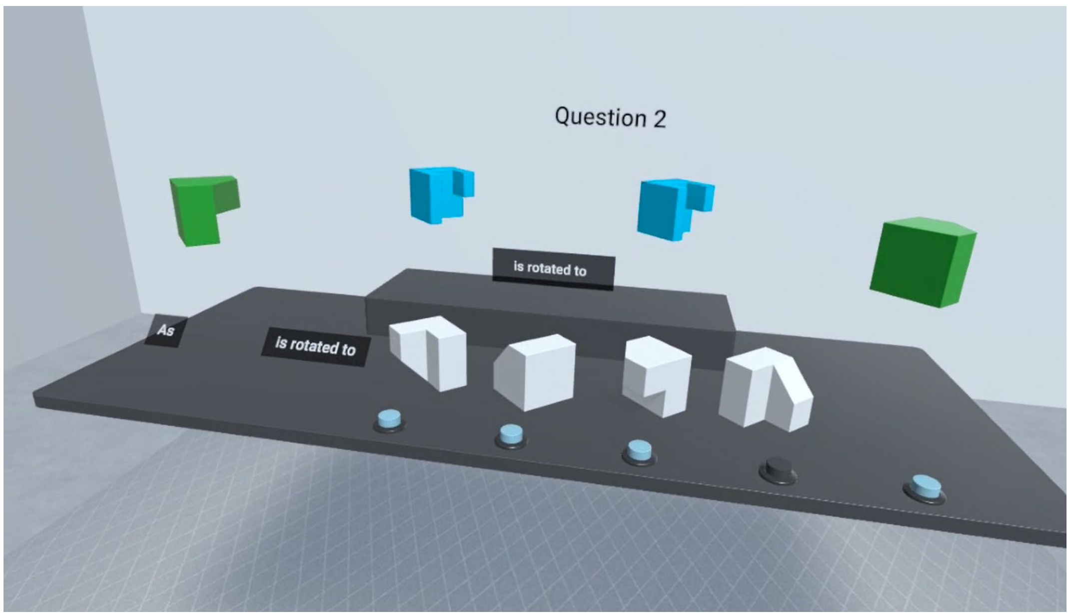

Figure 9. Completed animation demonstrating the correct answer.

Since the VRMRA application uses room-scale VR, the participants can walk back and forth to view different shapes on the virtual table up close or from different angles. The application does not allow the test-taker to pick up or interact with the virtual shapes, the shapes are static and stay in one position except for the automated animations following answer submissions. This way, the test-taker cannot manually rotate any shapes. All rotations must be performed mentally, in keeping with the spirit of the original tests.

The VRMRA application includes a short tutorial at the beginning to demonstrate the pressing of the buttons and the sounds for correct and incorrect answers. Following this tutorial, the application advances to an example question for part 1, the PSVT:R-style questions. This example question shows the correct answer with a green flashing button. After the test-taker presses this button, they will advance through a series of 12 PSVT:R-style questions of the same format. Then, a second example question is shown for part 2, the MRT-style questions. Since this section has two correct answers, the corresponding two buttons flash green. Once the test-taker presses these buttons, the application advances through the 12 questions in part 2. There is no time limit to answer the questions, but the time spent on each question is recorded, as are the answers submitted by the test-taker. This information is saved to a CSV file for analysis.

All questions in part 1 of the VRMRA were copied directly from the Revised PSVT:R with the exception of questions 1, 9, and 11 which were nearly the same question but with a different correct answer. The correct answer was modified by changing the direction of the example rotation. All answer choices were presented in the same order. Table 1 shows which questions in the VRMRA correspond to which questions in the Revised PSVT:R.

Table 1. *Questions in part 1 of the VRMRA and their corresponding question numbers in the revised PSVT:R.*

Question in VRMRA Part 1	**Corresponding question in Revised PSVT:R**
1	1 (modified to have answer "A")
2	12
3	17
4	13
5	26
6	24
7	27
8	23
9	25 (modified to have answer "E")
10	22
11	29 (modified to have answer "A")
12	30

All questions in part 2 of the VRMRA were copied directly from questions in the MRT, with the exception of question 5, which has the correct answers in positions B and D in the original MRT. Table 2 shows the questions in part 2 of the VRMRA and their corresponding question numbers in the MRT.

Table 2. *Questions in part 2 of the VRMRA and their corresponding question numbers in the MRT*

Question in VRMRA Part 2	**Corresponding Question in MRT**	**Correct Answers**
1	1	A,C
2	3	B,D
3	6	A,D
4	8	B,C
5	9	A,D (B and D in MRT)
6	10	A,D
7	11	B,D
8	12	B,D
9	13	B,D
10	15	B,D
11	16	B,C
12	18	A,D

3.3 Validation

A group of 68 graduate and undergraduate students with varying academic majors participated in our study. The majority of participants were from technology-related majors, and 65% had taken past courses in 3D graphics such as CAD or 3D computer modeling. Participants were not asked to report their ages, but we estimate that the majority of participants were in the age range of 18 - 34 years, with a few who were older. Participants were asked about their past experience using VR. Thirty-two percent had never used a VR headset before, 47% had used one once or twice before, and the remaining 21% used VR somewhat regularly. Gender and racial demographics of the participants are shown in Tables 3 and 4.

Table 3. *Gender demographics of participants*

	Number	**Percent**
Female	28	41.2
Male	38	55.9
Nonbinary/gender-fluid/gender nonconforming	2	2.9

Table 4. *Racial demographics of participants*

	Number	**Percent**
Black / African American	10	14.7
East Asian / Southeast Asian	11	12.2
Hispanic / Latino	5	7.4
South Asian / Indian	5	7.4
White	34	50.0
More than one race	2	2.9
Prefer not to say	1	1.5

Participants were asked to complete the entire VRMRA. Sessions were conducted with one participant at a time. Completion times and scores for each participant were recorded for subsequent comparative analysis with paper-based assessments. Prior to testing, participants were shown a video with the example questions for each part of the test. Alongside this video, the experimenter gave a verbal explanation of how to answer the questions. The video also showed how to press the virtual buttons. Participants were then assisted in adjusting the Meta Quest 2 headset to a comfortable position, including the clearest IPD setting of the three options in the headset. Participants were instructed to answer the questions as quickly as possible without making mistakes. Participants instructed that they were allowed to walk anywhere along the front and sides of the virtual table in order to look at the shapes from other angles. Participants were not allowed to walk through the virtual table because this could lead to an unintentional press of a button with the hand tracking. Participants were also not allowed to walk behind the table due to space limitations in the testing room. Following the testing with the VRMRA, participants answered a short survey with demographic questions.

4 RESULTS

4.1 General observations

Participants reported enjoying the experience of using the application. Many made unprompted comments saying that they wanted to use VR more often after completing the VRMRA. A few participants expressed frustration about not being allowed to walk behind the virtual table to view the shapes from the back, as they thought this would aid them in solving the questions. No participants reported any cybersickness. About five participants experienced minor difficulties related to the hand-tracking, most of which were resolved during the initial tutorial section. One difficulty was related to perceiving the depth of the button position, as two participants initially were not reaching far enough to activate the buttons. A second issue was related to unpressing the buttons. Because there was no physical button or physical table, test-takers could press "through" the button. When this happened, the button would be pressed but not unpressed, so the application would not advance. After observing this issue in early testing, we added more information about button pressing to the tutorial and reviewed the instructions for pressing the button in the initial video prior to testing, but some test-takers forgot this procedure and occasionally left their hands down at their sides after pressing the button, causing the application to not advance to the next question. This issue was quickly resolved with verbal instructions from the experimenter who was supervising the testing.

The two parts of the VRMRA application (PSVT:R questions and MRT questions) were analyzed separately to aid analysis. The results are reported next.

4.2 Part 1 (PSVT:R style) VRMRA Results

The average score on Part 1 of the VRMRA was 8.10 out of a possible 12 points (SD = 2.47, N = 68). The correct response rates are shown in descending order in Table 5. Question difficulty varied greatly, with 86.8% of respondents answering question 4 correctly and only 38.2% of respondents answering question 3 correctly. Table 5 also shows the degree of rotation required to solve the question and the correct response rate of the corresponding question in another study which used the Revised PSVT:R [38]. Questions could be solved in either one single rotation around a cardinal axis or two rotations around two different cardinal axes. Since the rotations in the PSVT:R style questions are at fixed increments around one or two of the cardinal axes, a simple correlation analysis between question difficulty and degree or rotation did not seem appropriate. Instead of using the degrees or rotation for our analysis, we first assigned each rotation type a "degree of rotation rank" on a scale of 1 (least complex) to 4 (most complex). A 90° rotation around a single axis was ranked 1, a 180° rotation around a single axis was ranked 2, a 90° rotation around one axis plus a 90° rotation around a second axis was ranked 3, a 90° rotation around one axis plus a 180° rotation around a second axis was ranked 4. For a second analysis, we also created a binary classification of the complexity of rotation, where the problems which involved a rotation around a single axis were considered simple rotations and the problems which involved rotations around two different axes were considered complex rotations.

Table 5. *Part1 VRMRA results compared to patterns seen in another study using the Revised PSVT:R*

Question in VRMRA Part 1	**% correct in VRMRA**	**% correct in Maeda et al.** [38]	**Rotation around first axis**	**Rotation around second axis**	**Degree of rotation rank (4 = most complex, 1 = least complex)**	**Complexity of Rotation**
4	86.8%	65.4%	180°	none	2	Simple
11	83.8%	55.4%	90°	none	1	Simple
2	75.0%	70.6%	180°	none	2	Simple
1	75.0%	90.8%	90°	none	1	Simple
8	73.5%	70.6%	90°	180°	4	Complex
9	72.1%	67.0%	90°	none	1	Simple
6	69.1%	77.2%	90°	180°	4	Complex
10	64.7%	45.6%	90°	90°	3	Complex
7	60.3%	63.4%	90°	180°	4	Complex

12	60.3%	32.7%	90°	180°	4	Complex
5	51.5%	64.6%	90°	180°	4	Complex
3	38.2%	68.9%	90°	90°	3	Complex

In comparison to the data from the Revised PSVT:R (percent correct in Maeda et al. [38]), some questions became easier when presented in the VRMRA, while others became more difficult. We performed a correlation analysis between the correct response rates and the degree of rotation ranks to investigate the association between the degree of rotation and the item difficulty. On the VRMRA, degree of rotation rank and correct response rate were moderately negatively correlated, though results were not statistically significant (r = -.55, p = .063). On the Revised PSVT:R, degree of rotation rank and the correct response rate had a small negative correlation, and results also were not statistically significant (r = -.29, p = .36).

We performed a t-test to see if there was a significant difference in item difficulty between the problems with rotations around a single axis ("simple" rotations) and the problems with rotations around two different axes ("complex" rotations). We coded the simple rotations, a "1" and the rotations around two axes, complex rotations, as "2," and did an independent samples t-test to compare the response rates of questions between the two groups. In the VRMRA, the independent-samples t-test showed a significant difference between correct response rates in the simple rotations versus complex rotations. Problems with simple rotations (M = 75.45% correct, SD = 6.35%) were significantly easier than the problems with complex rotations (M = 59.66% correct, SD = 11.79%, t = 3.231; df = 10, p = .005). We performed the same analysis on the Revised PSVT:R data reported by Maeda et al. [38]. In the Revised PSVT:R data, problems with simple rotations (M = 69.84%, SD = 13.0%) were not significantly easier than problems with complex rotations (M = 60.43%, SD = 15.66%, t = 1.097; df = 10, p = .149). Thus, the pattern of difficulty demonstrated in the VRMRA was more closely related to the complexity of rotation required to solve the problems in comparison with the Revised PSVT:R. While the rotational complexity as measured by number of rotational axes (one or two) appeared to affect item difficulty in the VRMRA, degree of rotation did not appear to matter, as questions requiring a single 90-degree rotation weren't necessarily easier than questions requiring a single 180-degree rotation, since our correlation analysis with the degree of rotation rank did not yield significant results.

4.3 Part 2 (MRT style) VRMRA Results

The average score on Part 2 of the VRMRA was 10.35 out of a possible 12 points (86.3%) using the scoring scheme of awarding a point only when both correct answers were identified in the question (SD = 1.62, N = 68). The average score was 22.25 out of a possible 24 points (92.7%) using the scoring scheme of awarding a point for each correctly identified answer (SD = 1.82, N = 68). The percentage of correct responses for each question in part 2 of the VRMRA is shown in Table 6, along with the information about whether the question contained mirror image distractors and a heterogeneous or homogeneous configuration in the shapes. Caissie et al. [9] called MRT figures with two blocks at one end segment and three blocks at the other end segment heterogeneous, and the figures with three blocks at both end segments homogeneous. They predicted that heterogeneous configurations were likely to help with figure identification and mental rotation and make questions easier. In the experiment by Caissie and colleagues, the MRT was given with a time limit, and many participants did not complete all the questions. Their subsample 1b included the participants who completed all or all but one of the questions. Thus, their subsample is likely to have an overrepresentation of high performers. Correct response rates on the VRMRA were all within 10 percentage points of Caissie and colleagues' sample, but the order of difficulty of questions was not the same.

Table 6. *Part2 VRMRA results compared to patterns seen in another study using the MRT*

Problem in the VRMRA	Problem in MRT	Answer	Mirror image shapes?	Configuration
1	1	A,*C*	y	hetero
2	3	B,*D*	n	hetero
3	6	A,*D*	y	hetero
4	8	B,*C*	n	hetero
5	9	A,*D*	y	homogeneous
6	10	A,*D*	y	homogeneous
7	11	B,*D*	n	homogeneous
8	12	B,*D*	n	homogeneous
9	13	B,*D*	y	homogeneous
10	15	B,*D*	n	homogeneous
11	16	B,*C*	n	homogeneous
12	18	A,*D*	y	homogeneous

Caissie and colleagues performed an analysis to look for an association between angle of rotation between the model shape and the answer shape and question difficulty. They found no significant association between angular disparity and difficulty; however, they used the X, Y, and Z rotation angles individually, rather than the calculated most efficient rotation around a skewed axis. On the VRMRA, degree of rotation in the X direction and correct response rate were moderately negatively correlated (r = -.526, p = .008). Significant correlations were not found between the degrees of rotation in the Y or Z directions and the correct response rate; (r = -.158, p = .462) for Y and (r = -.073, p = .735) for Z. (In the VRMRA, the X axis is horizontal, the Y axis is vertical, and the Z axis is coming toward the viewer). We wondered if using the skewed axis rotation would lead to a more significant correlation between angular disparity and difficulty, as solving skewed axis rotations was found to involve the same skill as solving rotations around cardinal axes [46]. The skewed axis rotation represents the most direct path to rotate the shape from one position to another. However, a significant correlation was also not found between the degree of rotation around the skewed axis and the correct response rate (r = -.054, p = .801). Answer choices, rotation angles, and correct response rates for part 2 of the VRMRA are shown in Table 7.

Table 7. *Degree of rotation and correct response rates for answer choices in part 2 of the VRMRA*

Answer choice in VRMRA	X-axis rotation	Y-axis rotation	Z-axis rotation	Skewed-axis rotation	Percent correct in VRMRA
2B	6°	159°	10°	160°	100.0%
4B	0°	178°	0°	178°	100.0%
2D	8°	80°	2°	80°	98.5%
4C	0°	99°	0°	99°	98.5%
8D	21°	106°	164°	173°	98.5%
1A	11°	80°	3°	81°	97.1%

7D	3°	63°	5°	63°	97.1%
3A	10°	147°	20°	149°	95.6%
12A	18°	8°	123°	125°	95.6%
3D	1°	79°	21°	81°	94.1%
7B	8°	174°	3°	174°	94.1%
11C	38°	56°	46°	90°	94.1%
11B	14°	119°	92°	148°	92.7%
12D	14°	24°	97°	102°	92.7%
6D	22°	51°	106°	124°	91.2%
5B	104°	135°	27°	173°	89.7%
6A	4°	72°	166°	171°	88.2%
8B	48°	79°	77°	134°	88.24%
9B	22°	138°	38°	148°	88.24%
10D	9°	37°	31°	51°	88.24%
10B	6°	126°	62°	137°	86.76%
9D	29°	86°	10°	93°	85.29%
1C	6°	180°	4°	180°	80.88%
5D	110°	163°	37°	160°	76.47%

A paired-samples t-test showed a significant difference between accuracy on heterogeneous and homogeneous items in the VRMRA (t = 3.066; df = 67, p = .003). Performance on heterogeneous items (M = .92, SD = .15) was better than on homogeneous items (M = .84, SD = .18). A paired-samples t-test showed a significant difference between accuracy on items with mirrored distractors and items with non-mirrored distractors on the VRMRA (t = 2.695; df = 67, p = .009). Performance on items with non-mirrored distractors (M = .90, SD = .14) was better than on items with mirrored distractors (M = .82, SD = .22).

Multiple regression analysis was used to test if mirrored/non-mirrored distractors and configuration type significantly predicted correct response rates on the individual answer choice options in part two of the VRMRA. Items with mirrored distractors were coded as 1 and items with structural, non-mirrored distractors were coded as 0. Heterogeneous configurations were coded as 1 and homogeneous configurations were coded as 0. The results of the regression indicated the two predictors explained 35.5% of the variance in correct response rates (R^2 = .355, $F(2,21)$ = 5.777, p = .01). Mirrored distractors significantly predicted item difficulty (β = -.051, p = .022), as did heterogeneous configurations (β = .051, p = .030). In a linear model created by Caissie and colleagues, occlusion and configuration type (heterogeneous or homogeneous) explained 36% of the variance in item difficulty of the individual answer choice figures, and other predictors such as distractor type (mirrored or structural) and angle of rotation did not improve the fit of the model [9]. Similarly, in the case of the VRMRA, adding degree of rotation as a predictor did not improve the fit of the model.

4.4 Full Test Results of VRMRA

We performed a correlation analysis between the PSVT:R style questions and the MRT style questions on the VRMRA. The PSVT:R section scores and MRT section scores (out of 12 points) were found to be moderately positively correlated,

($r(67) = .52$, $p < .001$). This is about the same as what other researchers have found in correlation analysis of the MRT and Revised PSVT:R tests administered on a computer [57].

The Cronbach's alpha coefficient of internal consistency for the full VRMRA was found to be 0.734, which indicates acceptable score reliability when used with a population like the one in our study. The magnitude of reliability depends on the sample of individuals who took the assessment [38] but it also depends on the number of questions. We are not aware of any other studies which calculated an internal consistency coefficient for a combined assessment that included questions derived from both the PSVT:R and MRT, so we do not have a direct point of a comparison for our test. The full Revised PSVT:R on its own was found to have a Cronbach's of 0.839 [38].

5 DISCUSSION

Part 1 of the VRMRA was derived from the Revised PSVT:R. Though the Revised PSVT:R is considered to be a mental rotation test, item difficulty in the Revised PSVT:R does not correspond to the complexity of the rotation needed to solve the problems [38]. Others have suggested that that this fact could be due to the complexity of the shapes affecting item difficulty [38]. However, the VRMRA uses the same shapes as the Revised PSVT:R, and in the VRMRA, item difficulty corresponded more closely to the complexity of rotation. On the VRMRA, questions which required rotations around two different axes were significantly more difficult than the questions which required a rotation around a single axis. Although the number of axes appeared to be important in the VRMRA, the degree of rotation was not found to be significantly correlated to the item difficulty in either the Revised PSVT:R or the VRMRA. Our results suggest that by leveraging VR for a more realistic 3D presentation of the shapes, the VRMRA may be a more accurate test of the skill of mental rotation than the Revised PSVT:R, though neither test shows a perfect relationship between item difficulty and degree of rotation required to solve the items. The reasons for the lack of clear relationship between item difficulty and degree of rotation in both the VRMRA and the Revised PSVT:R might be due to the use of an analogous format in the PSVT:R and the use of rotations around cardinal axes. It could be that the design of the PSVT:R question format is not one which is well-suited to demonstrably measure mental rotation skills without other factors also contributing to question difficulty.

Compared to the Revised PSVT:R, some questions in part 1 of the VRMRA became relatively easier, while others became relatively harder. One element of the VRMRA that might have made some PSVT:R-style questions more difficult is the fact that each shape is no longer seen from a fixed viewpoint as in the original instrument. In the Revised PSVT:R, each shape is seen from an isometric viewpoint, with an axis facing the viewer. In our instrument, all shapes are in a fixed position and are seen from the viewpoint of wherever the test-taker is standing, and while the axes of the shapes are all aligned to one another, the viewer would have to move directly in front on each shape individually to see the shapes from the same viewpoints that they are presented in the Revised PSVT:R. This scheme could make it slightly harder to judge the relative rotations since more information may need to be retained in working memory.

Part 2 of the VRMRA was based on the MRT. Our findings indicated that the VRMRA and MRT have some similarities in the qualities that predict item difficulty, but also some differences. In the MRT, degree of rotation in the X, Y, and Z direction was not found to correlate with item difficulty [9]. In the VRMRA, degree of rotation in the X direction was found to be moderately negatively correlated with item difficulty correlated ($r = -.526$, $p = .008$), but significant correlations were not found with degree of rotation around the Y or Z direction or around the skewed axis. The fact that degree of rotation does not predict item difficulty in the MRT has been considered a very problematic aspect of the test since the theory of mental rotation originally devised by Shepard & Metzler relies on a linear relationship between the time it took to solve a question and the degree of rotation needed to solve it [9]. We do not have a definite explanation for why only the rotation in the X direction and not the rotations in the Y or Z direction, or especially the most direct path rotation around the skewed axis, were predictive of item difficulty. One possibility could be the fact that the X rotations were least visible to the participants. Since the Z axis comes toward the viewer, any rotation around the Z axis would be very easy for the viewer to notice since no parts of the shape would become hidden from view with that rotation. Since the Y axis is vertical, a participant could view the rotated parts of the shape by walking to either side. With a rotation around the X axis, a feature that was initially facing the viewer could move to the bottom or back side of the shape, making it become more hidden. This theory could be tested in future experiments by rotating identical shapes by identical amounts around different axes in different questions.

Part 2 of the VRMRA showed different contributors to item difficulty in comparison to the MRT. Multiple researchers have found the occluded items were more difficult than non-occluded items in the MRT [14,67]. Occlusion was not considered a factor in the VRMRA since participants could view each 3D shape from multiple angles. On the VRMRA, performance was significantly better on items with heterogeneous structures versus homogeneous structures and was significantly better on items with structurally different distractors versus mirrored distractors. A linear model with shape configuration (heterogeneous vs. homogeneous) and mirrored versus structural distractors was found to account for 50.6% of the variance in question difficulty in the VRMRA. In contrast, 53% of the variance in question difficulty was predicted by occlusion and shape configuration in the MRT [9]. Others have also found that items with mirror-image distractors are more difficult than those with structural distractors on the MRT [25]. Thus, the VRMRA is similar to the MRT in that mirror image distractors and homogeneous configurations contribute to item difficulty. However, importantly, the presentation in VR removed the factor of occlusion. The inclusion of structural distractors in the MRT might be a barrier to this test format ever demonstrating a direct relationship between item difficulty and degree of rotation, because structural distractors might make it too easy for test-takers to use strategies that do not involve mental rotation, as they can identify the shapes that are not alike. Thus, our study results also indicate that the MRT test format might not be appropriate for measuring mental rotation.

The fact that participants reported enjoying the experience of taking the VRMRA stands in contrast to comments we have received when administering the Revised PSVT:R to similar groups of students in the past. For some participants, this was their first experience using VR and the novelty factor may have added to their engagement in the test. Additionally, the fact that the test provided feedback about the right or wrong answer through a sound and an animation may have aided in performance for anyone who did not fully understand the directions at first. Our observation during the pilot testing was that some participants can be quite confident in their answers that later turn out to be wrong. Since the MRT and Revised PSVT:R do not include any feedback, it is possible that someone could go through the questions thinking they are doing well when they actually are not. Therefore, the feedback provided in the VRMRA may lead to improved performance since subjects are less likely to proceed through the test with misplaced confidence in their ongoing performance.

Future work

Our test may have been easier than the MRT because the performance in our test closely matched that of Caissie and colleagues' subsample which likely included mostly high performers. However, the purpose of our analysis was not to compare difficulty, just response patterns. Future work could compare the difficulty of the PSVT:R and MRT with the VRMRA in a within-subjects or between-subjects experimental design. A future study could also examine the impact of feedback in the VRMRA on test performance by A/B testing versions with and without feedback. Another comparison could be made with a version of the VRMRA that contains a time limit, since the original MRT and PSVT:R are often administered with time limits. We did not include a time limit in this initial pilot study since we did not know how long it would take people to answer questions, but now that this study is complete, we have some data about the timing which could be used to inform future versions of the instrument.

Future work should also study more closely what factors contribute to item difficulty in part 1 of the VRMRA. One element that may have contributed to item difficulty in part 1 was the orientation of the correct answer choice items and which features were facing away from the viewer. For example, question 3 was the most difficult, and in this question the right answer has no features on the three sides facing the viewer, which may have led people to overlook this answer option. Multiple participants expressed discontent about not being allowed to walk behind the table to view the shapes from the back. In a future experiment, we could use a bigger room and modify the design of the table so that all shapes are visible from all four sides and allow participants to walk behind the table as well. Or, we could remove the table and have the shapes floating in mid-air, which would also allow participants to view the undersides of the shapes. We could also add eye tracking, which many VR headsets are capable of, to better understand how participants were taking in information about the shapes from different viewpoints. It is possible that participants who walked around more and looked at the shapes from multiple viewpoints performed better, but we did not track this. Test-takers may adopt different strategies in the VRMRA than in the traditional tests, and this difference in strategy may also have contributed to the relative changes in question difficulty. Future work could investigate strategy use in the VRMRA compared to the traditional test formats. Future work could seek to better understand the differences between mental rotation processes in a PSVT:R-style task, which uses solely 90- and 180-degree rotations around the cardinal axes but uses a variety of types of shapes, with the MRT style task which uses a much narrower variety of shapes but freeform rotation positions. A future version of our application could mix and match shape styles and rotation styles in order to learn more about this.

Limitations

This study used university students as participants, most of whom were in technology-related majors, 56% of whom were male and 50% of whom were white. This group of people is not representative of the general population, which may limit the generalizability of our findings. The sample size of 68 people is also relatively small compared to other studies of mental rotation.

This study was designed based on our theory that the lack of clarity of the drawings used in the original MRT and PSVT:R impacts the tests' ability to assess mental rotation skills, as the lack of clarity causes item difficulty to not relate to the degree or complexity of rotation needed to solve the questions. We designed the VRMRA to enable test takers to more clearly understand the 3D shapes presented in each question by presenting the shapes with more realistic 3D shading and in an immersive VR environment where the viewer could view the shapes from multiple angles, with the goal of creating a more accurate measure of mental rotation skill. However, we recognize that there are other differences in our instrument in comparison to the original instruments besides merely the presentation of the shapes. These differences include the size of shapes relative to the viewer, the feedback provided to the viewer, and lack of a time limit (time limits are sometimes, but not always, used with the MRT and PSVT:R). While we believe that the presentation of the shapes is the reason for the changes in relative question difficulty between the VRMRA and the original instruments, it is possible that these other modifications may also have factored into our results.

Conclusion

In the present work, we sought to improve the accuracy of mental rotation assessments by leveraging VR technology. We created the VRMRA, an instrument derived from two popular tests of mental rotation: the Revised PSVT:R and the MRT. The VRMRA was designed to run on the Meta Quest 2 headset and presents mental rotation questions in a room-scale environment. We found that the response patterns for identical questions presented in the VRMRA and the traditional mental rotation tests were different in some critical areas. In the VRMRA, we found that problems with simple rotations were significantly easier than problems with complex rotations, which is not the case in the Revised PSVT:R. This suggests that the VRMRA is more accurately measuring mental rotation ability than the Revised PSVT:R, most likely due to the more realistic and clear presentation of the 3D shapes in VR. However, one persisting issue in both the Revised PSVT:R and the VRMRA is that the degree of rotation needed to solve a question is not significantly correlated with item difficulty. For the MRT-style questions in the VRMRA, a moderate negative correlation was found between the degree of shape rotation in the X direction and the item difficulty. The VRMRA appeared to eliminate the factor of occlusion which contributes to item difficulty in the MRT. However, the shape configuration type and the presence of mirrored versus structural distractors were still found to explain 50.6% of the variance in item difficulty in the VRMRA, and suggesting that factors besides mental rotation, such as shape recognition, continue to play a role.

In conclusion, the VRMRA is not a perfect measure of mental rotation ability, but it appears to be an improvement on existing assessments. Additionally, our findings demonstrate that the fundamental designs of the MRT and PSVT:R may not lend themselves well to the assessment of mental rotation skill. While adapting these long-popular tests to VR appeared to improve the relationship between degree of rotation and item difficulty, we found that some factors besides degree of rotation continued to contribute to item difficulty. Thus, future versions of mental rotation assessments may be more accurate if they abandon the precedents of the MRT and PSVT:R and follow entirely new designs.

Dissemination

The VRMRA is now publicly available on GitHub at: https://github.com/krisd1024/VRMRA-application

This link will remain publicly available and future updates to the application will be published to this repository. Users can download the .apk file from the “Builds” folder on the GitHub repository and use Meta Quest Developer Hub or a similar application to install the file onto a Meta Quest 2 headset. The headset should have “developer mode” enabled in order to install the application. The VRMRA application saves answers in a comma separated values (CSV) file on the headset at the end of the assessment, which can be retrieved by connecting the headset to a computer after each administration. Researchers are encouraged to contact the corresponding author if they have further questions about using the VRMRA application.

6 ACKNOWLEDGEMENTS

Part of this research work has been supported by the Spanish Ministry of Education and Vocational Training under a FPU fellowship (FPU19/03878). Additionally, the stay of author Almudena Palacios-Ibáñez at Purdue University was funded by the Universitat Politècnica de València (grants for mobility of doctoral students from the Universitat Politècnica de València for stays in 2022).

7 REFERENCES

[1] Sunita Ariali. 2020. Training of mental rotation ability in virtual spaces. *J. Tech. Educ.* 8, 2 (2020), 46–63.

[2] Sunita Ariali and Bernd Zinn. 2021. Adaptive Training of the Mental Rotation Ability in an Immersive Virtual Environment. 16, 9 (2021), 20–39.

[3] Kinnari Atit, David H. Uttal, and Mike Stieff. 2020. Situating space: using a discipline-focused lens to examine spatial thinking skills. *Cogn. Res. Princ. Implic.* 5, 1 (December 2020), 19. DOI:https://doi.org/10.1186/s41235-020-00210-z

[4] Kristin A. Bartlett and Jorge Dorribo Camba. 2022. Isometric projection as a threat to validity in the PSVT:R. Minneapolis, MN.

[5] Kristin A. Bartlett and Jorge Dorribo Camba. 2022. An argument for visualization technologies in spatial skills assessment. In *Learning and Collaboration Technologies. Designing the Learner and Teacher Experience. Lecture Notes in Computer Science*, Springer, Cham., Online due to COVID-19, 30–39. DOI:https://doi.org/10.1007/978-3-031-05657-4_3

[6] Kristin A. Bartlett and Jorge Dorribo Camba. 2023. The role of a graphical interpretation factor in the assessment of spatial visualization: A critical analysis. *Spat. Cogn. Comput.* 23, 1 (2023), 1–30. DOI:https://doi.org/10.1080/13875868.2021.2019260

[7] Douglas A. Bors and François Vigneau. 2011. Sex differences on the mental rotation test: An analysis of item types. *Learn. Individ. Differ.* 21, 1 (February 2011), 129–132. DOI:https://doi.org/10.1016/j.lindif.2010.09.014

[8] Theodore J Branoff. 2000. Spatial Visualization Measurement: A Modification of the Purdue Spatial Visualization Test - Visualization of Rotations. *Eng. Des. Graph. J.* 64, 2 (2000), 14–22.

[9] Andre F. Caissie, Francois Vigneau, and Douglas A. Bors. 2009. What does the mental rotation test measure? An analysis of item difficulty and item characteristics. *Open Psychol. J.* 2, 1 (December 2009), 94–102. DOI:https://doi.org/10.2174/1874350100902010094

[10] Chen-Wei Chang, Jun Heo, Shih-Ching Yeh, Hui-Ya Han, and Mengtong Li. 2018. The Effects of Immersion and Interactivity on College Students' Acceptance of a Novel VR-Supported Educational Technology for Mental Rotation. *IEEE Access* 6, (2018), 66590–66599.

[11] Jack Shen-Kuen Chang. 2017. The Design and Evaluation of Embodied Interfaces for Supporting Spatial Ability. In *Proceedings of the Eleventh International Conference on Tangible, Embedded, and Embodied Interaction*, ACM, Yokohama Japan, 681–684. DOI:https://doi.org/10.1145/3024969.3025033

[12] Jack Shen-Kuen Chang, Georgina Yeboah, Alison Doucette, Paul Clifton, Michael Nitsche, Timothy Welsh, and Ali Mazalek. 2017. Evaluating the effect of tangible virtual reality on spatial perspective taking ability. In *Proceedings of the 5th Symposium on Spatial User Interaction*, ACM, Brighton United Kingdom, 68–77. DOI:https://doi.org/10.1145/3131277.3132171

[13] Qian Chen, Lixia Deng, Tao Xu, and Yun Zhou. 2022. Visualized Cues for Enhancing Spatial Ability Training in Virtual Reality. Virtual Event, 299–300. DOI:https://doi.org/DOI 10.1109/VRW55335.2022.00067

[14] Randi A Doyle and Daniel Voyer. 2013. Bodies and occlusion: Item types, cognitive processes, and gender differences in mental rotation. *Q. J. Exp. Psychol.* 66, 4 (2013), 801–815.

[15] John Eliot and I Macfarlane Smith. 1983. *An International Directory of Spatial Tasks*. NFER-NELSON Publishing Company, Windsor, Berkshire.

[16] Michael C. Felix, Joshua D. Parker, Charles Lee, and Kara I. Gabriel. 2011. Real Three-Dimensional Objects: Effects on Mental Rotation. *Percept. Mot. Skills* 113, 1 (August 2011), 38–50. DOI:https://doi.org/10.2466/03.22.PMS.113.4.38-50
[17] Maryanne L. Fisher, Tami Meredith, and Melissa Gray. 2018. Sex differences in mental rotation ability are a consequence of procedure and artificiality of stimuli. *Evol. Psychol. Sci.* 4, 2 (June 2018), 124–133. DOI:https://doi.org/10.1007/s40806-017-0120-x
[18] Nora Argelia Aguilera González. 2018. Development of spatial skills with virtual reality and augmented reality. *Int. J. Interact. Des. Manuf. IJIDeM* 12, 1 (February 2018), 133–144. DOI:https://doi.org/10.1007/s12008-017-0388-x
[19] Renata Gorska and Sheryl A. Sorby. 2008. Testing Instruments for the Assessment of 3D Spatial Skills. In *2008 ASEE Annual Conference & Exposition Proceedings*, ASEE Conferences, Pittsburgh, Pennsylvania, 13.1196.1-13.1196.10. DOI:https://doi.org/10.18260/1-2--4411
[20] Roland B. Guay. 1976. Purdue Spatial Visualisation Test: Rotations.
[21] Tibor Guzsvinecz, Eva Obran-Mihalyko, Cecilia Sik-Lanyi, and Erika Perge. 2021. Investigation of spatial ability test completion times in virtual reality using a desktop display and the Gear VR. *Virtual Real.* (2021).
[22] Tibor Guzsvinecz, Éva Orbán-Mihálykó, and Erika Perge. 2020. Analyzing the Spatial Skills of University Students with a Virtual Reality Application using a Desktop Display and the Gear VR. *Acta Polytech. Hung.* 17, 2 (2020), 22.
[23] Tibor Guzsvinecz, Cecilia Sik-Lanyi, Eva Orban-Mihalyko, and Erika Perge. 2020. The Influence of Display Parameters and Display Devices over Spatial Ability Test Answers in Virtual Reality Environments. *Appl. Sci.* 10, 2 (January 2020), 526. DOI:https://doi.org/10.3390/app10020526
[24] Nathan W. Hartman, Patrick E. Connolly, Jeffrey W. Gilger, Gary R. Bertoline, and Justin Heisler. 2006. Virtual reality-based spatial skills assessment and its role in computer graphics education. In *ACM SIGGRAPH 2006 Educators program on - SIGGRAPH '06*, ACM Press, Boston, Massachusetts, 46. DOI:https://doi.org/10.1145/1179295.1179342
[25] Mary Hegarty. 2018. Ability and sex differences in spatial thinking: What does the mental rotation test really measure? *Psychon. Bull. Rev.* 25, 3 (June 2018), 1212–1219. DOI:https://doi.org/10.3758/s13423-017-1347-z
[26] Mary Hegarty and David A Waller. 2005. Individual Differences in Spatial Abilities. In *The Cambridge Handbook of Visuospatial Thinking*, P Shah and A Miyake (eds.). Cambridge University Press, New York, 121–169.
[27] D.D. Hoffman. 1998. *Visual Intelligence: How We Create what We See*. W. W. Norton & Company.
[28] Jon-Chao Hong, Ming-Yeh Hwang, Kai-Hsin Tai, and Chi-Ruei Tsai. 2018. Training Spatial Ability Through Virtual Reality. Wollongong, NSW, Australia, 1204–1205.
[29] Matt C Howard and Elise C. Van Zandt. 2021. A meta-analysis of the virtual reality problem: Unequal effects of virtual reality sickness across individual differences. *Virtual Real.* 25, (2021), 1221–1246.

[30] Hannes Kaufmann, Mathis Csisinko, Irene Strasser, Sabine Strauss, Ingrid Koller, and Judith Glück. 2008. Design of a virtual reality supported test for spatial abilities. Dresden, Germany.

[31] Jeffrey Kim and Javier Irizarry. 2021. Evaluating the Use of Augmented Reality Technology to Improve Construction Management Student's Spat. *Int. J. Constr. Educ. Res.* 17, 2 (2021), 99–116.

[32] Eric Krokos and Amitabh Varshney. 2022. Quantifying VR cybersickness using EEG. *Virtual Real.* 26, (2022), 77–89.

[33] P. Larson, A. A. Rizzo, J. G. Buckwalter, and A. Van Roo. 1999. Gender Issues in the Use of Virtual Environments. *Cyberpsychol. Behav.* 2, 2 (1999), 113–123.

[34] Po-Han Lin and Shih-Ching Yeh. 2019. How Motion-Control Influences a VR-Supported Technology for Mental Rotation Learning: From the Perspectives of Playfulness, Gender Difference and Technology Acceptance Model. *Int. J. Human–Computer Interact.* 35, 18 (November 2019), 1736–1746. DOI:https://doi.org/10.1080/10447318.2019.1571784

[35] Marcia C Linn and Anne C. Petersen. 1985. Emergence and Characterization of Sex Differences in Spatial Ability: A Meta-Analysis. *Child Dev.* 56, 6 (1985), 1479–1498.

[36] Xiaolong Lou, Xiangdong A. Li, Preben Hansen, and Peng Du. 2021. Hand-adaptive user interface: improved gestural interaction in virtual reality. *Virtual Real.* (2021), 16.

[37] Yukiko Maeda and So Yoon Yoon. 2013. A Meta-Analysis on Gender Differences in Mental Rotation Ability Measured by the Purdue Spatial Visualization Tests: Visualization of Rotations (PSVT:R). *Educ. Psychol. Rev.* 25, 1 (March 2013), 69–94. DOI:https://doi.org/10.1007/s10648-012-9215-x

[38] Yukiko Maeda, So Yoon Yoon, Gyenam Kim-Kang, and P. K. Imbrie. 2013. Psychometric Properties of the Revised PSVT:R for Measuring First Year Engineering Students' Spatial Ability. *Int. J. Eng. Educ.* 29, 3 (2013), 763–776.

[39] Jorge Martín-Gutiérrez, José Luís Saorín, Manuel Contero, Mariano Alcañiz, David C. Pérez-López, and Mario Ortega. 2010. Design and validation of an augmented book for spatial abilities development in engineering students. *Comput. Graph.* 34, 1 (February 2010), 77–91. DOI:https://doi.org/10.1016/j.cag.2009.11.003

[40] Sergo Martirosov, Marek Bures, and Tomas Zitka. 2022. Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality. *Virtual Real.* 26, (2022), 15–32.

[41] W McWilliams, C J Hamilton, and S J Muncer. 1997. On Mental Rotation in Three Dimensions. *Percept. Mot. Skills* 85, (1997), 297–298.

[42] Samuel Messick. 1995. Validity of Psychological Assessment. *Am. Psychol.* 50, 9 (1995), 741–749.

[43] Rafael Molina-Carmona, María Pertegal-Felices, Antonio Jimeno-Morenilla, and Higinio Mora-Mora. 2018. Virtual Reality Learning Activities for Multimedia Students to Enhance Spatial Ability. *Sustainability* 10, 4 (April 2018), 1074. DOI:https://doi.org/10.3390/su10041074

[44] M. Mon-Williams, J. P. Wann, and S. Rushton. 1995. Design factors in stereoscopic virtual-reality displays. *J. Soc. Inf. Des.* 3, 4 (1995), 207–201.

[45] Aljoscha C. Neubauer, Sabine Bergner, and Martina Schatz. 2010. Two- vs. three-dimensional presentation of mental rotation tasks: Sex differences and effects of training on performance and brain activation. *Intelligence* 38, 5 (September 2010), 529–539. DOI:https://doi.org/10.1016/j.intell.2010.06.001

[46] Nils Nolte, Florian Schmitz, Jens Fleischer, Maximilian Bungart, and Detlev Leutner. 2022. Rotational complexity in mental rotation tests: Cognitive processes in tasks requiring mental rotation around cardinal and skewed rotation axes. *Intelligence* 91, (March 2022), 101626. DOI:https://doi.org/10.1016/j.intell.2022.101626

[47] Almudena Palacios-Ibáñez, María Alonso-García, Manuel Contero, and Jorge D. Camba. 2023. The Influence of Hand Tracking and Haptic Feedback for Virtual Prototype Evaluation in the Product Design Process. *J. Mech. Des.* 145, 4 (April 2023), 041402. DOI:https://doi.org/10.1115/1.4055952

[48] T Parsons, Peter Larson, Kris Kratz, Marcus Thiebaux, Brendon Bluestein, J Galen Buckwalter, and Albert A Rizzo. 2004. Sex differences in mental rotation and spatial rotation in a virtual environment. *Neuropsychologia* 42, 4 (2004), 555–562. DOI:https://doi.org/10.1016/j.neuropsychologia.2003.08.014

[49] S. Pastel, D. Burger, C. H. Chen, K. Petri, and K. Witte. 2022. Comparison of spatial orientation skill between real and virtual environment. *Virtual Real.* 26, (2022), 91–104.

[50] M. Peters, B. Laeng, K. Latham, M. Jackson, R. Zaiyouna, and C. Richardson. 1995. A redrawn Vandenberg and Kuse mental rotations test-different versions and factors that affect performance. *Brain Cogn.* 28, 1 (1995), 39–58.

[51] Zygmunt Pizlo. 2008. *3D shape: its unique place in visual perception*. MIT Press, Cambridge, Mass.

[52] Zygmunt Pizlo, Yunfeng Li, and Robert M. Steinman. 2008. Binocular disparity only comes into play when everything else fails; a finding with broader implications than one might suppose. 21, 6 (2008), 495–508.

[53] Kimberly A Pollard, Ashley H. Oiknine, Benjamin T. Files, Anne M. Sinatra, Debbie Patton, Mark Ericson, Jerald Thomas, and Peter Khooshabeh. 2020. Level of immersion affects spatial learning in virtual environments: results of a three-condition within-subjects study with long intersession intervals. *Virtual Real.* 24, (2020), 783–796.

[54] Albert A Rizzo, J Galen Buckwalter, Ulrich Neumann, Carl Kesselman, Marcus Thiebaux, Peter Larson, and Andre Van Rooyen. 1998. The Virtual Reality Mental Rotation Spatial Skills Project. *Cyberpsychol. Behav.* 1, 2 (1998), 8.

[55] Michèle Robert and Eliane Chevrier. 2003. Does men's advantage in mental rotation persist when real three-dimensional objects are either felt or seen? *Mem. Cognit.* 31, 7 (October 2003), 1136–1145. DOI:https://doi.org/10.3758/BF03196134

[56] Anahita Sanandaji, Cindy Grimm, and Ruth West. 2017. Inferring cross-sections of 3D objects: a 3D spatial ability test instrument for 3D volume segmentation. In *Proceedings of the ACM Symposium on Applied Perception*, ACM, Cottbus Germany, 1–4. DOI:https://doi.org/10.1145/3119881.3119888

[57] Erli Sarilita, Yurika Ambar Lita, Dani Rizali Firman, Tracey Wilkinson, Sri Susilawati, Risti Saptarini, Dudi Aripin, and Endang Sjamsudin. 2022. Spatial ability and anatomy learning performance among dental students. *Korean J. Med. Educ.* 34, 4 (December 2022), 309–318. DOI:https://doi.org/10.3946/kjme.2022.239

[58] Roger N. Shepard and Jacqueline Metzler. 1971. Mental Rotation of Three-Dimensional Objects. *Sci. New Ser.* 171, 3972 (1971), 701–703.

[59] Sheryl A. Sorby, Norma Veurink, and Scott Streiner. 2018. Does spatial skills instruction improve STEM outcomes? The answer is 'yes.' *Learn. Individ. Differ.* 67, 2018 (October 2018), 209–222. DOI:https://doi.org/10.1016/j.lindif.2018.09.001

[60] Kay Stanney, Cali Fidopiastis, and Linda Foster. 2020. Virtual Reality Is Sexist: But It Does Not Have to Be. *Front. Robot. AI* 7, (January 2020), 1–17. DOI:https://doi.org/10.3389/frobt.2020.00004

[61] Mike Stieff, Andrea Origenes, Dane DeSutter, Matthew Lira, Lukas Banevicius, Dylan Tabang, and Gervacio Cabel. 2018. Operational constraints on the mental rotation of STEM representations. *J. Educ. Psychol.* 110, 8 (November 2018), 1160–1174. DOI:https://doi.org/10.1037/edu0000258

[62] George Takahashi and Patrick Connolly. 2012. Impact of Binocular Vision on the Perception of Geometric Shapes in Spatial Ability Testing. In *67th EDGD Midyear Meeting Proceedings*, Limerick, Ireland, 26–31.

[63] Adam J. Toth and Mark J. Campbell. 2019. Investigating sex differences, cognitive effort, strategy, and performance on a computerised version of the mental rotations test via eye tracking. *Sci. Rep.* (2019). DOI:https://doi.org/10.1038/s41598-019-56041-6

[64] Róbert Tóth, Marianna Zichar, and Miklós Hoffmann. 2020. Gamified Mental Cutting Test for enhancing spatial skills. Online on MaxWhere 3D Web, 000229–000304.

[65] Emiko Tsutsumi, Wakana Ishikawa, Hiroshi Sakuta, and Kenjiro Suzuki. 2008. Analysis of Causes of Errors in the Mental Cutting Test – Effects of View Rotation. *J. Geom. Graph.* 12, 1 (2008), 109–120.

[66] Steven G. Vandenberg and Allan R. Kuse. 1978. Mental rotations, a group test of three-dimensional spatial visualization. *Percept. Mot. Skills* 47, 2 (December 1978), 599–604. DOI:https://doi.org/10.2466/pms.1978.47.2.599

[67] Daniel Voyer and Junjie Hou. 2006. Type of items and the magnitude of gender differences on the Mental Rotations Test. *Can. J. Exp. Psychol. Can. Psychol. Expérimentale* 60, 2 (2006), 91–100. DOI:https://doi.org/10.1037/cjep2006010

[68] Daniel Voyer, Susan Voyer, and M P Bryden. 1995. Magnitude of Sex Differences in Spatial Abilities: A Meta-Analysis and Consideration of Critical Variables. *Am. Psychol. Assoc.* 117, 2 (1995), 250–270.

[69] So Yoon Yoon. 2011. Psychometric Properties of the Revised Purdue Spatial Visualization Tests: Visualization of Rotations (The Revised PSVT:R) [Doctoral Dissertation, Purdue University] (Publication Number 3480934).

[70] S.Y. Yoon. 2011. Revised Purdue Spatial Visualization Test: Visualization of Rotations (Revised PSVT:R) [Psychometric Instrument].

[71] Jianping Yue. 2006. Spatial visualization by isometric drawing. In *Proceedings of the 2006 IJME*, Union, New Jersey.

[72] Jianping Yue. 2008. Spatial Visualization by Realistic 3D Views. *Eng. Des. Graph. J.* 72, 1 (2008), 28–38.