

# Instructor-Led Approach to Integrating an Augmented Reality Sandbox into a Large-Enrollment Introductory Geoscience Course for Nonmajors Produces No Gains

Scott Giorgis,<sup>1</sup> Nancy Mahlen,<sup>1,a</sup> and Kirk Anne<sup>2</sup>

## ABSTRACT

The augmented reality (AR) sandbox bridges the gap between two-dimensional (2D) and three-dimensional (3D) visualization by projecting a digital topographic map onto a sandbox landscape. As the landscape is altered, the map dynamically adjusts, providing an opportunity to discover how to read topographic maps. We tested the hypothesis that the AR sandbox is a more effective tool for teaching topographic maps than the traditional, paper-based approach alone. Students enrolled in an introductory geology course were randomly divided into two groups. The control group ( $N = 82$ ) completed the paper-based topographic maps laboratory course. The experimental group ( $N = 94$ ) completed the same laboratory course but also used the AR sandbox. Students spent 15–20 min working on an instructor-guided exercise to explore concepts such as contours, landforms, profiles, and gradients. The following week, all students took a topographic maps assessment. As in previous studies, these data suggest that male students, those with prior map-reading experience, and those students with higher 3D visualization skills tended to score better on the assessment. Unfortunately, we identified no significant gains in the experimental group over the control group, even when the results were subdivided by gender, prior map experience, and/or spatial visualization skill. These results suggest a short instructor-guided approach is not the most effective way to use the AR sandbox. Future research should focus on determining the amount of time needed to make the AR sandbox effective and/or the efficacy of using the AR sandbox multiple times during the semester for shorter periods of time. © 2017 National Association of Geoscience Teachers. [DOI: 10.5408/17-255.1]

**Key words:** augmented reality sandbox, topographic maps

## INTRODUCTION

Topographic maps are a commonly used medium for describing the shape of the surface of Earth. Topographic data are key factors needed to intelligently manage a wide variety of natural hazards (e.g., flooding, landslides, sea-level rise, etc.); therefore, a basic understanding of how to read and interpret topographic data is necessary not only for future geoscientists, but citizens in general. For that reason, most introductory courses in the Earth Sciences include a module intended to teach students how to read and interpret topographic maps. Numerous activities are available through science education Web sites (such as the Science and Education Resource Center at Carleton College [SERC]) that promote building a connection between two-dimensional (2D) paper maps and three-dimensional (3D) visualizations of the map features. Some paper-based examples include constructing a topographic profile of a pyramid (SERC, 2017a), using geographic information systems (GIS) and digital elevation models (SERC, 2017b), and building contours with playdough (SERC, 2017c). Many students, however, struggle with visualizing data presented

in this static manner (Chang et al., 1985; Ishikawa and Kastens, 2005; Rapp et al., 2007).

While topographic maps may be an abstract concept to some students, experience playing in a sandbox is far more common. In a sandbox, students can intuitively create hills, valleys, craters, etc., and then change those landforms. The KeckCAVES consortium at University of California–Davis has developed a process to quantify the landscape in a sandbox, calculate a topographic map, and then project that map back on the sand (i.e., the “augmented reality [AR] sandbox”; Reed et al., 2016; KeckCAVES, 2017; Fig. 1). This turns a static, 2D paper topographic map into an interactive, dynamic 3D tool. As students alter its landscape, the topographic map dynamically adjusts to match the new surface. This allows students to make predictions about the topography (e.g., What direction does water flow?) and then test those predictions instantly and persistently. This is exactly the strategy Newcombe et al. (2015) suggested would be most effective for teaching topographic map reading skills.

The AR sandbox has generated intense enthusiasm at professional geoscience conferences (Giorgis et al., 2016; Rost et al., 2016). Undergraduates enrolled in an introductory geology class for nonmajors overwhelmingly “loved” using the AR sandbox as part of a topographic maps laboratory course (Fig. 2). Woods et al. (2016) noted a similar level of zeal for the AR sandbox in their study. This enthusiasm has resulted in the construction of over 150 AR sandbox installations around the world (Reed et al., 2016). This widespread development is likely because the software on the University of California–Davis KeckCAVES Web site is free, the installation instructions are clear, and there are

Received 22 February 2017; revised 25 May 2017; accepted 26 May 2017; published online 7 August 2017.

<sup>1</sup>Department of Geological Sciences, State University of New York–Geneseo, 1 College Circle, Geneseo, New York 14454, USA, giorgis@geneseo.edu

<sup>2</sup>Computing and Information Technology, State University of New York–Geneseo, 1 College Circle, Geneseo, New York 14454, USA, kma@geneseo.edu

<sup>a</sup>Present address: STEM instructor, North Dakota State University, 404 Ceres, Fargo, North Dakota 58108, USA, nancy.mahlen@ndsu.edu

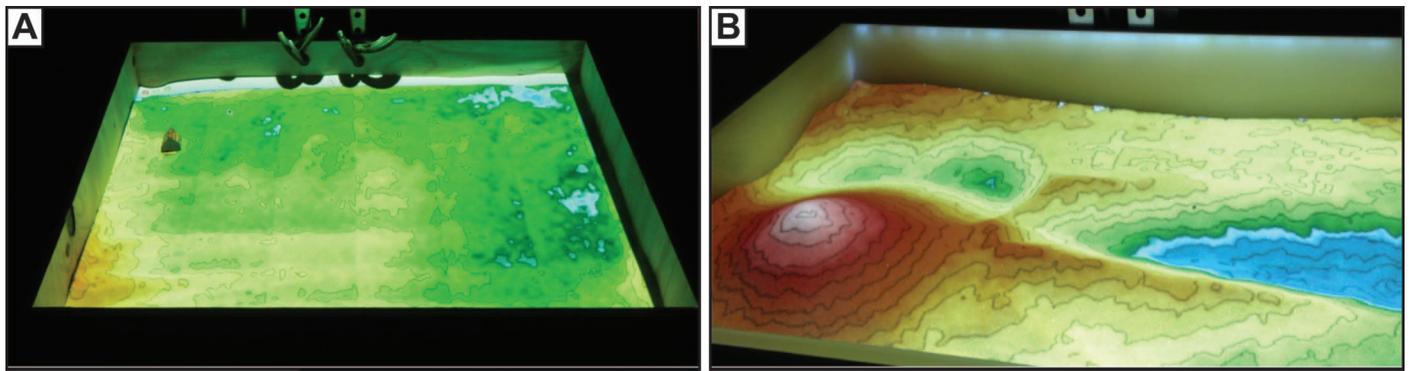


FIGURE 1: View of the augmented reality sandbox. The sand landscape begins as gently sloping surface (A). Subsequently, the sand was sculpted to form a hill, ridge, and a series of depressions (B). The image projected on the sand adjusts in real time to display the new topographic map.

many resources for constructing the sandbox (Ryker et al., 2016a; Woods et al., 2016; KeckCAVES, 2017).

The popularity of the AR sandbox implies that its use in geoscience classes should be rigorously evaluated to determine the best pedagogical approach. Ryker et al. (2016a) suggested a number of critical research objectives that should be addressed, including the following: (1) Is structured questioning effective for learning about topographic maps when students are using the sandbox? (2)

What are the characteristics of the students who get the most out of the sandbox? In this contribution, we sought to address these two questions. Specifically, the experiment presented here attempted to test the hypothesis that a structured, instructor-guided AR sandbox exercise is a more effective tool for teaching topographic maps than the traditional, paper-based approach alone. Previous studies have identified two groups that could potentially benefit most from the AR sandbox: (1) students with no experience reading maps, and (2) students with lower 3D visualization skills. Students with prior experience reading maps tend to do better on topographic map assessments (Chang et al., 1985; Rapp et al., 2007). Reading topographic maps requires 3D visualization skills; therefore, students with lower 3D visualization skills tend to struggle more with topographic map reading tasks. Female students on average tend to do slightly worse than male students on some 3D visualization tasks (Linn and Petersen, 1985), although this is not always the case with the specific task of reading topographic maps (Rapp et al., 2007). During the development of their topographic maps assessment, Newcombe et al. (2015), however, found that the female students as a group did not score as well as their male counterparts. Therefore, in addition to our broad hypothesis of testing the efficacy of the AR sandbox as a teaching tool, we sought to determine if would be particularly helpful to novice map readers, female students, and/or students with lower 3D visualization skills.

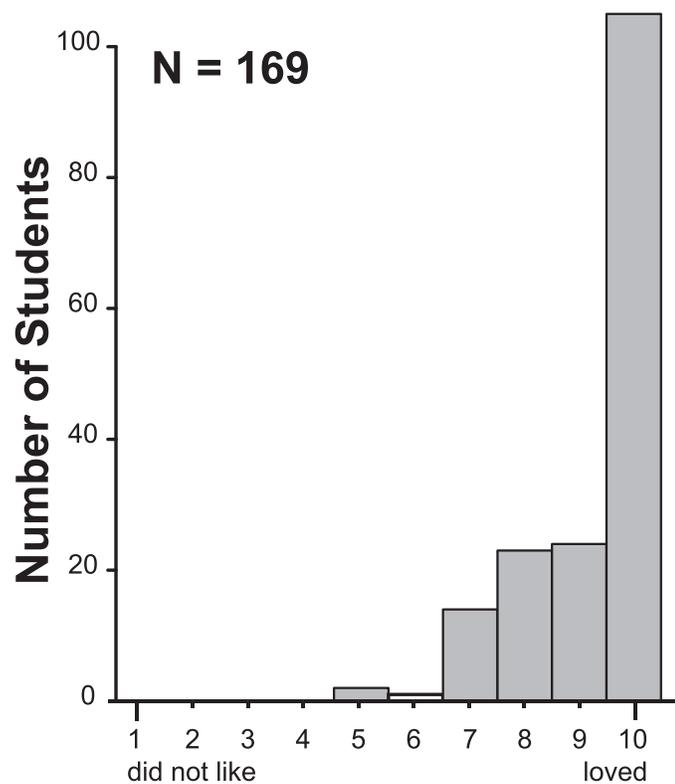


FIGURE 2: Histogram of student responses when prompted to rate their overall experience using the AR sandbox on a scale of 1 (“I did not like using the AR sandbox”) to 10 (“I loved using the AR sandbox”). Note these data were collected in the same course but during a different semester (fall 2017) than the remainder of the data presented here.

## METHODOLOGY

### Study Population and Design

The study took place during the spring 2016 semester at an undergraduate, liberal arts college populated by approximately 5,000 students mostly ranging from 18 to 21 years old with an average SAT score of 1,275 for students enrolled during the study period. Students who participated in this study self-reported a mean SAT score of 1,263 (Fig. 3). Approximately 15%–20% of the student body during the 2015–2016 academic year self-identified as an ethnic minority. The student body for the college as a whole was 58% female and 42% male, and the study group closely reflected this distribution ( $N = 176$ ; 60% female, 40% male). This study took place within an introductory geology laboratory course for nonmajors that satisfies the natural

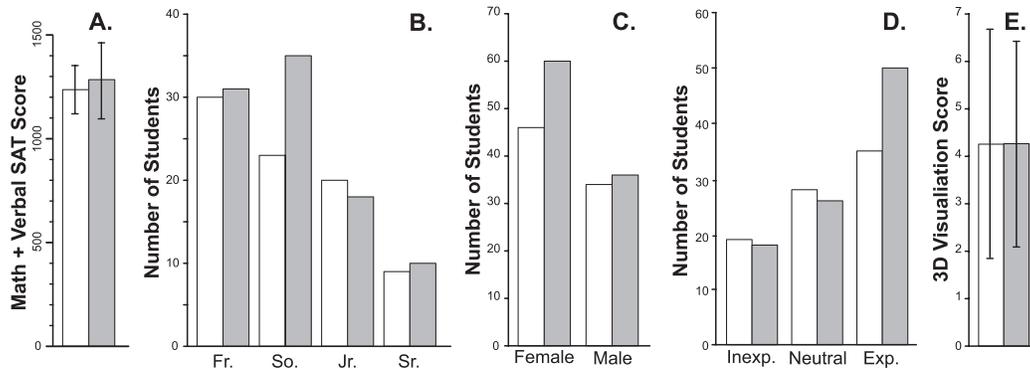


FIGURE 3: Population statistics for the control (white) versus experimental (gray) groups. (A) SAT scores. (B) Academic class (Fr. = freshman, So. = sophomore, Jr. = junior, Sr. = senior). (C) Self-identified gender. (D) Self-assessed prior experience using topographic maps (Inexp = inexperienced, neutral, Exp. = experienced). (E). Score out of 10 on the mental rotation puzzles portion of the assessment (see Fig. 4). Uncertainties on the SAT and 3D visualization scores are one standard deviation. Note that SAT scores were self-reported. Some students did not recall their SAT scores or took the ACT; therefore, the mean score reported is based on 120 out of the 176 students that participated in the study.

science general education requirement; therefore, two thirds of students were in their first or second year of college (34% freshman, 33% sophomores, 22% juniors, 11% seniors).

This study followed a quasi-experimental design (Trochim et al., 2015), and students were enrolled in 10 laboratory course sections. At five different times during the week, two laboratory sections met simultaneously in adjacent rooms. The AR sandbox exercise was introduced into one of these two rooms. At the beginning of laboratory course, students from both laboratory classes were pooled into one group ( $N = 48$ ) and randomly selected to either the control laboratory section (no AR sandbox;  $N = 24$ ) or experimental laboratory section (AR sandbox;  $N = 24$ ). Students in both the experimental and control groups completed the same paper-based topographic maps laboratory course. Those students in the experimental group also completed the AR sandbox exercise described below. Since students in the control group did not complete the AR sandbox exercise, they finished the laboratory class 10–15 min earlier than those in the experimental group. One week after completion of the laboratory, students' topographic map reading skills were assessed.

### Augmented Reality Sandbox Exercise

The AR sandbox exercise consisted of a set of nine questions that students worked to answer under the direction of an undergraduate teaching assistant or a laboratory instructor (see Supplement A for a full copy of the exercise, available in the online journal and at <<http://dx.doi.org/10.5408/17-255s1>>). The exercise aimed to have the students “discover” the following concepts by interacting with the AR sandbox:

1. the relationships among color and elevation, contour lines, and contour intervals,
2. the difference between hills and depressions in both map and profile view,
3. how variations in gradient are portrayed on maps, and

4. how surface water flows and pools in relation to patterns of topographic contours.

The paper-based topographic maps exercise engaged these same topics (contour lines, landform identification, gradient, flow direction of surface water, and topographic profiles); however, other parts of the paper-based laboratory were not incorporated into the AR sandbox laboratory (latitude, longitude, fractional scale, and graphical scales).

Laboratory periods lasted for 2 h, and an individual laboratory had up to 24 students; therefore, ideally, six groups of four students were able to use the AR sandbox for 20 min blocks of time. In practice, group size ranged from four to six, and each group spent 15–20 min with the sandbox. Undergraduate teaching assistants and faculty instructors took part in an hour-long training session to learn how to operate the AR sandbox and how to lead the students through the exercise. A teaching assistant and/or faculty instructor were present at all times during the laboratory to address technical difficulties and guide the students through the exercise at a pace sufficiently quick to finish the assignment in the allotted time.

### Assessment

Students were given 1 week following completion of the topographic maps laboratory to complete the assessment online through Google Forms outside of class (see Supplement B for a full copy of the assessment, available in the online journal and at <<http://dx.doi.org/10.5408/17-255s2>>). All students were offered extra credit for completion of the assessment, but they were not graded for their responses. Of 240 possible respondents, 176 (73%) took the assessment, which consisted of two parts. First, students were asked to complete mental rotation puzzles from the Purdue Visualization of Rotations Test (PVRT; Guay, 1976). The full PVRT contains 30 questions; however, in the interest of time, we used a subset of 10 questions (Fig. 4) following Titus and Horsman (2009). While the PVRT does not completely capture the full spectrum of 3D visualization skills, we chose the instrument for two reasons: (1) It measures a facet of spatial visualization that is not explicitly developed by either

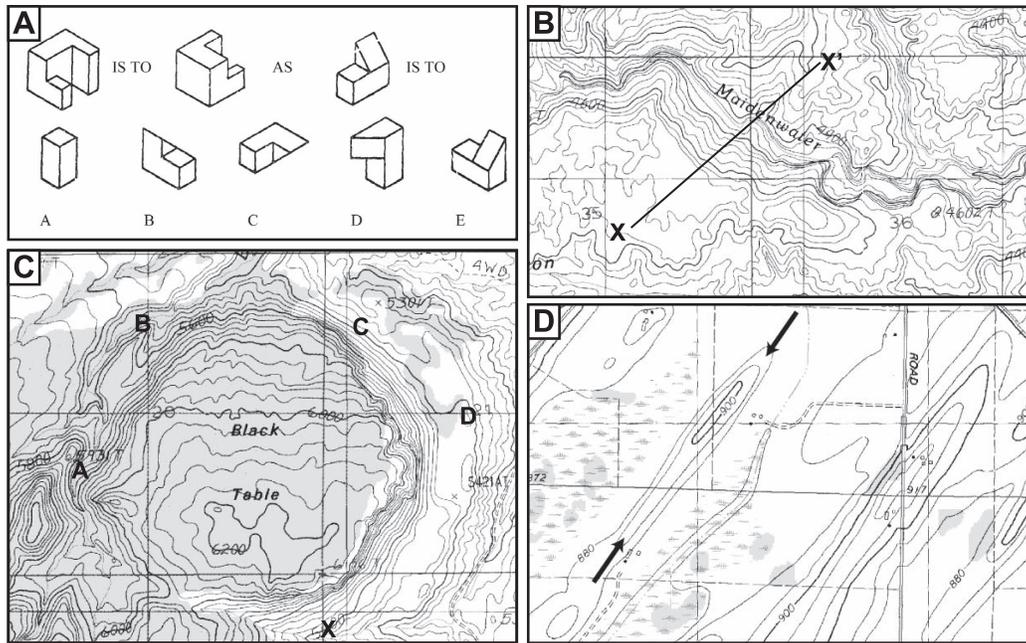


FIGURE 4: Representative questions from the postintervention assessment (see Supplement A, available in the online journal and at <http://dx.doi.org/10.5408/17-255s1>). (A) 3D visualization skills tested using a set of 10 mental rotation puzzles from the Purdue Visualization of Rotations Test (PVRT). (B) Given a choice of four possible topographic profiles, choose the one that best matches X-X'. (C) Imagine you are standing at position X, and there are campfires at positions A, B, C, and D. Determine which campfire you can see. (D) Identify the noted landform by choosing between the options of hill, ridge, valley, or depression.

the paper-based or AR sandbox laboratories, so it should provide some estimate of the students' 3D thinking skills prior to completing this laboratory class. (2) It is well established and quick to administer. Correct answers earned one point, and unanswered questions received a score of zero. Second, students answered three different types of topographic map questions (Fig. 4):

1. Identify a landform.
2. Choose the correct topographic profile.
3. Make a line-of-sight judgment from a given location (i.e., a campfire exercise).

The assessment consisted of five of each type of question, where correct answers earned one point, and unanswered questions or incorrect answers received a score of zero.

Evaluation of the internal reliability of the assessment using the "alpha" function in the Psychology package in R suggested it is a consistent measure of topographic map reading ability. Overall, the assessment yielded an acceptable Cronbach's alpha of  $0.66 \pm 0.7$  ( $2\sigma$ ). Item analysis indicated dropping any one question yielded equal or slightly lower Cronbach alpha values in the range of 0.63–0.66; therefore, we retained all 15 questions in our analysis.

### Data Analysis

The data collected were subjected to a standard statistical analysis in Rstudio (version 0.99.486) to address the following questions: (1) Is there any difference in the scores on the topographic map assessment between groups that engaged in a short (20 min) interaction with the AR sandbox versus those that did not? (2) Is this relation a function of gender, prior

knowledge, or 3D visualization skills? All subgroups analyzed were tested for a normal distribution using the Shapiro-Wilk test. The Shapiro-Wilk test evaluates the null hypothesis that a population comes from a normal distribution. A score  $<0.05$  indicates that the null hypothesis should be rejected, and the population likely comes from a nonnormal distribution. Means and standard deviations were calculated for the larger subgroups (e.g., gender or experimental/control), and differences between those means were evaluated with a simple unpaired *t*-test and the Mann-Wilcox test, which does not assume a normal distribution. *P*-values  $<0.05$  were considered to indicate statistically significant differences between subgroups. Where statistically significant differences were identified, Cohen's *d* effect size was calculated. A two-factor analysis of variance (ANOVA) was then used to compare between the control versus experimental group and various subgroups. The raw data are presented in Supplement C (available in the online journal and at <http://dx.doi.org/10.5408/17-255s3>).

### RESULTS

The control and experimental groups yielded mean and median scores of 8 out of 15 on the topographic map assessment (Fig. 5 and Table I). The experimental group had a broader distribution of scores, but there was no difference in their central values. Subdivision by gender (Fig. 6 and Table I) suggested male students consistently scored higher than female students, regardless of whether or not they used the AR sandbox. Analysis of both gender and control/experiment, however, revealed no significant interaction between these factors (Fig. 6 and Table II). Those students

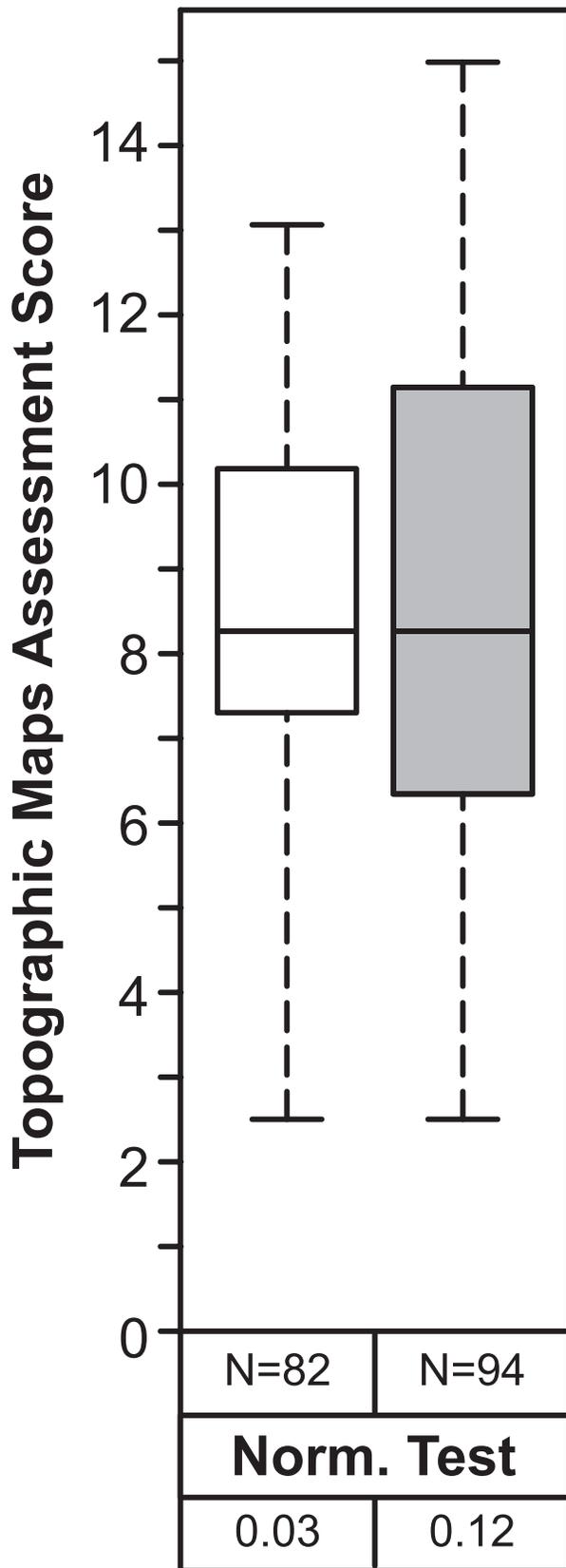


FIGURE 5: Box plot of control group (white) and experimental group (gray) scores on the topographic map assessment test. Bold line marks the median score, box encloses the middle 50% of the data, and the whiskers mark the highest and lowest values.

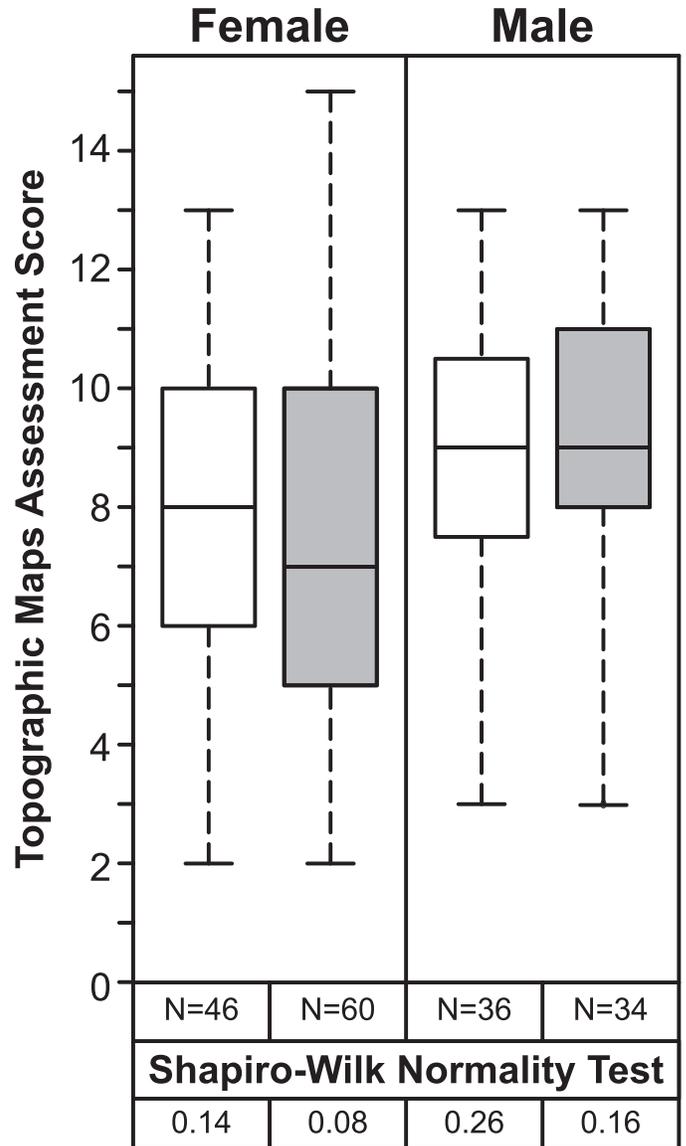


FIGURE 6: Box plot of control group (white) and experimental group (gray) scores on the topographic map assessment test subdivided by gender. Bold line marks the median score, box encloses the middle 50% of the data, and the whiskers mark the highest and lowest values.

with higher self-assessed prior knowledge of topographic maps prior to the class scored better on the assessment (Table I); however, there was no interaction between this factor and the experiment/control groups (Fig. 7 and Table II). Last, those students who scored higher scores in 3D visualization ability as measured by the PVRT also scored higher on the assessment (Table I and Fig. 8). Examination of visualization ability and the control/experimental groups again yielded no statistically significant interaction (Fig. 8 and Table II).

## DISCUSSION

The data show some clear patterns that have been identified in previous studies. On average, female students

TABLE I: Results subdivided by subgroup.<sup>1</sup>

	N	Shapiro	Mean	SD	T-test	Mann-Wilcox	d
Experimental	94	0.12	8.3	3.0	0.83	0.8	–
Control	82	0.03	8.4	2.8			
Male	70	0.03	9.0	2.4	0.01	<0.01	0.37
Female	106	0.02	7.9	3.1			
Experienced	85	<0.01	9.2	2.7	<0.01	<0.01	0.84
Inexperienced	37	0.19	6.9	2.6			
High 3D vis. score	78	0.03	9.7	2.6	<0.01	<0.01	0.89
Low 3D vis. score	98	0.08	7.3	2.7			

<sup>1</sup>N = number of individuals in the population; Shapiro = Shapiro-Wilk test for normality; Mean = mean value; SD = one standard deviation about the mean; T-test = *p*-value calculated from the *t*-test; Mann-Wilcox = *p*-value calculated from the Mann-Wilcox test; *d* = Cohen's *d* value.

tended to do slightly worse than male students on the topographic maps assessment (Fig. 6). Previous studies have found that female students score marginally lower on some 3D visualization tasks (Linn and Petersen, 1985). Similar to Hegarty et al. (2006) and Ishikawa and Kastens (2005), there

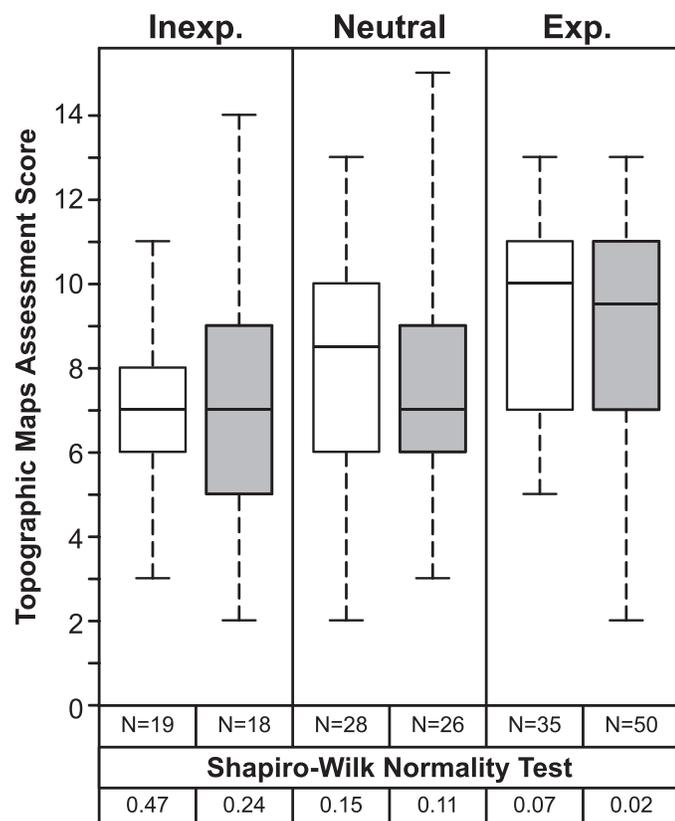


FIGURE 7: Box plot of control group (white) and experimental group (gray) scores on the topographic map assessment test subdivided by self-assessed knowledge of topographic maps prior to class. Students were asked to rate the following statement: “I was comfortable using topographic maps prior to taking this class” on a scale from “disagree” to “neutral” to “agree.” Inexp. = inexperienced = disagree, Neutral = neutral, Exp. = experienced = agree. Bold line marks the median score, box encloses the middle 50% of the data, and the whiskers mark the highest and lowest values.

was a large amount of variance in the data, with many female students scoring higher than their male counterparts. On average, however, the data indicate that female students tended to have lower scores than male students, similar to the results of Newcombe et al. (2015), but this was a relatively minor effect ( $d = 0.37$ ; Table I). The data also support the conclusion that prior map reading experience is a significant factor ( $d = 0.84$ ; Table I) in determining topographic map reading scores (Fig. 8; Chang et al., 1985; Rapp et al., 2007). Last, 3D visualization skill, as measured by the PVRT, also appeared to have a large effect ( $d = 0.89$ ; Table I) on the topographic map assessment outcomes (Fig. 8).

Despite these encouraging signs, and a great deal of enthusiasm for the AR sandbox among both educators and students (Reed et al., 2016; Woods et al., 2016; Fig. 2), the data clearly indicate that topographic map reading skills are not higher for those students who used the AR sandbox in our experiment (Fig. 5 and Table 1). This disappointing result does not change even if subgroups with little to no experience using topographic maps (Fig. 7) or those with lower 3D visualization skills (Fig. 8) are examined. These results could be due to differences in the randomly determined experimental versus control populations (Fig. 3). The two groups were similar with respect to SAT scores and 3D visualization scores, but other differences did exist. The experimental group was larger and had more second-year (sophomore-level) students. The data suggest, however, that academic class was not a significant factor in determining topographic map reading ability. There were more female students in the experimental group (Fig. 3), and our analysis does suggest that gender played a significant role in determining assessment scores (Fig. 6). Direct comparison of the female (or male) students in the control versus experimental groups, however, indicated that gender cannot be used to explain the lack of gains in the

TABLE II: Two factor analysis of variance comparison of experimental versus control groups within subgroups.<sup>1</sup>

	Df	SS	MS	F	<i>p</i>
Gender	1	4.8	4.8	0.6	0.45
Prior knowledge	2	1.7	0.9	0.1	0.89
Visualization score	1	3.2	3.2	0.5	0.50

<sup>1</sup>Df = degrees of freedom; SS = sum of the squares; MS = mean of the squares; F = *F*-value; *p* = *p*-value.

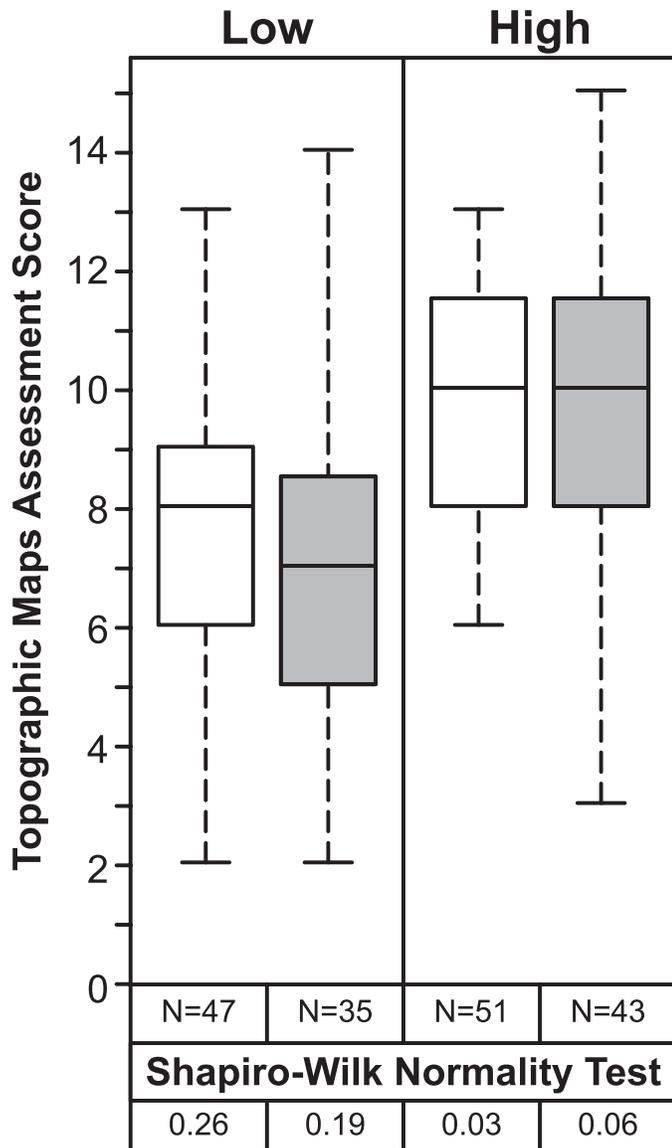


FIGURE 8: Box plot of control group (white) and experimental group (gray) scores on the topographic map assessment test subdivided by 3D visualization score. Low = scored at or below the class median of 4. High = scored above the class median. Bold line marks the median score, box encloses the middle 50% of the data, and the whiskers mark the highest and lowest values.

experimental group. Last, the experimental group did contain a larger proportion of students who self-identified as being comfortable with topographic maps prior to the laboratory class. If anything, this variation should have improved the scores of the experimental group, which it clearly did not (Fig. 5).

**Limitations**

The results presented here are clearly limited by the size of the study population, particularly once it was divided into subgroups. The smallest group, inexperienced topographic map users in the experimental group, had only 18 students (Fig. 7). Our hypothesis anticipated that those students with

lower 3D visualization skills would benefit most from use of the AR sandbox, but that group only had 35 individuals (Fig. 8). Replication of the same experiment on a much larger group could tease out a statistically significant gain from the experimental group; however, our results suggest that any such gains are likely to be modest at best.

The nonnormal distribution of the three subgroups (control group; experimental with prior experience reading topographic maps; control group with high 3D visualization skills) is probably a result of their sample sizes ( $N = 82, 50, 51$ , respectively). A larger study population might result in normal distributions for these subgroups, but their non-normal distribution does not affect the results presented here. Although the ANOVA analysis applied here does assume a normal distribution, violation of that assumption could lead to a false positive. Our ANOVA results (Table II), unfortunately, suggest no statistically significant gains; therefore, we are not in danger of identifying a false positive.

Last, the results of this analysis are limited by the assessment itself. Following Titus and Horsman (2009), we used only a subset of the rotation puzzles from the PVRT. We limited it to 10 puzzles in an effort to limit the time necessary to take the assessment to about 20 min. Students took the assessment outside of class; therefore, unlike the original PVRT, the test was untimed. Allowing the students to take an unlimited amount of time could theoretically result in higher 3D visualization scores. Anecdotally, however, students reported taking approximately 20 min to complete the entire survey. This seems reasonable and is likely true for the study population as a whole, because they took the assessment on their own and were assured that their score on the assignment was based solely on completion.

Beyond the rotation puzzles, the assessment could be flawed because it was not explicitly aligned with the laboratory materials. Both the paper-based and AR sandbox portions of the laboratory required students to engage in questions about landforms and topographic profiles. Students did not, however, explicitly answer the same kinds of questions on the assessment as in laboratory. For example, they were required to draw a topographic profile during the paper-based portion of the laboratory, but the assessment asked them to identify the correct topographic profile from a list of possible choices. Similarly, students were asked to visualize the landscape from topographic maps (AR and paper) to answer questions about gradient, but they were never explicitly asked to solve a line-of-sight campfire exercise in the laboratory exercise. Regardless, while an explicit alignment between the exercises and assessment would likely increase scores, if the AR sandbox is an effective means of teaching students to read topography, then student scores should reflect this on the assessment presented here.

**Implications**

Students clearly enjoyed using the sandbox (Fig. 2; e.g., Reed et al., 2016; Woods et al., 2016). Van der Hoeven Kraft et al. (2011) suggested that students who experience positive emotions, such as joy or success, and fewer negative emotions, such as boredom or anxiety, are more likely to engage, persist, and learn. Anecdotally, students seemed to be more open to asking instructors and teaching assistants questions about topographic maps using the AR sandbox as a prop. With the goal of using the positive emotions

generated by the AR sandbox in the most effective way possible, our results suggest three interpretations: (1) The assessment was not well aligned with the AR sandbox exercise; (2) the AR sandbox is not an effective teaching tool; or (3) the AR sandbox exercise described here is an ineffective approach. Given how recently this technology became available, and the overwhelming enthusiasm of students for it, we strongly favor the third option. The challenge moving forward is to develop a pedagogical approach that makes use of the unique advantages of the AR sandbox, fits the constraints of a large-enrollment introductory-level course, and effectively improves students' topographic map reading skills. We suggest the following avenues to pursue.

- Allow the students to have more time working with the AR sandbox during the laboratory class. In hindsight, it may have been naïve to think that a short, 20 min exercise would lead to statistically significant gains in map reading skills. Time, however, is limited in large-enrollment classes. Offering students more than 20 min to work with the box in a laboratory period requires either larger groups (6–8 students) or construction of more sandboxes.
- Allow students to use the AR sandbox repeatedly throughout the semester, as described by Woods *et al.* (2016). Although they were working with a small group ( $N = 12$ ), their approach of integrating the AR sandbox into seven different laboratories should be scalable to larger classes. Repeated, short interactions as part of laboratories on floods, shorelines, groundwater, etc., could be a more effective means of making use of the AR sandbox.
- Allow the students to have more freedom to interact with the AR sandbox on their own terms rather than having an instructor-guided exercise (i.e., the “free play” of Ryker *et al.*, 2016b). The laboratory described here was developed with the intention that an undergraduate teaching assistant or a laboratory instructor would guide the students through it. This was primarily to ensure that a large number of students would all have a chance to use the sandbox. The real draw of the AR sandbox, however, is as a device to facilitate student-driven inquiry. The instructor-guided exercise described here severely limited students' freedom in this capacity. Opening the exercise to be entirely free play, however, has also been found to be ineffective (Ryker *et al.*, 2016b).

## CONCLUSIONS

The AR sandbox is an effective means for generating a large amount of enthusiasm from students with respect to understanding and using topographic maps (Fig. 2). While enthusiasm is an important part of engaging students (Van der Hoeven Kraft *et al.*, 2011), we found that a short (20 min) instructor-driven AR sandbox exercise was not sufficient to produce statistically significant gains in topographic map reading ability. Future studies should investigate the following questions: (1) How much time do students need with the AR sandbox for it to be an effective tool? (2) Can that time be parceled over several weeks and still be

effective? (3) What is the most effective balance between student-led free play and instructor guidance?

## Acknowledgments

Thanks go to Oliver Kreylos and the KeckCAVES group at University of California–Davis for making the augmented reality (AR) sandbox software freely available and easy to use. The AR sandboxes at State University of New York–Geneseo were constructed with an Innovative Instruction Technology Grant (IITG) award from the State University of New York. Kris Anderson, Diane Lounsbury, Joe Dolce, and Clint Cross were instrumental in their construction at Geneseo. Talor Walsh, Margaret Reitz, and Nicolas Warner assisted in the collection of assessment data. Gregg Hartvigsen and Janelle Goeke provided much-needed assistance with R and statistics in general. The AR sandbox research group (Rachel Atkins, Christine Clark, Nicole LaDue, Karen McNeal, Katherine Ryker, and Shelly Whitmeyer) has been very supportive and welcoming to those new to education research. This contribution was greatly improved by insightful critiques from two anonymous reviewers and an associate editor.

## REFERENCES

- Chang, K.-T., Antes, J., and Lenzen, T. 1985. The effect of experience on reading topographic relief information: Analyses of performance and eye movements. *Cartographic Journal*, 22:88–94.
- Giorgis, S., Mahlen, N., and Anne, K. 2016. Integration of an augmented reality sandbox into a large enrollment introductory geoscience course for non-majors. *Geological Society of America Abstracts with Programs*, 48(7):284–12.
- Guay, R. 1976. Purdue Spatial Visualization Test. West Lafayette, Indiana, Purdue Research Foundation.
- Hegarty, M., Montello, D.R., Richarson, A.E., Ishikawa, T., and Lovelace, K. 2006. Spatial abilities at different scales: Individual differences in aptitude-test performance and spatial-layout learning. *Intelligence*, 34:151–176.
- Ishikawa, T., and Kastens, K.A. 2005. Why some students have trouble with maps and other spatial representations. *Journal of Geoscience Education*, 53(2):184–197.
- KeckCAVES. 2017. KeckCAVES. Available at <https://arsandbox.ucdavis.edu/> (accessed 12 January 2017).
- Linn, M.C., and Petersen, A.C. 1985. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56:1479–1498.
- Newcombe, N., Weisberg, S.M., Atit, K., Jacovina, M.E., Ormand, C.J., and Shipley, T.F. 2015. The lay of the land: Sensing and representing topography. *The Baltic International Yearbook of Cognition, Logic and Communication*, 10:1–57.
- Rapp, D.N., Culpepper, S.A., Kirby, K., and Morin, P. 2007. Fostering students' comprehension of topographic maps. *Journal of Geoscience Education*, 55(1):5–16.
- Reed, S., His, S., Kreylos, O., Yikilmaz, M.B., Kellogg, L.H., Schladow, S.G., Segale, H., and Chan, L. 2016. Augmented reality turns a sandbox into a geoscience lesson. *Eos*, 97. Available at <https://doi.org/10.1029/2016EO056135> (accessed 25 June 2017).
- Rost, R., Giorgis, S., and Anne, K. 2016. Visualizing topography and dipping strata using an augmented reality sandbox. *Geological Society of America Abstracts with Programs*, 48(7):67–1.
- Ryker, K., McNeal, K., Atkins, R., LaDue, N., and Clark, C. 2016a. Facilitating spatial thinking with augmented reality sandboxes. *In the Trenches*, 6(4):1–4.
- Ryker, K., McNeal, Karen, S., Atkins, R., LaDue, N., and Clark, C. 2016b. The impact of an AR sandbox on map-reading skills

- using a “free-play” experience. *Geological Society of America Abstracts with Programs*, 48(7):284–13.
- Science Education Resource Center (SERC). 2017a. Topographic maps. Available at <http://serc.carleton.edu/NAGTWorkshops/careerdev/AcademicCareer2010/activities/40981.html> (accessed 25 May 2017).
- Science Education Resource Center (SERC). 2017b. Landscape interpretation and visualization using a GIS. Available at <http://serc.carleton.edu/NAGTWorkshops/gis/activities2/47999.html> (accessed 25 May 2017).
- Science Education Resource Center (SERC). 2017c. Contour mapping with playdough. Available at <http://serc.carleton.edu/NAGTWorkshops/intro/activities/67200.html> (accessed 25 May 2017).
- Titus, S., and Horsman, E. 2009. Characterizing and improving spatial visualization skills. *Journal of Geoscience Education*, 57:242–254.
- Trochim, W., Donnelly, J.P., and Arora, A. 2015. *Research methods: The essential knowledge base*, 2nd ed. Boston, MA: Cengage Learning. p. 448.
- van der Hoeven Kraft, K.J., Shogi, L., Husman, J., Semken, S., and Fuhrman, M. 2011. Engaging students to learn through the affective domain: A new framework for teaching in the geosciences. *Journal of Geoscience Education*, 59:71–84.
- Woods, T.L., Reed, S., His, S., Woods, J.A., and Woods, M.R. 2016. Pilot study using the augmented reality sandbox to teach topographic maps and surficial processes in introductory geology labs. *Journal of Geoscience Education*, 64:199–214.