VR Geoscience Education: Building Spatial Reasoning Skills

Katharine E. Johanesen*
Department of Geology
Juniata College

J. Adam Jones†
High Fidelity Virtual Environments Lab (Hi5 Lab)
Mississippi State University

Territa Poole
Department of Psychology
Juniata College

Katherine Ryker
Christopher Green
School of the Earth, Ocean and Environment
University of South Carolina

2024 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW) | 979-8-3503-7449-0/24/$31.00 ©2024 IEEE | DOI: 10.1109/VRW62533.2024.00034

Abstract

We tested the impact of a 15-minute VR training on spatial skills and performance on a geoscience task with a control group. The VR group improved more on the Water Level Task—a measure of understanding of horizontal (B = 0.68, p=0.008). Both groups performed equally on the geology task, except for an orientation rule not well instructed in the VR module (B = -1.33, p=0.0057). In the post-survey, the VR group reported higher ability to link knowledge (X^2=4.45, p=0.035) and more interest than in past activities (X^2=8.47, p=0.004). This is encouraging, given the brevity of the VR lesson.

Index Terms: K.6.1 [Management of Computing and Information Systems]: Project and People Management—Life Cycle; K.7.m [The Computing Profession]: Miscellaneous—Ethics

1 Introduction

Geoscience requires strong spatial reasoning abilities, especially 3D skills like sense of direction, spatial visualization, and mental rotation. Not surprisingly, geologists performed among the highest in spatial reasoning skills among academic disciplines [3, 6]. Further, a longitudinal study shows that spatial ability in adolescence correlates with persistence in STEM fields [21]. However, these skills are not directly trained in a typical undergraduate geoscience curriculum. If spatial reasoning skills are so important for success in geoscience and STEM fields, how can we better support their acquisition in the classroom? We tested the use of Virtual Reality (VR) to train spatial reasoning skills in college students both in the abstract and applied to geoscience content.

Spatial reasoning is a set of cognitive skills that we use to understand and visualize objects or scenes in 2D or 3D. Research has demonstrated multiple, separate skills on which individuals can have variable ability levels. The skills that comprise spatial reasoning, according to Newcombe and Shipley [16], are:

- Disembedding: The ability to pick out specific information from a complicated image or scene.
- Spatial Visualization: Includes the ability to translate between 2D and 3D and to imagine changes to the shape of objects.
- Mental Rotation: The ability to visualize rigid-body rotations of objects.
- Spatial Perception: Understanding extrinsic abstract spatial concepts such as vertical and horizontal orientations.
- Perspective Taking: The ability to visualize an object or scene from different points of view.

Ability levels in these spatial skills vary at the individual level and have been demonstrated to correlate with upbringing and activities from childhood onward [2, 8, 15]. Some studies show disparities in gains in spatial performance by sex and race [22], indicating that differences in acquisition of spatial reasoning skills may affect equity, diversity, and inclusion efforts in STEM. Spatial reasoning ability levels are malleable, with improvements shown through direct training [13, 19].

One specific task almost every college geoscience student must learn is measuring the orientation of an inclined plane. This is a common task performed by geologists in the field to gather data about the position of folded rock layers and other features. It requires a strong command of 3-dimensional spatial reasoning. The orientation of an inclined plane is reported using two numbers: 1) the map direction of a horizontal line on that plane reported as degrees from north, called the strike, and 2) the angle of inclination of the plane measured from horizontal, called the dip. Measurements are taken using a handheld compass/clinometer tool. To successfully take a measurement, geologists must identify the inclined plane in question, visualize a horizontal plane intersecting with it, and position the compass parallel to the line of intersection. Then, they turn the tool on its side and use the clinometer to measure the slope of the steepest line down the plane or the intersection of the plane with the vertical. Each stage of this task requires multiple spatial skills, including disembedding (isolating the plane to measure), spatial perception (identifying an imaginary horizontal plane intersecting it), spatial visualization (imagining a line of strike at that intersection), mental rotation (imagining the possible ways to rotate the measurement tool) and perspective taking (determining the right side from the tool's point of view).

We developed a VR lesson to build the requisite spatial skills and then teach the task of measuring strike and dip. This study tests the effectiveness of the VR lesson in improving spatial abilities and teaching the strike and dip measurement task. In addition, we explore responses to a short, post-experience survey.

2 Methods

The experiment was conducted in a large university introductory geology lab course, in which students participated in a two-hour lab session during their regularly scheduled lab time. Four graduate student Instructional Assistants (IAs) were trained to administer the lab activities, and the researchers ran the VR training in a separate room nearby. The participants were given a pre-test online prior to their lab session and a post-test and demographic survey online afterwards. Entire lab sections were assigned to either VR or Standard training, balancing groups for time of day, day of week, and ensuring that each IA taught at least one lab with each training type. Both groups watched a series of videos about measuring strike and dip and received instruction from the IA about how to make a geologic map. Next, the IAs administered a short paper survey of the Water

*e-mail: johanesen@juniata.edu
†e-mail: jadamj@acm.org

979-8-3503-7449-0/24/$31.00 ©2024 IEEE
DOI 10.1109/VRW62533.2024.00034

Level Task (WLT). Next, each group completed a training activity. Students who had not completed the online survey prior to class were asked to do so before starting the training activity. Students in the standard group were asked to practice their measurement skills in the classroom. Students in the VR group spent 10-15 minutes in a VR module designed to train strike and dip measurement with a Brunton-style compass-clinometer. After completing their assigned training, all students completed a geologic mapping lab activity that involved measuring planes, identifying rocks, and plotting their data to create a simple geologic map. Post-test activities included an online survey administered outside of class time as well as a paper WLT test completed at the beginning of lab the following week.

2.1 Participants and Recruitment

All participants were students in a large introductory geology course serving mostly non-majors at a public university. Students participated in the activity as part of their regularly scheduled laboratory session. All students were asked to complete the pre- and post-tests as part of the laboratory activity. They were provided an informed consent document and were allowed to opt out of their responses being included in the study. All researchers completed Institutional Review Board (IRB) training in advance of the experiment and the procedure was determined to be exempt by the university's IRB.

Approximately 200 students attended the class the week of the study, of whom 160 consented to being included in the study. The number of participants varies between tests due to some participants skipping questions or not completing the post-test.

2.2 Materials

The pre- and post-tests consisted of three spatial measures: the Spatial Vocabulary test, the Visualization of Views test, and the Water Level task. In addition, measures of social desirability, verbal ability, and spatial anxiety were taken on the pre-test. Demographic and user experience questions were collected with the post-test. Part of the lab activity consisted of a post-training Brunton test activity.

2.2.1 Spatial Vocabulary Test

The spatial vocabulary test was designed for this study to determine whether participants understood the terms used in questions about spatial concepts. We focused the questions on the terms used in descriptions of orientation of objects. The questions are adapted from terminology used in the class and were piloted with geoscience majors and nonmajors to refine the instrument. The six-item matching question asked participants to find the best definition for the terms Vertical, Flat, Horizontal, Parallel, Orientation, and Perpendicular. The definitions and terms remained the same for the pre- and post-tests, but the order was changed.

2.2.2 Visualization of Views Test

We used a modified Visualization of Views (VoV) Test based on the test first constructed by Guay [5] and subsequently modified by Eliot and Smith [4]. This test measures perspective taking (visualizing an object from a new viewing direction). Each test item consists of a 2D sketch of a 3D object in a box and an alternate view of the object at a new orientation. The corners of the box represent the possible viewpoints to observe the original object and participants are instructed to choose the viewing direction that would produce the alternate view presented. This modified VoV test has been utilized with undergraduate populations to successfully measure spatial abilities [1].

2.2.3 Water Level Task

The Water Level Task (WLT) questions are modeled after Piaget and Inhelder [17] and continued research use [11, 14, 20]. Liben et al. [12] demonstrated construct validity of WLT by relating scores

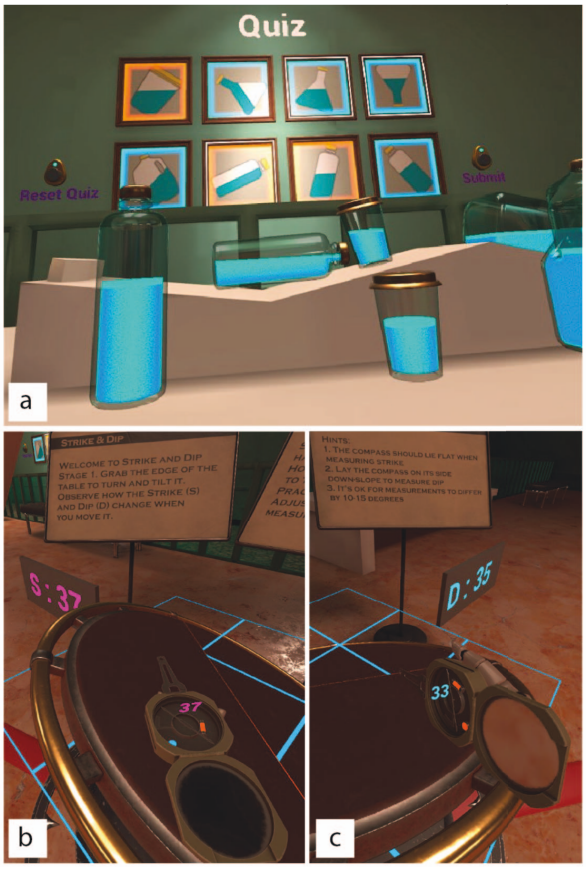

Figure 1: Screenshots of the VR Geoscience module: a) a table of vessels with an interactive WLT quiz on the wall behind; b) the compass tool held in position to measure strike on the tilting table, with automated strike readout in view; c) the tool held in position to measure dip, with automated dip readout in view.

on the WLT with strike and dip tasks in geoscience, which are recognized to require an understanding of the horizontal.

The participant is instructed "Below, you will see a series of tilted water bottles on a horizontal line which represents a table. Fill in the correct water line if the bottle was about half full and resting at this angle." The bottle is drawn tilted at five different angles with a horizontal line representing the table beneath it. An example of the test adapted to multiple choice for the VR module is pictured in Figure 1.

The WLT questions were piloted with a separate group of 26 undergraduate students. Based on the pilot responses, we set a cut-off of no more than 10 degrees from parallel for the angle of the water and no more than 1 mm amplitude for curvature of the water surface. The tests were scanned and angles from horizontal measured using an image analysis tool. Amplitude is measured as the largest deviation from a linear average of the curved line.

2.2.4 User Experience Questions

In the online post-activity survey, participants were asked to answer questions describing their learning experience and comparing it to other lab sessions in the course. The VR and classroom training groups each received a separate version of the questions. These user experience questions are provided in Table 11. A subset of questions was asked only of the VR training group.

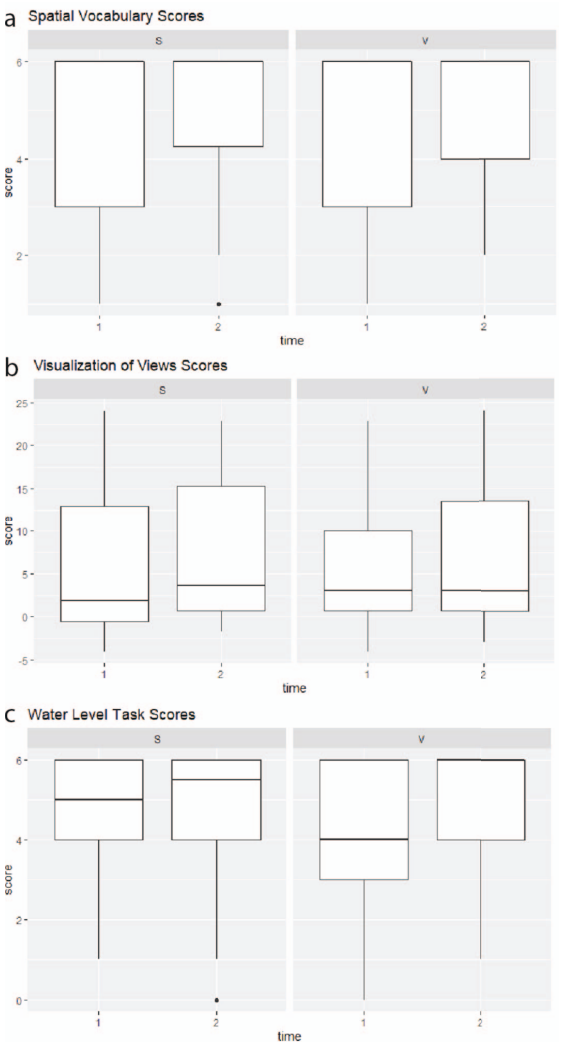

Figure 2: Box plots of pre (1) and post-test (2) scores divided by treatment condition (S = standard, V = VR) for the three spatial tasks: a) Spatial Vocabulary, b) Visualization of Views, and c) Water Level Task.

2.2.5 Instructional Videos

Prior to beginning the lab activity, students watched a series of three videos and were given some instruction on creating a geologic map by their Instructional Assistants (IAs). The first video, "Interpreting Strike and Dip" [7] provides a simple overview of the concept of strike and dip. The second video, "Intro to the Brunton Compass" [9] provides detail on the relevant parts of the compass and when they are used. The third video, "Using a Brunton" [10] explains the steps to measure the strike and the dip of a plane using a compass/clinometer.

2.2.6 VR equipment

All participants in the VR group used a Meta Quest 2 Virtual Reality headset. Researchers pre-loaded the VR module and guided users in wearing and adjusting the equipment using a standardized script.

2.2.7 VR module

The VR module was developed in Unreal Engine 4. A guiding concept of this work was to break down a complicated task into the requisite pieces to support acquisition of the skills being taught. This practice is supported by Cognitive Load Theory [18].

Upon starting the module, participants are immersed in a museum-like space. Signs near the starting location guide participants through basic use of the controllers to navigate. A solid line on the floor traces a path through several tutorial stations and then on to the training activity. Each tutorial station teaches a new interaction: picking up objects, using the laser pointer to select options and submit a quiz. One table is covered in cubes the user can pick up, the other has a submit and reset button with a wall of picture frames behind it. Participants are instructed to select correct answers (images of a check mark), then press the submit button. Upon selection, the framed images are ringed in a white glowing light. After pressing submit, the responses that are correctly selected (and correctly not selected) are highlighted in blue, while incorrectly selected answers (and incorrectly not selected answers) are highlighted in orange. The white light remains when the scoring colors appear so that users can view their own selections and reflect on their answers. No participants reported color blindness during this activity. Instructions also suggest that users should try out some incorrect answers to see what happens. A sign at the final tutorial activity instructs participants to move on to the water level activity when they are ready.

At the water level station, a series of clear vessels that are half-filled with liquid are posed on the table leaning on a crooked shelf in various inclined positions. A sign instructs participants to pick up the vessels and experiment with how their orientation affects the liquid inside. On the wall to the left, the word "vertical" is painted vertically in the style of an art exhibit. When participants move closer, they can read a simple definition of vertical. On the floor along that wall, "horizontal" is painted, also accompanied with a definition placard. Behind the table of vessels is a wall of picture frames that show a set of random Water Level Task multiple choice items (Figure 1). A sign near the frames asks participants to choose all the items that have the correct water level orientation, then click the submit button to check their answers. The selection and scoring mechanism is the same as in the tutorial. Signs ask participants to try the quiz several times before moving on to the next station.

The next station features a display with several oversized Brunton-style compasses in various positions. One is placed in a horizontal position with the dial facing the ceiling. One is placed on its side to illustrate setting up for measuring dip. Two more are arranged on tilted planes with labels: one in the proper orientation for measuring strike and the other oriented to measure dip. Signs instruct the student to observe these and re-familiarize themselves with the compass layout and correct positions for measurement.

The final station features a round table that can be tilted and turned. Numbers beside the table automatically update to reflect its current orientation. Signs near the table instruct students to move and tilt it and observe how the strike and dip numbers change, then use the compass attached to their hand to measure the plane and try to reproduce the numbers in the display (Figure 1). The signs instruct the student to adjust the table and repeat the measurement at least four times before exiting the VR experience. In total, participants spent 10-20 minutes in the VR module.

2.2.8 Standard Assignment

Participants in the standard groups first watched the same introductory videos and completed the same WLT pre-test as the VR groups. After this, the IA reviewed the basics of measuring with a compass/clinometer and asked the whole class (each with a compass in hand) to point north, then east, to practice basic compass directions. The IA then instructed students to practice measuring strike and dip on small cardboard planes at their seats. The students were asked to measure the various sides of the cardboard wedge at four different orientations. When this was complete, students then moved on to the Geologic Mapping Activity.

Table 1: Responses to the post-activity user experience questions (n=94: 49 VR and 45 Standard group), reporting numbers of a) True (T) and False (F) responses and percent true responses for each treatment condition (VR = experimental virtual reality, Std = Standard classroom instruction), b) collapsed Likert scale responses (A = somewhat agree + strongly agree, D = somewhat disagree + strongly disagree, neutral not shown) and percent agree for each treatment condition, and c) binary responses to questions about the ease and interest in the laboratory activity. Chi-square of differences and p are reported for all questions asked of both treatment groups: $p < 0.01$ in bold with *, $p < 0.05$ in bold.

Question			VR	Std		
a. True/False	**T**	**F**	**%True**		**X^2**	**p**
I became more interested to learn about geologic mapping	63	31	73%	60%	1.36	0.243
I am confident and can understand the basic concepts of Strike and Dip	82	12	88%	87%	0.00	1.000
Measuring Strike and Dip is fun to do	53	41	65%	47%	2.60	0.107
I can apply what I've learned in this lab in a real context	75	19	86%	73%	1.53	0.216
I was able to link new knowledge with my previous knowledge and experiences	**78**	**16**	**92%**	**73%**	**4.45**	**0.035**
I expect to do well in this course	83	11	86%	91%	0.24	0.623
This VR experience helped me learn	43	6	88%	NA		
This VR experience allowed me to be more responsive and active in the learning process	40	9	82%	NA		
Learning how to use this VR application was too complicated and difficult for me	23	26	47%	NA		
b. Likert Scale (reduced to agree/neutral/disagree)	**A**	**D**	**%Agree**			
I gained a good understanding of the basic concepts of geologic mapping	67	10	63%	80%	4.98	0.083
The ability to manipulate the objects within the virtual environment makes learning more motivating and interesting	32	6	65%	NA		
I was satisfied with this [VR-based] learning experience	59	9	63%	62%	0.09	0.954
This [VR/learning] experience was boring	31	38	29%	38%	3.12	0.210
This VR experience allowed me to have more control over my own learning	29	7	59%	NA		
This VR experience helped me engage in the learning activity	33	6	67%	NA		
The realism of the VR experience helps to enhance my understanding	29	5	59%	NA		
Overall, I think this VR application is easy to use	30	5	61%	NA		
c. Binary Responses	**+**	**-**	**%+**			
Compared to other labs this semester, this lab was: easier (+), more difficult (-)	52	42	65%	44%	3.33	0.068
Compared to other labs this semester, this lab was: more interesting (+), less interesting (-)	**72**	**22**	**90%**	**62%**	**8.47**	**0.004***

2.2.9 Geologic Mapping Activity

Following the experimental or standard training session, students used the remaining time in lab to construct a simplified geologic map in the classroom using mock outcrops. The mock outcrops consist of five stations with tilted boards and rock samples, arranged in the shape of a folded rock structure. At each station, students identified the rock and measured the orientation of the plane. Finally, they plotted their measurements on a simplified map (with numbered positions for each station) and drew in contacts between the rock units.

The strike and dip measurement portion was designed to be a post-test activity. Because of this, IAs were instructed to have students work alone for the measurement portion of the activity. Unfortunately, this was not achieved, evidenced by nearly identical answer sets in many lab sessions.

3 Results

These results focus on direct comparison of spatial test and Brunton test scores between the two treatment groups, as well as their response to user experience questions. Other factors, including the influence of demographic factors, spatial anxiety, or interactions between multiple factors are beyond the scope of this paper.

3.1 Spatial Tests

Linear regression shows no difference between treatment groups on pre-test scores on the Spatial Vocabulary (n=149) and Visualization of Views (VoV) (n=102) tests, indicating that the two groups were relatively equivalent in these spatial abilities (Table 2). The VR group's pre-test scores on the Water Level Task (WLT) were significantly lower than that of the standard training group (B= -0.63, p=0.024), indicating a pre-existing difference in this ability. Both treatment groups showed minor improvement from pre-test to post-test for Spatial Vocabulary. Both groups showed no significant difference in pre- vs. post-test scores on the VoV test (Figure 2). Only the VR training condition group showed improvement on the WLT.

To further explore the impact of the two training conditions on spatial scores, we calculated difference scores for each test (post-pre). Linear regression analysis of difference scores (Table 3) shows no relationship between difference scores and training condition for the Spatial Vocabulary (p=0.387, n=86) and VoV (p= 0.737, n=49) tests. The VR group showed 11% higher improvement scores on the WLT compared to the standard group (p = 0.008, n=95).

3.2 Strike and dip task

Initial scoring and analysis of the strike and dip task responses revealed that the majority of students in eight of the lab sections had nearly identical responses within the section. As identical measurements of strike and dip are unlikely even among experts due to natural variations in the surface and measurement tool, we concluded that these lab sections collaborated and that their responses do not reflect the individual's comprehension of the content in question. To resolve this issue, we used a subset of four lab sections in which students submitted unique responses (n=41).

Table 4 reports linear regression results for each component of the measurement task, with two ways of scoring the strike responses. We found no significant differences on the strike and dip task except with

Table 2: Simple linear regression results for pre-test scores on the three spatial tests as predicted by training condition. $p < 0.05$ in bold.

y	x	B	SE	t	p	F(df)	R^2 (adj)
vocab	(intercept)	4.62	0.20	22.65	<0.001	0.012 (1, 145)	0.91
	training	-0.03	0.27	-0.11	0.913		
VoV	(intercept)	5.92	1.22	4.86	<0.001	0.022 (1, 100)	-0.01
	training	-0.23	1.59	-0.15	0.883		
WLT	(intercept)	4.76	0.21	23.05	<0.001	5.20 (1, 150)	0.03
	training	**-0.63**	**0.28**	**-2.28**	**0.024**		

Table 3: Simple linear regression results for difference scores on the three spatial tests as predicted by training condition. $p < 0.01$ = bold with *.

y	x	B	SE	t	p	F(df)	R^2 (adj)
Vocab	(intercept)	0.31	0.28	1.09	2.770	0.7551 (1, 84)	0.00
	training	0.33	0.38	0.87	0.387		
VoV	(intercept)	1.48	1.01	1.47	0.148	0.1145 (1, 47)	-0.02
	training	0.46	1.36	0.34	0.737		
WLT	(intercept)	-0.02	0.18	-0.13	0.900	7.448 (1, 93)	0.06
	training	**0.68**	**0.25**	**2.73**	**0.008***		

application of the right-hand rule (RHR). The classroom instruction group performed 26% higher on the strike task when the RHR was taken into account (p = 0.006), but this difference does not persist when scoring is agnostic to the RHR (no RHR in table 4).

3.3 User experience responses

Responses by training condition are provided in Table 1. Responses from both training groups were generally positive. The total number of responses used in this portion of the study was 94: 49 from the VR group and 45 from the classroom group. The VR group gave strong positive responses to the True/False questions "This VR experience helped me learn" (88% true) and "This VR experience allowed me to be more responsive and active in the learning process" (82% true). Chi-squares of the True/False questions show no significant differences between the VR and classroom training groups, except for the question "I was able to link new knowledge with my previous knowledge and experience," in which the VR group scored 19% higher.

Likert-scale responses were collapsed into three categories for analysis: somewhat agree and strongly agree were collapsed into "agree" and somewhat and strongly disagree collapsed into "disagree." Total agree and disagree responses and percent agree for each training group is reported in Table 1, as well as Chi-squares for the Likert-scale questions that were asked of each group. No significant differences were found between groups for the three Likert-scale questions (n=94).

Of the two final, binary response questions asked of all participants, the VR group scored 38% higher (more interesting) on the question "Compared to other labs this semester, this lab was more interesting/less interesting."

Table 4: Simple linear regression results for strike and dip scores as predicted by training condition. $p < 0.01$ = bold with *.

y	x	B	SE	t	p	F (1, 39)	R^2 (adj)
strike (w/ RHR)	(intercept)	3.12	0.35	8.99	<0.001	8.56	0.01
	training	**-1.33**	**0.45**	**-2.93**	**0.0057***		
strike (no RHR)	(intercept)	3.53	0.32	11.02	<0.001	0.02	-0.03
	training	0.05	0.42	0.13	0.898		
dip	(intercept)	2.53	0.40	6.27	<0.001	2.83	0.04
	training	0.89	0.53	1.68	0.101		

4 Discussion

The VR lesson successfully improved participant scores on the Water Level Task (WLT), but had no effect on Spatial Vocabulary or Visualization of Views (VoV) scores. Spatial Vocabulary scores were high in both groups on the pre-test, which may make it difficult to measure any improvement due to a ceiling effect.

The measured improvement on the WLT should be interpreted with caution. The standard group received no instruction on the WLT, while the VR group received practice on this task within the lesson. It is therefore reasonable that they would have higher scores on this task. Another possible reason we see improvement in the VR group only on the WLT may be that they started with a lower mean score on this test (on the pre-test). If more of the standard group scored the maximum score on the pre-test, their mean difference score or improvement would be lower.

Convenience sampling and dividing by labs leads to less control over group composition. We were unable to prevent a between-groups difference in WLT pre-test scores. We believe that the benefits of preserving ecological validity outweigh these limitations.

One possible explanation for the lack of observed improvement on Spatial Vocabulary and VoV tests include the short duration of training. The VR and classroom lesson portion of this experiment lasted between 10-20 minutes. It would be unreasonable to expect dramatic changes in mental abilities from such a short training period. An additional complicating factor is the low participation rate on the VoV task. Many participants skipped these questions on the post-test, resulting in complete results for only 49 participants. A simple solution to this issue in future studies would be to force response to all spatial task questions in both the pre- and post-tests, though this might lower overall test completion rates due to participants quitting early.

The VR lesson appears to be as good as classroom instruction in teaching the skill of measuring strike and dip, except for application of the right-hand-rule to strike measurements. The VR training did not have an interactive mechanism to explain the right-hand rule. This was explained in the introductory videos (watched by all participants) and on signage within the VR lesson. The classroom practice group had the opportunity to ask questions about this concept and receive feedback from their IA, which may have given them an advantage on this aspect of the strike and dip measurement.

The collaboration issue on the post-training strike and dip measurement task emphasizes the need to work closely with IAs to determine the most feasible procedures. Perhaps labeling this activity as a "quiz" would reduce collaboration. Another option would be to send a few students at a time to a separate room (with supervision) to complete the task, however this may be challenging in a two-hour laboratory session.

The VR training group showed considerably higher positive responses to the true/false question "I was able to link new knowledge with my previous knowledge and experiences" and the binary question "Compared to other labs this semester, this lab was (more/less) interesting." Responses were more neutral to questions about the difficulty or ease of use of the VR lesson. These responses affirm the importance of user experience in the design of VR lessons. Some improvements we believe would help improve this score are audio instructions, pop-up reminders of the how-to instructions when a user is idle for too long or "wanders off," and more dynamic instructions for using the strike and dip tool (with 3D animation, for example). We were not able to separate the effect of novelty from the value of the VR lesson in this study.

5 Conclusions

This experimental study evaluated the use of a geology-specific VR spatial training module to improve spatial skills and train geology skills in a geology class. The results are encouraging, but not decisive, regarding the effectiveness of the VR module in training spatial skills and strike and dip measurement.

- Students in the VR group showed higher levels of improvement on the Water Level Task (WLT) measure of horizontal, however caution should be used in interpreting this since the VR group had lower pre-test scores on this test. This shows that VR has the potential to improve spatial skills, though further study is needed to isolate the magnitude of this improvement.
- There was no difference between groups in performance on the geology-specific task of measuring strike and dip, except when the "right hand rule" of strike orientation was taken into account. This indicates both VR and classroom instruction teach the skill equally well, though the VR software was less specific about the right hand rule.
- In user experience questions, students in the VR group reported a higher degree of ability "to link new knowledge with... previous knowledge and experiences" and found this lab activity to be more interesting than other labs that semester to a higher degree than the classroom experience students.

Given the positive outcomes and responses despite the brevity and user-reported difficulty, the VR lesson appears to be a promising tool for training novice geoscience students. Several aspects of the program need to be further refined, however, to increase accessibility and maximize the benefits. Future work should include longer duration or repeated practice VR lessons to increase the impact on geoscience spatial skills.

Acknowledgments

The authors would like to thank Nicole LaDue for her support in developing the research design and grappling with spatial reasoning. We thank Cynthia Baldwin and Royal Yu for their help assembling the Qualtrics survey and piloting the study questions. We also thank Ekaterina Rojas, Gabriel de Souza, Gabriel Faton, Breanna Hirosky, and Reece Hammond for their assistance in conducting the research study and collecting data. Development of the VR module used in this study was funded by the Innovative Educational Initiatives Grant at Juniata College. The research component of this study was funded by National Science Foundation Education and Human Resources Division Core Research Program: Building Capacity in STEM Education Research award (FAIN 2125377).

References

[1] C. A. Cohen and M. Hegarty. Sources of difficulty in imagining cross sections of 3d objects. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, vol. 29, 2007.

[2] C. A. Cohen and M. Hegarty. Visualizing cross sections: Training spatial thinking using interactive animations and virtual objects. *Learning and Individual Differences*, 33:63–71, 2014.

[3] B. A. Colaianne and M. G. Powell. Developing transferrable geospatial skills in a liberal arts context. *Journal of Geoscience Education*, 59(2):93–97, 2011.

[4] J. Eliot and I. Smith. An international directory of spatial tests windsor, berkshire: Nfer-nelson; atlantic highlands, nj: distributed in the usa by humanities press, 1983, 1983.

[5] R. Guay. *Purdue spatial vizualization test*. Educational testing service, 1976.

[6] M. Hegarty, R. D. Crookes, D. Dara-Abrams, and T. F. Shipley. Do all science disciplines rely on spatial abilities? preliminary evidence from self-report questionnaires. In *Spatial Cognition VII: International Conference, Spatial Cognition 2010, Mt. Hood/Portland, OR, USA, August 15-19, 2010. Proceedings 7*, pp. 85–94. Springer, 2010.

[7] C. Hoch. Interpreting Strike and Dip. `https://youtu.be/K5vsoqDpu5Q?si=3rL1pcpsQlUO_MWw`, 2023. [Online; accessed 24-January-2024].

[8] M. Hoffman, U. Gneezy, and J. A. List. Nurture affects gender differences in spatial abilities. *Proceedings of the National Academy of Sciences*, 108(36):14786–14788, 2011.

[9] K. Johanesen. Intro to the Brunton compass. `https://youtu.be/GbcOFfX-JlA?si=a2wQFeheWW3JWFxz`, 2021. [Online; accessed 24-January-2024].

[10] K. Johanesen. Using a Brunton. `https://youtu.be/aB8m980AsCE?si=yuWtUlC0J150autY`, 2021. [Online; accessed 24-January-2024].

[11] L. S. Liben and S. L. Golbeck. Adults' demonstration of underlying euclidean concepts in relation to task context. *Developmental Psychology*, 22(4):487, 1986.

[12] L. S. Liben, L. J. Myers, and A. E. Christensen. Identifying locations and directions on field and representational mapping tasks: Predictors of success. *Spatial Cognition & Computation*, 10(2-3):105–134, 2010.

[13] T. R. Lord. Enhancing the visuo-spatial aptitude of students. *Journal of research in science teaching*, 22(5):395–405, 1985.

[14] A. V. McGillicuddy-De Lisi, R. De Lisi, and J. Youniss. Representation of the horizontal coordinate with and without liquid. *Merrill-Palmer Quarterly of Behavior and Development*, 24(3):199–208, 1978.

[15] A. Moè, P. Jansen, and S. Pietsch. Childhood preference for spatial toys. gender differences and relationships with mental rotation in stem and non-stem students. *Learning and Individual Differences*, 68:108–115, 2018.

[16] N. S. Newcombe and T. F. Shipley. Thinking about spatial thinking: New typology, new assessments. In *Studying visual and spatial reasoning for design creativity*, pp. 179–192. Springer, 2014.

[17] J. Piaget, B. Inhelder, F. Langdon, and J. Lunzer. *The Child's Conception of Space*. Routledge, 1957.

[18] J. Sweller. Cognitive load theory, learning difficulty, and instructional design. *Learning and instruction*, 4(4):295–312, 1994.

[19] D. H. Uttal, N. G. Meadow, E. Tipton, L. L. Hand, A. R. Alden, C. Warren, and N. S. Newcombe. The malleability of spatial skills: a meta-analysis of training studies. *Psychological bulletin*, 139(2):352, 2013.

[20] R. Vasta and L. S. Liben. The water-level task: An intriguing puzzle. *Current Directions in Psychological Science*, 5(6):171–177, 1996.

[21] J. Wai, D. Lubinski, and C. P. Benbow. Spatial ability for stem domains: Aligning over 50 years of cumulative psychological knowledge solidifies its importance. *Journal of educational Psychology*, 101(4):817, 2009.

[22] J. Wilhelm, M. Toland, and C. Merryn. Evaluating middle school students' spatial-scientific performance within earth/space astronomy in terms of gender and race/ethnicity. *Journal of Education in Science Environment and Health*, 3(1):40–51, 2017.