

ISSN: 2149-214X

**Journal of Education in Science,
Environment and Health**

www.jeseh.net

**Evaluating Middle School Students'
Spatial-Scientific Performance within
Earth/Space Astronomy in Terms of
Gender and Race/Ethnicity**

Jennifer Wilhelm, Michael D. Toland, Merryn Cole
University of Kentucky

To cite this article:

Wilhelm, J., Toland, M.D., & Cole, M. (2017). 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 (JESEH)*, 3(1), 40-51.

This article may be used for research, teaching, and private study purposes.

Any substantial or systematic reproduction, redistribution, reselling, loan, sub-licensing, systematic supply, or distribution in any form to anyone is expressly forbidden.

Authors alone are responsible for the contents of their articles. The journal owns the copyright of the articles.

The publisher shall not be liable for any loss, actions, claims, proceedings, demand, or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of the research material.

Evaluating Middle School Students' Spatial-Scientific Performance within Earth/Space Astronomy in Terms of Gender and Race/Ethnicity

Jennifer Wilhelm, Michael Toland, Merryn Cole

Article Info

Article History

Received:
30 August 2016

Accepted:
21 November 2016

Keywords

Middle school education
Geometrical and spatial
thinking
Gender and race/ethnicity
Astronomy education

Abstract

Differences were examined between groups of sixth grade students' spatial-scientific development pre/post implementation of an Earth/Space unit. Treatment teachers employed a spatially-integrated Earth/Space curriculum, while control teachers implemented their Business as Usual (BAU) Earth/Space units. A multi-level modeling approach was used in a hierarchical manner to evaluate student performance on the Purdue Spatial Visualization: Rotation test (PSVT-Rot) and on the Lunar Phases Concept Inventory (which included four spatial domains), while controlling for two variables (gender and race/ethnicity) at the student level and one variable (teaching experience) at the teacher level. Results showed Treatment girls achieved higher LPCI Periodic Patterns (PP) spatial domain post-scores than girls in the BAU group. A gender gap was also observed (in favor of boys) within the BAU group for PP domain post-scores, while no gap was shown within the Treatment group. In addition, results for PP suggest Students of Color tended to have lower PP scores than White students (Effect Size = .29), and that higher pretest PP scores tended to lead to higher posttest PP scores, after adjusting for other student and teacher characteristics. The only statistically significant predictor of the PSVT-Rot posttest scores were scores on the respective pretest.

Introduction

Research has shown gender differences on students' spatial understandings in favor of males, particularly for spatial visualization and mental rotation (Kaufman, 2007). Linn and Petersen (1985) determined that males outperformed females at all age levels on mental rotation tasks. Numerous studies have also shown a substantial gap in mathematical achievement between Black and White students (Lee & Wong, 2004; Reyes & Stanic, 1988) which is further intensified among Hispanic and White students (Lubienski, 2002). However, research focusing on spatial reasoning and visualization among students of color is underdeveloped.

Studies have shown relationships between students' spatial abilities and their understanding of scientific phenomena (Black, 2005), especially in the area of Earth/Space science. Rudmann (2002) found students' inclination to learn scientific explanations for the cause of the seasons was restricted by their spatial aptitude. Similarly, Wellner (1995) reported students were more likely to describe a correct cause of lunar phases when they had a strong spatial sense. Other studies claimed understanding celestial motion demands the skill of moving between frames of references (Plummer, Wasko, & Slagle, 2011; Plummer, 2014).

This study builds on previous research (Wilhelm, 2009; Wilhelm, Jackson, Sullivan, & Wilhelm, 2013) and examined differences between groups of students' spatial-scientific development from pre to post implementation of an Earth/Space unit. Wilhelm's (2009; 2013) prior research found that students who participated in spatial experiences within an Earth/Space unit made significant gains on lunar-related concepts. Females tended to lead in significant content development concerning geometric spatial test items. One group of students experienced a purposeful, spatially-integrated Earth/Space unit while the other experienced their Business as Usual (BAU) Earth/Space unit. Differences in spatial-scientific understanding by gender groups and racial/ethnic groups were also investigated within and between BAU and Treatment groups.

The Argument for Developing Spatial Skills in STEM

Research articles in the 1990s have reported a link between students' abilities to report the correct cause of lunar phases with their projective spatial skills (Reynolds, 1990; Wellner, 1995; Bishop, 1996). Other research

correlated students’ success on science assessments with their spatial ability (Hake, 2002; Sorby, 2006). In addition to this, studies have shown students’ improvement in the areas of Chemistry, Geoscience, Physics, and Calculus after they received spatial training (Sanchez, 2012; Miller & Halpern, 2014; Sorby, Casey, Veurink, & Dulaney, 2013).

Recent research has claimed that well-developed spatial thinking is necessary for understanding many astronomical concepts such as celestial motions and lunar phases (Plummer, 2014; Wilhelm, 2009; Wilhelm et al., 2013). Table 1 outlines claims made over the last 25 years linking spatial ability to scientific understanding especially in the area of astronomy.

Table 1. Research that links spatial ability to scientific understanding

| Author(s)/Year | Findings |
|---|---|
| Reynolds (1990); Wellner (1995); Bishop (1996) | Students were more likely to report a correct cause of lunar phases when they had strong projective spatial skills. |
| Pribyl & Bodner (1987); Hake (2002); Sanchez (2012); Sorby (2006); Sorby, Casey, Veurink, & Dulaney (2013); Miller & Halpern (2014) | Students’ scores and success on science assessments in the areas of Chemistry, Physics, Geoscience, and Calculus were correlated to their spatial ability. |
| Black (2005); Plummer (2009, 2014); Plummer, Wasko, & Slagle (2011); Wilhelm (2009); Wilhelm, Jackson, Sullivan, & Wilhelm (2013) | Well-developed spatial thinking is necessary for understanding astronomical concepts such as celestial motions and lunar phases. Spatial thinking includes: Mental rotations, Perspective, Geometric Spatial Visualization, Spatial Projection, Periodic Patterns, and Cardinal Directions. |

Black (2005) “hypothesized that mental rotation is the most important in understanding Earth science concepts...humans are handicapped by their single vantage point from Earth of the moving bodies in outer space” (p. 403). Plummer, Wasko, and Slagle (2011) argued that children have difficulties learning to explain daily celestial motion since it requires an understanding across moving frames of references. A mismatch between students’ description of apparent motion and their explanation may be due to limited ability to use the necessary spatial abilities to make the logical connection. Instruction may have differentially supported high spatial ability students over low spatial ability (Plummer et al., 2011, p. 1986).

We contend that one cannot understand many astronomical concepts without a developed understanding of four specific spatial domains defined as follows (Wilhelm et al., 2013): a) *Geometric Spatial Visualization (GSV)*-Visualizing the geometric spatial features of a system as it appears above, below, and within the system’s plane; b) *Spatial Projection (SP)*-Mentally projecting to a different location on an object and visualizing from that global perspective; c) *Cardinal Directions (CD)*-Distinguishing directions (N, S, E, W) in order to document an object’s vector position in space; and d) *Periodic Patterns (PP)*-Recognizing occurrences at regular intervals of time and/or space.

All four of these domains are driven by the facility to mentally rotate objects over time when posed within an astronomical context. For example, the GSV domain concerns visualizing and manipulating the Earth/Moon/Sun system; the SP domain involves mentally maneuvering the sky throughout a day’s viewing from various Earthly perspectives; CD domain includes mapping, recording, and predicting lunar positions over time; and PP domain involves noticing the repeated nature of celestial orbital motions.

Gender and Racial Gaps in Spatial Ability

The literature has shown gender differences on students’ spatial understandings in favor of males (Kaufman, 2007; Kerns & Berenbaum, 1991; Silverman, Choi, & Peters, 2007; Ansell & Doerr, 2000). Results from the 1996 National Assessment of Educational Progress (NAEP) for United States (US) grade 4 and grade 12 students showed males having significantly higher scale scores than females in the areas of Measurement, and

Geometry and Spatial Sense. “An item-level analysis of percent-correct values revealed some historically common, research-based patterns of difference such as males performing better than females on items that required spatial visualization, the use of measurement tools such as rulers, and working with rational numbers” (Ansell & Doerr, 2000, p. 75). McGraw, Lubienski, and Strutchens (2006) analyzed NAEP results from 1990 – 2003 and found that not only was there a gender gap favoring males in the areas of measurement and geometry, but also that this gap was concentrated at the higher end of score distributions and was most consistent with White students.

Males scored significantly higher than females on tests of spatial visualization as well as 3D mental rotation (Kaufman, 2007). Wilhelm (2009) found that pre-teen female students scored significantly lower than pre-teen male students on spatial pre-tests. However, following a spatially-focused intervention that utilized STEM integrated lessons with many situational opportunities to experience 2D and 3D stimuli, females achieved significantly higher gain scores than their male counterparts. The study speculated that the initial sex differences (on pretests) could be explained by the faster maturation (during preteen years) of the male brain’s anatomical regions that handle spatial visual reasoning (Giedd et al. 1999). The implication of the Wilhelm study was that the 2D and 3D instructional intervention allowed females to develop their spatial skills resulting in significant achievement.

In addition to gender differences, research studies have also shown differences in mathematical performance between Black and White students (Lee, 2004; Lee & Wong, 2004; Lubienski, 2002; Reyes & Stanic, 1988) and between Hispanic and White students (Lubienski, 2002). McGraw et al. (2006) analyzed the 2003 NAEP assessment for gender gaps in achievement by race/ethnicity and found “that the differences in scale scores were much greater between racial/ethnic groups than between males and females within the same racial/ethnic group” (p. 140). McGraw et al. (2006) argued that one must examine gender and race/ethnicity as well as social economic status together; otherwise differences within groups will not be documented and interactions will not be found. Despite calls for further research in this area, studies exploring gender and racial/ethnic differences in mathematical performance with potential research-based solutions towards closing the achievement gap have been severely limited.

In order to add to the research base on these issues we examined the following questions: *In what ways will students’ curricular and instructional Earth/Space experiences affect their spatial-scientific learning? What, if any, differences in spatial-scientific performance will be observed between gender groups and racial/ethnic groups?*

Methodology

Participants and Instructional Curriculum

Research subjects were sixth-grade students from three US middle schools (Juniper, Butternut, and Willow). Juniper had two Treatment groups ($N = 187$) taught by teachers with 4 and 9 years’ experience. The Juniper BAU group ($N = 58$) was taught by a teacher with 3 years’ experience. Butternut had three Treatment groups ($N = 228$) taught by two first year teachers and one teacher with 11 years of experience. A group of 26 students comprised the Butternut BAU group taught by a teacher with 12 years of experience. Willow had one Treatment group ($N = 53$) taught by a teacher with 13 years’ experience. Table 2 displays the teacher and student characteristics. *Pseudonyms were used for all schools; each school self-selected its BAU and Treatment teachers. This, unfortunately, resulted in small BAU numbers (including no BAU classroom at Willow), which was beyond the researchers’ control.*

All groups studied Earth/Space concepts related to the Solar System. Treatment teachers employed a spatially-oriented, STEM-integrated Earth/Space curriculum while BAU teachers implemented their regular Earth/Space lessons (see Table 3). The spatially-oriented curriculum (Treatment instruction) was designed to: (A) Foster students’ understanding of Earth-Space science concepts and ‘big ideas’ (such as planetary geologic activity and celestial motions and patterns) through the development of innovative projects, lessons, and learning communities; (B) Create experiences for students to do mathematics by challenging them to: i) represent situations graphically and geometrically, ii) observe patterns and functional relationships to make predictions, and iii) develop and employ spatial visualization skills to model phenomena; and (C) Construct opportunities for students to engage in authentic project work, modeling, and data collection and interpretation. The BAU curriculum and instruction tended to utilize videos, simulations, texts, and modeling. Table 3 outlines the time spent on Earth/Space content by each group, the content implemented, and the instructional methods. Juniper

teachers executed their Earth/Space units over a nine-week period while Butternut and Willow teachers implemented theirs in approximately four weeks.

Table 2. Teacher and student characteristics

| | Teachers | |
|--------------------------|-----------------------------|------------------------|
| | Control (BAU) (n = 2) | Treatment (n = 6) |
| Gender | | |
| Male | 0 | 1 |
| Female | 2 | 5 |
| Ethnicity | | |
| Caucasian | 2 | 6 |
| Yrs teaching | | |
| Mean | 7.50 | 6.50 |
| Highest degree earned | | |
| BA, BS | 0 | 3 |
| MA, MS | 2 | 3 |
| | Students | |
| | Control (n = 84) | Treatment (n = 384) |
| Gender | | |
| Boys | 38 | 198 |
| Girls | 46 | 186 |
| Grade | | |
| 6 | 84 | 384 |
| Race/Ethnicity | | |
| Caucasian (Non-Hispanic) | 55 | 244 |
| African American | 6 | 21 |
| Asian American | 3 | 21 |
| Native American | 3 | 10 |
| Hispanic American | 5 | 25 |
| Asian (Not American) | 1 | 21 |
| Other | 11 | 42 |

Research Questions and Measures

Spatial-scientific reasoning was assessed via pre/post content surveys. The research questions that drove this study were: *In what ways will students’ curricular and instructional Earth/Space experiences (Treatment versus BAU) affect their spatial-scientific learning? What, if any, differences in spatial-scientific performance will be observed between gender groups, racial/ethnic groups, and Treatment and BAU groups?*

Due to the small numbers of students comprising groups other than Caucasian, we classified two groups of students: Students of Color (SoC) and White. We acknowledge that analysis at this level has limitations due to the small number of student in these categories. This quasi-experimental study utilized quantitative measures to document students’ understanding before and after implementation. The quantitative data sources were the Lunar Phases Concept Inventory (LPCI, Lindell & Olsen, 2002), a multiple-choice survey which assessed eight science domains and four spatial-mathematics domains (*Periodic Patterns (PP)*, *Geometric Spatial Visualization (GSV)*, *Cardinal Direction (CD)*, *Spatial Projection (SP)*) as shown in Table 4, and the Purdue Spatial Visualization Test: Rot (PSVT-Rot, Bodner & Guay, 1997), a 20-item multiple choice instrument which assessed mental rotation ability.

Table 3. Time spent on Earth/Space content by each group, content implemented, and instructional method used

| Week | Juniper | | | | Butternut and Willow | | | |
|--------|------------------------------------|---|--|---|---|--|--|--|
| | Business as Usual | | Treatment | | Business as Usual | | Treatment | |
| | Lesson | Method | Lesson | Method | Lesson | Method | Lesson | Method |
| Week 1 | Big Bang Theory; Solar System | PPT Modeling Expanding Universe Balloons | <i>Overview of Universe*</i> Why does the Moon appear to change its shape? | “Many Moons” by Thurber, Moon Journaling (5 weeks), Stellarium (planetarium software) | Intro to Solar System | Lecture and note taking | Why does the Moon appear to change its shape? Measuring distance between objects in the sky. Altitude and Azimuth Angles | Moon Journaling (4 weeks) Stellarium (planetarium software) Activity with measurement and graphing |
| Week 2 | Gravity | YouTube video Textbook Centripetal Motion PhET Simulations | How do I measure the distance between objects in the sky? Altitude and Azimuth Angles | Activity with measurement and graphing | Angular measures and measuring the diameter of the Moon; How Far to the Star? Parallax Effect) | Lab work, note taking, and whole class discussion | How can I say where I am on the Earth? Longitude/Latitude <i>Rotation/Revolution and Seasons*</i> | Earth Globe Activity PPT Modeling Activity |
| Week 3 | Stars | Parallax Activity Stellarium | How to say where I am on the Earth. Longitude/Latitude <i>Rotation/Rev.</i> | Earth Globe Activity PPT Modeling Activity | Why is Earth the only possible place for life? Seasons Reasons | Lab work using probeware | What can we learn by examining the Moon’s surface? Scaling Earth/Moon/Mars | Exploration of Lunar Images PPT Scaling Activity using Balloons |
| Week 4 | Planets; Earth (day/night) | Foam ball models. Graphing | What can we learn from the Moon’s surface? | Exploration of Lunar Images | Moon Phases Eclipses; Tides | Oreo Moon Phases; 3D Earth/Moon/Sun Activity; Gizmos | Modeling Earth/Moon/Sun System <i>Tides*</i> | PPT 2D and 3D Modeling Activity |
| Week 5 | Seasons | PPT Demos | Scaling Earth/Moon/Mars | PPT Scaling Activity using Balloons PPT | | | | |
| Week 6 | Green House Effect; Water Cycle | Mythbusters Video Book Review | Earth/Moon/Sun System <i>Tides*</i> | 2D and 3D Modeling Activity | | | | |
| Week 7 | Moon Phases | Phase Simulations Moon Activity | What Makes a Planet Geo. Active? | Lab Investigations | | | | |
| Week 8 | Eclipses | PPT | Crater Number Density | Lab Investigations | | | | |
| Week 9 | Projects | Student projects | Experts Lesson on Mars | Video of NASA Expert Scientist; | | | | |

* Not part of the STEM-integrated Treatment curriculum

Statistical Analysis

This study involved a three-level cross-sectional sample consisting of 468 students (level-1) nested within 8 teachers (level-2) nested within 3 schools (level-3). Note that teachers had either 1 to 3 class periods, but due to the missing data on this variable or students not reporting the correct class period this nested level was not considered and all class periods were collapsed within a teacher. In addition to this, since the race/ethnicity group numbers were small for all non-Caucasian racial groups (as shown in Table 2), it was decided to group the non-Caucasian students into a group category of Students of Color (SoC). Thus, a three-level cross-sectional multilevel model (MLM; Hox, 2010; Raudenbush & Bryk, 2002) was used to examine the effects of pretest score (mean centered), gender (0 = girl, 1 = boy), and race/ethnicity (0 = White, 1 = SoC) (level-1) and Treatment condition (0 = control, 1 = experimental) (level-2) on raw scores. A model-building approach was used to determine the nature and statistical significance of pretest score, gender, race/ethnicity, and Treatment condition on LPCI, each spatial domain that made-up the LPCI (PP, GSV, CD, SP), and PSVT:Rot raw scores. Specifically, a series of multilevel models (MLMs) were specified, estimated, tested, and compared in a hierarchical manner to arrive at the final MLM.

Table 4. LPCI Question Topics and Spatial and Scientific Domains

| Question Topics | Spatial Domain | | Scientific Domain |
|---|-------------------------------------|--------------------|--|
| A: Time to complete one orbit | Periodic Patterns | | Periodicity of Moon’s Earthly orbit |
| B: Time between phases (i.e., time between full and first quarter Moon) | Periodic Patterns | | Periodicity of Moon’s phases |
| C: Direction of orbit above the North Pole | Geometric Visualization; Projection | Spatial Spatial | Moon’s orbit direction around Earth as viewed from space |
| D: Direction of Moon rise and Moon set | Cardinal Directions | | Moon Motion |
| E: Alignment to produce various phases such as waxing crescent | Geometric Visualization | Spatial | Phase and Earth/Moon/Sun positions |
| F: Time at which various Moon phases rise and set | Cardinal Directions | | Phase – sky location - time |
| G: Explanation of why the Moon’s appearance changes over time | Geometric Visualization | Spatial | Cause of phases |
| H: How does the Moon’s appearance change when viewed around the world on the same day | Spatial Projection | | Effect of lunar phase with change in Earthly location |

First, an unconditional (null) model consisting of no predictors was fit to the data. Second, a covariate or main effects only model was fit to the data that consisted of pretest score, gender, race/ethnicity, and treatment condition. Third, a model including the two interactions of primary interest (gender by treatment and race/ethnicity by treatment) were added to the model. To test the difference between nested MLMs, a likelihood ratio test (LRT) or sometimes referred to as deviance difference test was used to test whether each subsequent, larger (i.e. more complicated) model was statistically significantly better than a previous, smaller (i.e. simpler) model. If a model including additional parameters was deemed better fitting than a previous model, it was retained and interpreted. If no difference was found between two subsequent models, then the smaller (reduced) model was retained. If a model including both interactions was deemed better than the main effects only model, then it is known that at least one interaction term was important. To determine which interaction term or both was statistically significantly contributing to the model a backward elimination strategy was used. That is, if the difference in fit between a model without an interaction term versus a model with both interaction terms is nonsignificant, then that interaction term can be eliminated. If the difference in fit is statistically significant, then the interaction term should be retained. The LRTs were based on a full information maximum likelihood estimation method (FIML), while random effects (variances) and fixed effects were estimated using restricted maximum likelihood estimation (REML). Fixed effects were then tested using the convenient Wald test. All statistical significance tests were performed at an alpha level of .05. Hedge’s g (corrected for small sample size) was used as an effect size (ES) measure for specific mean comparisons, with MLM coefficient estimates used as the numerator and respective groups posttest variances. All statistical analyses were conducted via SAS version 9.3. In addition to the MLM analysis, we also conducted descriptive statistics to determine gain scores by group for the overall LPCI, each LPCI spatial domain, and the PSVT:Rot. Descriptive results included students by treatment, race/ethnicity, and gender. Including descriptive results allowed us to shed further light regarding how well each student group performed by domain.

Results

Measures

All quantitative assessments were given to both the Treatment and BAU groups immediately prior to and at the conclusion of their Earth/Space unit implementation. Reliability was calculated using the *Cronbach's alpha*; this measures the instrument's internal consistency. The coefficient alpha was calculated for 0.68 and 0.74 for the overall LPCI and the PSVT:Rot assessments, respectively. LPCI and PSVT:Rot values were acceptable. The subset items making up the spatial domains PP (5 test items), GSV (7 test items), SP (4 test items), and CD (5 test items) had coefficient alphas calculated for 0.64, 0.54, 0.41, and 0.17, respectively. The very low alpha for the CD domain illustrates unreliability with these test items; these items have been historically quite difficult for students. For this paper, we will focus on the overall LPCI, the sub-domains PP, GSV, SP, and the PSVT:Rot.

Multilevel Model Results

Table 5 contains the final MLM results for PSVT:Rot, LPCI overall and the LPCI spatial domains (PP, GSV, and SP). We interpreted the MLM results for each outcome as follows. The results for the LPCI overall score showed the best fitting model to the data was the main effects only model. Specifically, LPCI pretest scores, gender, and race/ethnicity were each statistically significant predictors of LPCI posttest scores regardless of treatment condition. That is, higher pretest LPCI scores tended to lead to higher LPCI posttest scores, boys tended to have higher LPCI posttest scores than girls ($ES = .18$), and Students of Color (SoC) tended to have lower LPCI posttest scores than White students ($ES = .23$), after adjusting for other student and teacher characteristics.

The results for spatial-mathematics domain PP showed the best fitting model was a model including the interaction term of gender by treatment, which was statistically significant. This interaction term can be understood as meaning that differences in BAU and Treatment groups were dependent on gender of the student, after adjusting for PP pretest scores and student race/ethnicity. Specifically, it could be understood as meaning that gender differences were dependent on treatment condition. That is, boys scored higher than girls in the BAU group (Mean difference = .92, $ES = .68$), but this gender difference was not maintained in the Treatment group (Mean difference = -0.01, $ES = 0$). Or, it could be understood to meaning that girls in the Treatment group scored higher than girls in the BAU group (Mean difference = 0.44, $ES = .31$), while boys in the Treatment group scored lower than boys in the BAU group (Mean difference = -0.49, $ES = .36$).

In addition, results for PP suggest SoC tended to have lower PP scores than White students ($ES = .29$), and that higher pretest PP scores tended to lead to higher posttest PP scores, after adjusting for other student and teacher characteristics. Results for domain GSV showed the best fitting model to the data was the main effects only model, which did not include any interaction terms. Results for GSV suggest boys tended to have higher GSV scores than girls ($ES = .19$) and that higher pretest GSV scores tended to lead to higher posttest GSV scores, after adjusting for other student and teacher characteristics. Results for SP and PSVT-Rot showed the best fitting model to the data was the main effects only model. The only statistically significant predictor of SP and PSVT-Rot posttest scores were scores on the respective pretest.

Descriptive Results

In order to unpack the MLM results, we graphed the gain scores by Treatment and BAU groups for the overall LPCI, the PP, GSV, and SP spatial domain items, and the PSVT:Rot test. Figure 1 displays all Treatment sub-groups (Treatment White and SoC Boys and Treatment White and SoC Girls) to be clustered with similar overall LPCI gain scores (similar clustering can be found for the Treatment sub-groups in Figures 2 - 4 displaying the LPCI spatial domain results). This is not the case for the BAU sub-groups. Within the BAU group, Figure 1 shows the BAU White Boys with the largest gain scores followed by the BAU SoC Boys. BAU White Girls displayed even less gain scores than the BAU Boys, and the BAU SoC Girls showed negative gains. Similar disparaging data is displayed for the BAU group's PP, GSV, and SP spatial domains in Figures 2, 3, and 4, respectively.

Conclusions and Significance

We compared Treatment and BAU groups by LPCI outcomes. Overall LPCI results showed pre-test scores predicted post-test scores, boys performed better than girls, and Whites performed better than Students of Color. We also compared Treatment and BAU groups by LPCI spatial domain outcomes. Domain SP showed no statistically significant differences were observed for gender, race/ethnicity, or treatment type. For domain GSV, it was found that boys, in general, tended to have higher GSV post-test scores. Recall each of the LPCI spatial domains contains mental rotation derivatives. As shown in the review of literature, boys often outperformed girls on mental rotation test items, so it is not surprising that boys, in general, had higher GSV post-scores than girls (Wilhelm, 2009). GSV descriptive results (shown in Figure 3) illustrate White and SoC Treatment students with similar gains, but the same cannot be said for the BAU group where only BAU White boys achieved similar gains to that of the Treatment group.

PP post-scores for the Treatment showed no gender gap. However, boys did outperform girls on PP post-scores within the BAU group. Additionally, Treatment girls scored better than BAU girls on this same domain. Research has shown students (especially females) benefit greatly from situated, project-enhanced learning experiences (Boaler, 2002; Morrow & Morrow, 1995) and this might explain why Treatment girls performed better than BAU girls on the PP domain and why no gender gap was observed within the PP domain for the Treatment group.

Table 5. Final MLM results for predicting LPCI, PP, GSV, SP, and PSVT:rot scores

| Parameter | LPCI (n = 462) | PP (n = 462) | GSV (n = 462) | SP (n = 462) | PSVT:Rot (n = 443) |
|---------------------|-------------------|-----------------|------------------|-----------------|-----------------------|
| Fixed effects | | | | | |
| Intercept | 7.65*** | 2.26** | 2.99** | 1.91** | 6.41** |
| Level-1 (Student) | | | | | |
| Pretest | 0.46*** | 0.41*** | 0.29*** | 0.17*** | 0.65*** |
| Gender | 0.61* | 0.92*** | 0.33* | 0.14 | 0.30 |
| Race/Ethnicity | -0.64* | -0.41*** | -0.17 | 0.01 | -0.27 |
| Gender by Treatment | | -0.93** | | | |
| Level-2 (Teacher) | | | | | |
| Treatment | 0.69 | 0.44 | 0.34 | 0.19 | 0.62 |
| Random effects | | | | | |
| Level-1 (Student) | | | | | |
| Residual variance | 8.25*** | 1.43*** | 2.19*** | 1.03*** | 7.51*** |
| Level-2 (Teacher) | | | | | |
| Intercept variance | 0.55 | 0.06 | 0.17 | < .001 | 0.18 |
| Level-3 (School) | | | | | |
| Intercept variance | 1.10 | 0.20 | 0.44 | 0.17 | 1.63 |

Note. LPCI = Lunar Phases Concept Inventory; PP = Periodic Patterns; GSV = Geometric Spatial Visualization; SP = Spatial Projection; PSVT:Rot = Purdue Spatial Visualization Test: Rotations; SoC = Students of Color; Pretest = scores on outcome variable prior to start of study; Gender = girl (0) or boy (1); Ethnicity = white (0) or SoC (1); Treatment = BAU (0) or Treatment condition (1)

* $p < .05$. ** $p < .01$. *** $p < .001$.

Due to limitations of this study (small *N* numbers within BAU groups as well as SoC groups), we can only speculate that the significantly higher scores for the Treatment girls (as compared to the BAU girls) could be due to their project work and the spatially-intensive learning experiences that included daily observations where Treatment students purposefully documented lunar position and appearance while noting patterns and periodicity in journals (Table 3). Although, the only statistically significant predictor of PSVT:Rot posttest scores was the score on the respective pretest, it is interesting to note all groups making similar gain scores except for the BAU girls as shown in Figure 4’s descriptive results. In other words, girls in Treatment group performed similarly to boys, but the same cannot be said of BAU girls. Perhaps, there is a way to close the notorious gender gap, after all, when it comes to orchestrating purposeful spatial experiences.

Effect sizes comparing the treatment conditions were estimated for each outcome and ranged from 0.17 (PSVT:Rot) to 0.31 (PP). Although these effect sizes may be small by most standards, they are similar to effect sizes reported elsewhere comparing two groups (McGraw, Lubinski, & Strutchens, 2006). There are obvious limitations to this study in terms of our small BAU numbers. However, our results warrant further studies to examine in more depth how well spatially-oriented, STEM-integrated Earth/Space curricula can advance students’ learning, especially for females and students of color.

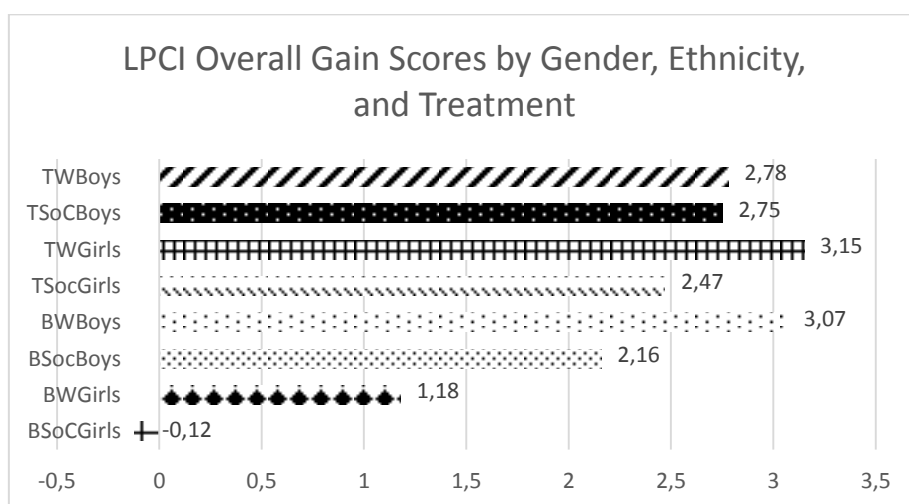


Figure 1. Gain scores by gender [girls/boys], race/ethnicity [white (W)/students of color (SoC)], and treatment [treatment(T)/BAU(B)] for overall LPCI

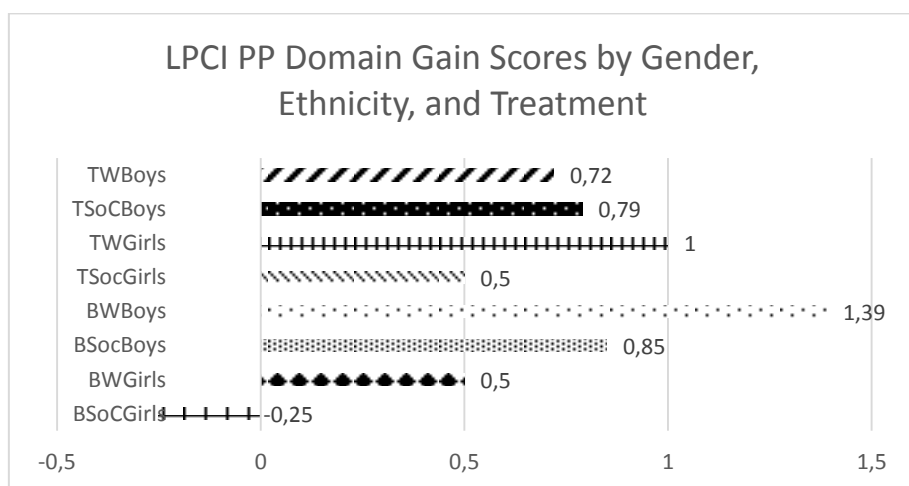


Figure 2. Gain scores by gender [girls/boys], race/ethnicity [white (W)/students of color (SoC)], and treatment [treatment (T)/BAU(B)] for PP domain

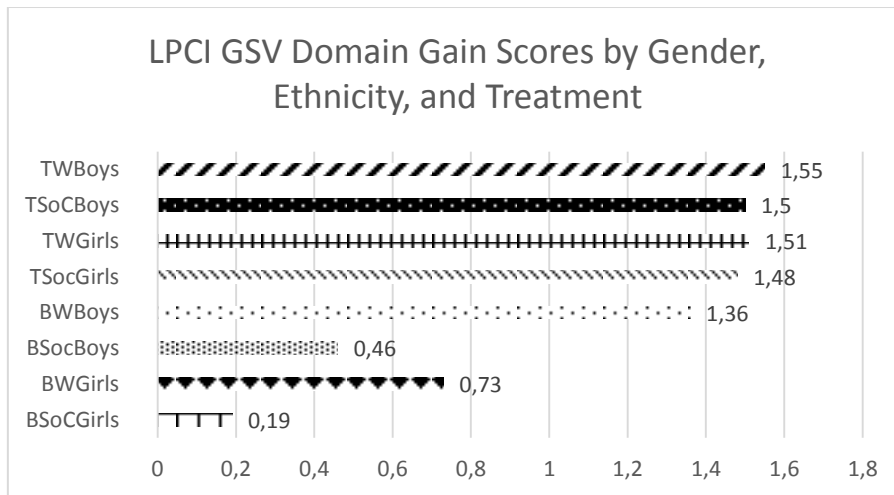


Figure 3. Gain scores by gender [girls/boys], race/ethnicity [white (W)/students of color (SoC)], and treatment [treatment(T)/BAU(B)] for GSV domain

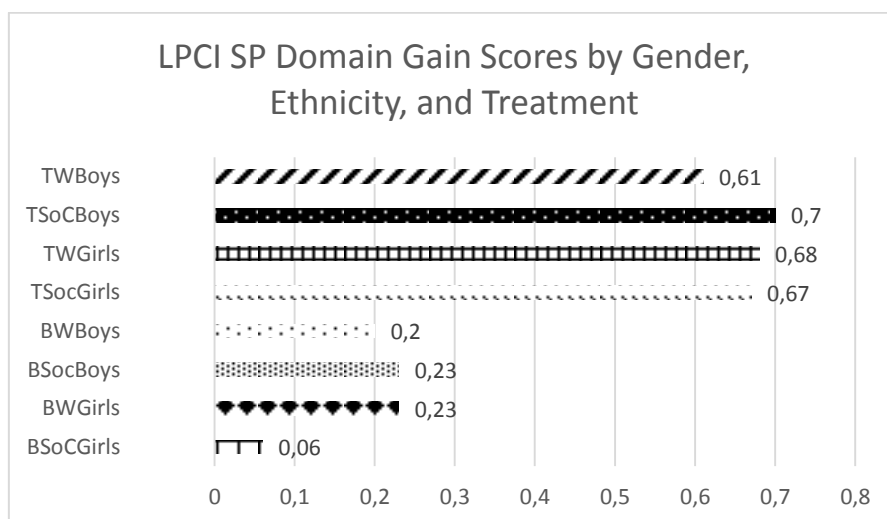


Figure 4. Gain scores by gender [girls/boys], race/ethnicity [white (W)/students of color (SoC)], and treatment [treatment(T)/BAU(B)] for SP domain

The PSVT:Rot gain scores displayed in Figure 5 show all groups making gains between 0.82 and 1.07 except for the BAU girls.

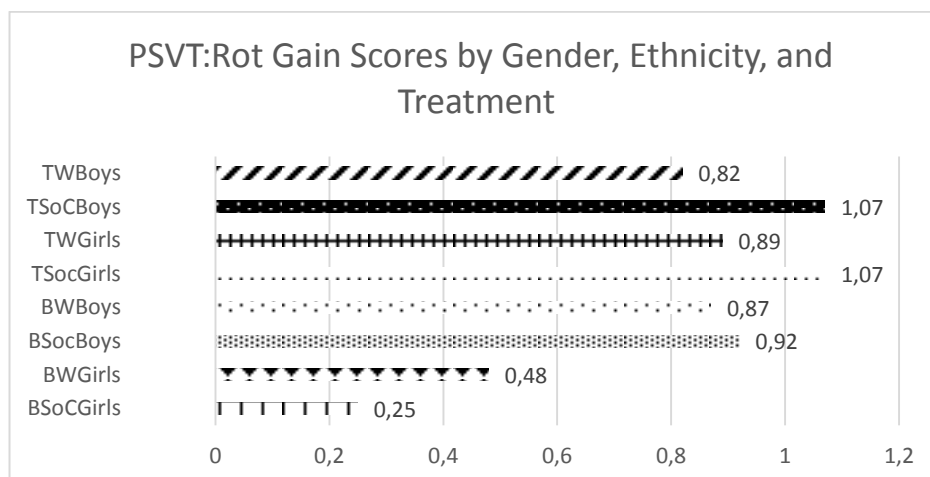


Figure 5. Gain scores by gender [girls/boys], race/ethnicity [white (W)/students of color (SoC)], and treatment [treatment(T)/BAU(B)] for PSVT:rot

The study is unique because it is amongst the first research studies to examine students' spatial-scientific development as they participate in Earth/Space science units. Making the study even more distinctive is discovering how curricular choice and instruction affects student spatial-scientific learning outcomes by gender and race/ethnicity. The authors claimed that one must have well-developed spatial skills in order to understand astronomical phenomena such as lunar phases. Students could come to the classroom already equipped with strong spatial reasoning, ready to understand complicated Earth/Space phenomena; or students will develop the necessary spatial ways of thinking as they make sense of the patterns, geometries, and celestial motions. If we better understand how and which curricular pieces and classroom experiences are instrumental in students' developmental understanding of scientific and spatial content and processes, we can provide more focused interventions to better promote spatial and scientific reasoning with an end effect of better preparedness for all students' STEM achievement.

References

- Ansel, E., & Doerr, H. (2000). NAEP findings regarding gender: Achievement, affect, and instructional experiences. *Results from the Seventh Mathematics Assessment of the National Assessment of Educational Progress*, 73-106.
- Bishop, J. (1996). Astronomy learning and student thinking. *Mercury: Journal of the Astronomical Society of the Pacific*, 25(2). Rocky Mountain Planetarium Association. Retrieved August 1, 2006, from <http://www.rmpadomes.org/library/articles/StudentThinking.pdf>.
- Black, A. (2005). Spatial ability and earth science conceptual understanding. *Journal of Geoscience Education*, 53(4), 402-414.
- Boaler, J. (2002). *Experiencing school mathematics: Traditional and reform approaches to teaching and their impact on student learning* (Volume in the Studies in Mathematical Thinking and Learning Series). Erlbaum Associates, Inc., Mahwah, New Jersey.
- Bodner, G. & Guay, R. (1997). The Purdue visualization of rotations test. *The Chemical Educator*, 2(4), 1-17.
- Giedd, J. N., Blumenthal, J., Jeffries, N. O., Castellanos, F. X., Liu, H., Zijdenbos, A., Paus, T., Evans, A. C., & Rapoport, J. L. (1999). Brain development during childhood and adolescence: a longitudinal MRI study. *Nature neuroscience*, 2(10), 861-863.
- Hake, R. (2002, August). Relationship of individual student normalized learning gains in mechanics with gender, high-school physics, and pretest scores on mathematics and spatial visualization. *Physics Education Research Conference* (Boise, ID).
- Hox, J. J. (2010). *Multilevel analysis: Techniques and applications* (2nd ed.). New York, NY: Routledge.
- Kaufman, S. (2007). Sex differences in mental rotation and spatial visualization ability: Can they be accounted for by differences in working memory capacity? *Intelligence* 35, 211-223.
- Kerns, K., & Berenbaum, S. (1991). Sex differences in spatial ability in children. *Behavior Genetics*, 21(4), 383 - 396.
- Lee, J. (2004). Multiple facets on inequity in racial and ethnic achievement gaps. *Peabody Journal of Education*, 79(2), 51-73.
- Lee, J., & Wong, K. K. (2004). The accountability on racial and socioeconomic equity: Considering both school resources and achievement outcomes. *American Educational Research Journal*, 41(4), 797-832.
- Lindell, R. & Olsen, J. P. (2002). Developing the Lunar Phases Concept Inventory. *Proceedings of the 2002 Physics Education Research Conference*, Eds. S. Franklin, J. Marx and K. Cummings. PERC Publishing, NY.
- Linn, M. & Petersen, A. (1985). Emergence and characterization of sex differences in spatial ability; A meta-analysis. *Child Development*, v.56, p. 1479-1498.
- Lubienski, S. T. (2002). Are we achieving "mathematical power for all?" A decade of national data on instruction and achievement. ERIC, document number ED463166.
- McGraw, R. Lubienski, S.T., & Strutchens, M. E. (2006). A closer look at gender in NAEP mathematics achievement and affect data: Intersections with achievement, race/ethnicity, and socioeconomic status. *Journal for Research in Mathematics Education*, 37(2), 129-150.
- Miller, D., & Halpern, D. (2014). The new science of cognitive sex differences. *Trends in Cognitive Sciences*, 18(1), 37-45.
- Morrow, C., and Morrow, J. (1995). Internal influences on gender differences in mathematics. In P. Rogers & G. Kaiser (Eds.), *Equity in mathematics education: Influences of feminism and culture*. London: Falmer Press, 13-26.
- Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models* (2nd ed.). Newbury Park, CA: Sage.

- Reyes, L. H., & Stanic, G. M. A. (1988). Race, sex, socioeconomic status, and mathematics. *Journal for Research in Mathematics Education*, 19, 26-43.
- Reynolds, M. (1990). Two-dimensional versus three-dimensional conceptualization in astronomy education, *Doctoral dissertation*, University of Florida.
- Rudmann, D. (2002). Solving astronomy problems can be limited by intuited knowledge, spatial ability, or both. *American Educational Research Association*.
- Plummer, J. (2009). A cross-age study of children's knowledge of apparent celestial motion. *International Journal of Science Education*, 31(12), 1571-1605.
- Plummer, J., Wasko, K., & Slagle, C. (2011). Children learning to explain daily celestial motion. *International Journal of Science Education*, 33(14), 1963-1992.
- Plummer, J. (2014). Spatial thinking as the dimension of progress in an astronomy learning progression. *Studies in Science Education*, DOI: 10.1080/03057267.2013.869039.
- Pribyl, J., & Bodner, G. (1987). Spatial ability and its role in organic chemistry: A study of four organic courses. *Journal of Research in Science Teaching*, 24(3), 229-240.
- Sanchez, C. (2012). Enhancing visuospatial performance through video game training to increase learning in visuospatial science domains. *Psychometric Bulletin and Review*, 19(1), 58-65.
- Silverman, I., Choi, J., & Peters, M. (2007). The hunter-gatherer theory of sex differences in spatial abilities: Data from 40 countries. *Archives of Sexual Behavior*, 36(2), 261-268.
- Sorby, S. (2006, March). Gender differences in spatial reasoning skills and their effects on success and retention in engineering programs. *Engineers Week*. Retrieved August 1, 2006, from http://www.eweek.org/2002/news/eweek/2006_marathon/buildingskills_2.ppt.
- Sorby, S., Casey, B., Veurink, N., & Dulaney, A. (2013). The role of spatial training in improving spatial and calculus performance in engineering students. *Learning and Individual Differences*, 26, 20-29.
- Wellner, K. (1995). A correlational study of seven projective spatial structures with regard to the phases of the moon, *Doctoral dissertation*, University of Iowa.
- Wilhelm, J. (2009). Gender differences in lunar- related scientific and mathematical understandings. *International Journal of Science Education*, 31(15), 2105-2122.
- Wilhelm, J., Jackson, C., Sullivan, A., & Wilhelm, R. (2013). Examining differences between preteen groups' spatial-scientific understandings: A quasi-experimental study. *The Journal of Educational Research*, 106(5), 337-351.

Author Information

Jennifer Wilhelm

University of Kentucky
 Lexington, Kentucky 40506, U.S.A.
 Contact e-mail: jennifer.wilhelm@uky.edu

Michael Toland

University of Kentucky
 Lexington, Kentucky 40506, U.S.A.

Merryn Cole

University of Kentucky
 Lexington, Kentucky 40506, U.S.A.
