Gender and Racial Differences: Development of Sixth Grade Students' Geometric Spatial Visualization within an Earth/Space Unit

Christa Jackson
Iowa State University

Jennifer Anne Wilhelm
University of Kentucky

Mary Lamar
University of Kentucky

Merryn Cole
University of Kentucky

This study investigated sixth-grade middle-level students' geometric spatial development by gender and race within and between control and experimental groups at two middle schools as they participated in an Earth/Space unit. The control group utilized a regular Earth/Space curriculum and the experimental group used a National Aeronautics and Space Administration-based curriculum. The quantitative data sources included the Lunar Phases Concept Inventory, Geometric Spatial Assessment, and the Purdue Spatial Visualization–Rotation Test. The results indicated the experimental males and females, and the students of color and white students in the experimental group showed significant gains in their understanding of geometric spatial visualization from pre- to post-implementation. However, for the control group, the significant gains were limited to the males and the white students. The findings reveal that support is needed for males, females, and all racial groups to have the opportunity to develop their spatial reasoning, which in turn, increases students' scientific understanding.

The mathematics education literature has used a variety of phrases to refer to spatial ability such as spatial reasoning, spatial visualization, and spatial skills (Baki, Kosa, & Guven, 2011). Other researchers distinguish between spatial ability and spatial skill, arguing that spatial ability is an innate ability and spatial skill is something that can be learned (Sorby, 1999). McGee (1979) defined spatial visualization as "the ability to mentally manipulate, rotate, twist, or invert a pictorially presented stimulus" (p. 893). The authors of the *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics (NCTM, 2000)) emphasize that students in prekindergarten through grade 12 need to "use visualization, spatial reasoning, and geometric modeling to solve problems" (p. 41). Young students begin developing their spatial skills by first manipulating objects, and then as they get older, those manipulations extend to mental images (NCTM, 2002). A student's spatial sense and the ability to problem solve is positively correlated, indicating that "spatial visualization is a good predictor of successful problem solving" (NCTM, 2002, p. 59). However, middle-level students have difficulties in mentally rotating objects and answering questions related to mental rotation (NCTM, 2002). Middle-level students need to be given opportunities to develop and connect spatial visualization to other subjects beyond mathematics (NCTM, 2002).

Delgado and Prieto (2004) argue that spatial ability is an essential component to both scientific thinking and mathematics. Black (2005) "hypothesized that mental rotation is the most important in understanding Earth science concepts" (p. 403) because mental rotation is the ability to look at an object and visualize its rotation in three-dimensional space. Although spatial ability is important in learning Earth Science concepts (Black, 2004), spatial visualization and spatial orientation are imperative when learning astronomy (Broadfoot, 1995). Yair, Schur, and Mintz (2003) claim that in order for students to understand astronomical phenomena such as moon phases, they must be able to visualize objects as they occur from different perspectives. Black (2004) states that "moon phases, an astronomical phenomenon involving movement of a half-lit body in space viewed from the unavoidable fixed position of Earth observers," is a difficult concept for students to learn (Mulholland & Ginns, 2008). For example, many individuals have the misconception that the phases of the moon are caused by the Earth casting a shadow on the moon (Baxter, 1989; Dai, 1991). In a study of seventh, eighth, and ninth graders in Australia, Danaia and McKinnon (2007) found that the shadow explanation was the most common misconception related to moon phases exhibited by the participants. Whereas 27.7% of seventh graders and 29.7% of eighth graders held this misconception, over half (51.2%) of the ninth graders in the study did as well. This misconception generally persists into adulthood (Parker & Heywood, 2007). To fully understand the causes of lunar phases, students must know (a) how the earth and moon revolve and their relative position to the sun as they move, (b) the moon is lit by the sun and we only see the illuminated side of the moon, (c) the illuminated side changes as the position of the moon and earth change relative to the sun, and (d) how to create a three-dimensional mental image of this motion in their minds (Bayraktar, 2009). Students must be able to visualize the positions and motions of celestial objects to describe observed lunar phenomena and make predictions based on these observations (Plummer, Wasko, & Slagle, 2011). We contend students must have a strong spatial sense in order to understand many astronomical concepts. In particular, students must have developed understanding in geometric spatial visualization (GSV; visualizing a system in space from above/below/within a system's plane).

Gender and Racial Differences

Gender differences in spatial ability have been well documented in the scientific literature (e.g., Coluccia & Louse, 2004; Halpern, 2007; Levine, Vasilyeva, Lourenco, Newcombe, & Huttenlocher, 2005), with males having the advantage on mental rotation tasks (Linn & Petersen, 1985; Lippa, Collaer, & Peters, 2010). Research has shown that the difference is more pronounced for certain types of spatial ability (e.g., mental rotation, spatial perception) than for spatial visualization (Voyer, Voyer, & Bryden, 1995). However, the authors pointed to a lack of consensus in the literature as to how to categorize spatial abilities, leading to some difficulty in categorizing what the assessments were measuring.

Males have shown higher test results on science and mathematics assessments that have spatial reasoning components. For example, males score higher than females on standardized exams (e.g., SAT and Graduate Record Exam [GRE]) in science and mathematics, and this performance has been correlated with males' spatial sense (Halpern et al., 2007). Consequently, "mental rotation skill has been shown to be a meaningful mediator of gender differences in mathematics by the end of high school" (Casey, Nuttall, & Pezaris, 2001, p. 34).

Campbell (1989) and Manning (1998) argue that one should also examine racial differences when discussing the relationship of gender and academic performance. Researchers (Casey, Colón, & Goris, 1992; Eals & Silverman, 1994) reported males have stronger spatial ability than females and spatial ability favors males regardless of racial classification. However, McGraw, Lubienski, and Strutchens (2006) analyzed the 2003 National Assessment of Educational Progress (NAEP) by race, gender, and socioeconomic status within the *Principles and Standards of Mathematics* (NCTM, 2000) five content standards: number sense and operations, data analysis and probability, algebra, geometry, and measurement, and found a significant difference favoring African American females in geometry, which includes spatial visualization on the fourth- and eighth-grade sections of the NAEP. No other significant differences were found within this content area by other racial or gender groups at the fourth- and eighth-grade levels.

Currently, few research studies speak directly to gender differences on astronomical assessments and only one had specifically examined spatial development before and after an integrated mathematics/science curricular intervention (Wilhelm, 2009). No research, to our knowledge, has examined racial differences in this context. Thus, the purpose of this study was to examine differences between two groups (i.e., experimental and control) of sixth-grade students' scientific content knowledge and geometric spatial reasoning from pre- to post-implementation of an astronomy unit. Specifically,

1. To what extent do sixth-grade middle-school students' GSV change after participating in an integrated Earth/Space unit?
2. What gender differences exist in GSV between students as they reason and consider lunar phases within and between control and experimental groups?
3. What differences exist in GSV among sixth-grade middle-level students from different racial backgrounds as they reason and consider lunar phases between control and experimental groups?

Research Methods

Participants

Research subjects were sixth-grade middle-school students from two middle schools (Juniper and Butternut: all names are pseudonyms) in the south-central region of the United States. The demographics of Juniper Middle School (n = 890) were 83% White, 6% African American, 4% Hispanic, 4% Asian, and 3% Other; and 30% were eligible for free and reduced-price meals, whereas students' self-reported races were 58.6% White, 5% African American, 10% Hispanic, 1.2% Native American, 8.4% Asian, 9.2% Other, and 7.6% Declined to answer. For the purpose of this study, the students' self-reported demographical data will be used. Mrs. Stevens and Mrs. Castle taught the experimental group (n = 141) with four and nine years of teaching experience, respectively. Ms. Branson, the control group teacher (n = 65), taught for four years.

The demographics of Butternut Middle School (n = 1,131) were 74% White, 10% African American, 8% Hispanic, 5% Asian, and 3% Other; and 26% were eligible for

Table 1
Butternut's Earth/Space Unit Timeline by Group With Lesson Content and Method of Implementation

Week	Butternut			
	Control		Experimental	
	Lesson	Method	Lesson	Method
Week 1	Introduction 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	Angular measurement and measuring the diameter of the moon How Far to the Star? (The Parallax Effect)	Lab work, note taking, and whole class discussion	How can I say where I am on the Earth? Longitude/latitude Rotation/revolution and seasons*	Earth globe activity PPT Modeling activity
Week 3	Why is Earth the only possible place for life?	Lab work using probeware	What can we learn by examining the moon's surface? Scaling	Exploration of lunar images PPT Scaling activity using PPT
Week 4	Moon phases Eclipses; tides	Oreo moon phases; 3D Earth/moon/sun activity; explore- learning gizmos	Modeling Earth/moon/sun system Tides*	2D and 3D modeling activity

Note. *Means this lesson was not part of the REAL Curriculum.

free and reduced price-meals, whereas students' self-reported races were 55.8% White, 6.3% African American, 2.4% Hispanic, 4.9% Asian, 4.9% Native American, 14.6% Other, and 11.1% Declined to answer. Again, for the purpose of this study, the students' self-reported demographical data will be used. Ms. Apple, Mr. Land, and Mrs. Roling, with 1, 1, and 11 years of teaching experience, respectively, taught the experimental group (n = 221). Mrs. Fehr, the control group teacher (n = 28), taught for 12 years.

The control groups at Juniper and Butternut studied the astronomical unit using their regular Earth/Space curriculum, and the experimental group implemented a National Aeronautics and Space Administration (NASA)-based curriculum called Realistic Exploration in Astronomical Learning (REAL) (Wilhelm & Wilhelm, 2007), which is an interdisciplinary curriculum that integrates mathematics and science content. Both student groups studied Earth/Space concepts related to the Solar System within their units. The teachers from Butternut implemented the Earth/Space unit for four weeks, whereas the teachers from Juniper implemented the unit for nine weeks. Tables 1 and 2 (Wilhelm, Jackson, Toland, Cole, & Wilhelm, 2013) outline the time spent on the unit by each group, the content implemented, and the method of implementation.

Measures

This research focused on students' geometric spatial content and scientific knowledge from pre- to post-implementation of an Earth/Space unit. Students' spatial and scientific understanding was assessed pre and post via survey instruments. The quantitative data sources included three instruments, the Lunar Phases Concept Inventory (LPCI) (Lindell & Olsen, 2002), the Geometric Spatial Assessment (GSA, Wilhelm & Wilhelm, 2007), and the Purdue Spatial Visualization—Rotation Test (PSVT-Rot) (Bodner & Guay, 1997). The LPCI is a 20-question multiple-choice instrument that assesses eight science domains and four mathematical domains. However, for the purpose of this study, we discuss three of the eight science domains: (a) Domain C: Direction of the moon's orbit around Earth; (b) Domain E: Phase due to sun/Earth/moon positions; and (c) Domain G: Causes of lunar phases due to their relationship to GSV. Additionally, we focus on one of the four mathematics domains: Geometric Spatial Visualization. This resulted in analyzing 7 of the 20 multiple-choice questions of the LPCI. The GSA is a 16-question multiple-choice instrument that assesses the same four mathematics domains as the LPCI. Again, for the purpose of this study, we focus only on GSV, which resulted in analyzing four questions. The PSVT-Rot is a 20-question multiple-choice instrument that measured students' rotational reasoning.

Table 2
Juniper's Earth/Space Unit Timeline by Group With Lesson Content and Method of Implementation

Week	Juniper			
	Control		Experimental	
	Lesson	Method	Lesson	Method
Week 1	Big Bang Theory; Solar system	PPT Modeling Expanding Universe Balloons	Overview of universe* Why does the moon appear to change its shape?	"Many Moons" by Thurber, moon journaling (5 weeks), Stellarium (planetarium software)
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
Week 3	Stars	Parallax Activity Stellarium	How to say where I am on the Earth? Longitude/latitude Rotation/revolution	Earth globe activity PPT Modeling activity
Week 4	Planets; Earth (day/night)	Foam ball models Graphing	What can we learn by examining the moon's surface?	Exploration of lunar images
Week 5	Seasons	PPT Demos	Scaling Earth/moon/Mars	PPT Scaling Activity using Balloons
Week 6	Green House Effect; Water Cycle	Mythbusters video Book review	Earth/moon/sun system Tides*	PPT 2D and 3D modeling activities
Week 7	Moon phases	Phase simulations	What makes a planet geologically 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 scientists; projects

Note. *Means this lesson was not part of the REAL Curriculum.

Data Analysis

To answer the research questions a repeated measures analysis of variance (RM-ANOVA) was conducted with the factors being teacher, experimental/control group, white students in the experimental/control group, students of color in the experimental/control group, and gender and the dependent variables for each factor being pre/post scores. All of the assumptions for the RM-ANOVA were confirmed prior to analyzing the data. The data were continuous with matched pairs. No significant outliers were part of the analyzed data. The distribution of data were approximately normally distributed. Because RM-ANOVA was used with pre- and post results (two conditions), sphericity was not an issue with these analyses. The analysis was conducted for the three science domains of the LPCI, the GSV of the GSA, and the overall scores of the PSVT-Rot. In addition, the frequency of correct and incorrect answers was calculated for each question on the LPCI and GSA to assess the change in students' knowledge of GSV and students' scientific understanding. Finally, descriptive statistics of students' knowledge of GSV on each assessment were calculated for race and gender.

Assessments

For the purpose of this study, the reliability was not calculated on the full LPCI and GSA assessment. Instead, reliability for the LPCI and GSA only focused on the questions related to GSV. Consequently, reliability for the LPCI (7 questions), GSA (4 questions), and PSVT-Rot (20 questions) was calculated using the Cronbach's alpha, which measures the instrument's internal consistency. The coefficient alpha for Butternut Middle School was calculated for .51, .45, and .63 for the LPCI, GSA, and PSVT-Rot post-assessment scores, respectively. The coefficient alpha for Juniper Middle School was calculated for .38, .44, and .70 for the LPCI, GSA, and PSVT-Rot post-assessment scores, respectively. The coefficient alpha was low to moderate low for both middle schools for the LPCI and GSA because it was calculated on a portion of each assessment, whereas the coefficient alpha, which was calculated on the full PSVT-Rot assessment, was moderate.

Results

Butternut's LPCI Results

Students were assessed pre/post on the LPCI on three science domains. An RM-ANOVA showed a significant increase in mean values of overall scores for the experimental group with large effect sizes (Huck, 2012) on Domain C: Direction of the moon's orbit around Earth, $F(1, 220) = 133.2$, $p < .001$, partial $\eta^2 = .38$; Domain E: Phase and sun/Earth/moon positions, $F(1, 220) = 152.19$, $p < .001$, partial $\eta^2 = .41$; and Domain G: Cause of lunar phases, $F(1, 220) = 56.38$, $p < .001$, partial $\eta^2 = .20$. The students of color (i.e., African Americans, Asians, Native Americans, and Hispanics) in the experimental group ($n = 72$) revealed a significant increase with large effect sizes from pre to post on Domain C, $F(1, 71) = 32.95$, $p < .001$, partial $\eta^2 = .40$; Domain E, $F(1, 71) = 34.78$, $p < .001$, partial $\eta^2 = .33$, and Domain G, $F(1, 71) = 2.61$, $p < .001$, partial $\eta^2 = .23$. Similarly, the white students ($n = 131$) showed a significant increase on Domain C, $F(1, 130) = 86.03$, $p < .001$, partial $\eta^2 = .40$; Domain E, $F(1, 130) = 130.13$, $p < .001$, partial $\eta^2 = .50$; and Domain G, $F(1, 130) = 36.71$, $p < .001$, partial $\eta^2 = .22$. There was a significant increase from pre to post on each science domain for experimental males and females (see Table 3). As we look at the increase by experimental teacher, both males and females show a significant increase on domains C and E; and only Mrs. Roling's and Ms. Apple's males and females show a significant increase on Domain G. The control group revealed significant increases in overall scores with a large effect size from pre to post on Domain C, $F(1, 27) = 6.53$, $p < .05$, partial $\eta^2 = .20$, and only the control females showed a significant increase on Domain C. The white students in the control group ($n = 15$) revealed a significant increase only on Domain G, $F(1, 14) = 6.14$, $p < .05$, partial $\eta^2 = .31$.

In examining students' post responses to the seven individual questions, a majority of the students (both male and female) selected the correct response. For example, a question from Domain E asked students to determine the correct alignment for the Earth, moon, and sun that would produce a full moon if they were looking down from a point located above the Earth's north pole. Fifty-seven percent of the experimental males and 65% of the experimental females correctly answered this question. Although 50% of the experimental males selected the correct response to the correct alignment of the Earth, moon, and sun to see a waxing crescent in Figure 1, only 16% of the experimental females selected the correct response (choice B), whereas 16% chose A, 19% C, 9% D, 6% E (none of the above), and 8% F (more than one of the above).

On the same question, 46% of the control males correctly answered the question and 31% of the control females. However, 38% of the control females selected choice C, 23% D, and 7% E. The experimental group overall, experimental males, experimental females, the control group overall, and the control females showed a significant increase in their scores from pre to post on the GSV domain of the LPCI. Table 4 displays the significant increase of all gender groups by teacher.

In order to fully determine if students had a deep understanding on the direction the moon orbits around the earth, students were asked two questions: one at the beginning of the LPCI and one at the end. The question at the beginning stated: "The Moon orbits around the Earth; in which direction does it orbit if observed from a point directly above the Earth's North Pole?" The question at the end asked, "Which direction did the moon travel around the Earth?" Both questions had the same multiple-choice options: clockwise, counterclockwise, or either direction. Of the experimental group, 79% correctly responded to the first question on the post-assessment with the correct answer of counterclockwise, and 91% correctly responded on the second question, again with the correct answer of counterclockwise. When examining the males' and females' responses from pre to post, the males' percentage gain of 44% was slightly higher than the females (39.5%), although both genders' pre-percentage scores were similar (males 36% and females 36.5%). However, on the second question, the females' pre-percentage score (63%) was higher than the males (49%), but the males made a larger gain (40%) on the post than the females (30%). The control group males' pre- and post-percentage on the first question was the same (67%) showing no gain, whereas the females had a 46% gain from pre to post. On the last question, the males and females scored the same percentage on the pre (46%), whereas on the post, the females percentage increased by 31% and the males increased by 21%. Overall, both the experimental and control groups made the highest gains on the last question, and the control females made higher gains than the control males on both questions.

As we examine students' average percentage scores from pre to post by race, we see an increase in all demographics for the experimental group. Yet, we see a smaller gain from pre to post for the control group (see Figure 2).

As we look at the experimental group more generally by whites and students of color, we find the average percentage pre (29.88%) for white students increased to 57.14%, and students of color average pre (26.35%) increased to 56.98%. However, for the control group, the students of

Table 3
Butternut's Percentage Correct on Science Domains Pre to Post

Group	*N*	C: Direction of the Moon's Orbit Around Earth					E: Phase and Sun/Earth/Moon Positions					G: Cause of Lunar Phases				
		Pre (*SD*)	Post (*SD*)	F	*p* Value	Partial η^2	Pre (*SD*)	Post (*SD*)	*F*	*p* Value	Partial η^2	Pre (*SD*)	Post (*SD*)	F	*p* Value	Partial η^2
All experimental	221	49 (38)	84 (29)	133.2	.00**	.38	21 (24)	53 (34)	152.19	.00**	.41	18 (28)	40 (38)	56.38	.00**	.20
Experimental males	98	48 (39)	84 (30)	59.15	.00**	.38	24 (26)	56 (35)	63.45	.00**	.40	15 (23)	43 (40)	43.95	.00**	.31
Experimental females	123	50 (38)	85 (28)	73.45	.00**	.38	19 (23)	51 (33)	88.62	.00**	.42	21 (31)	37 (36)	17.83	.00**	.13
Mr. Land males	21	52 (43)	79 (34)	5.99	.02*	.23	21 (22)	44 (43)	5.07	.04*	.20	14 (23)	31 (33)	4.38	.05	.18
Mr. Land females	27	48 (43)	78 (29)	9.3	.01*	.26	22 (23)	47 (36)	10.36	.00**	.29	24 (38)	37 (33)	2.75	.11	.10
Mrs. Rong males	40	49 (40)	84 (33)	18.92	.00**	.33	28 (27)	58 (29)	31.12	.00**	.44	13 (22)	46 (40)	28.69	.00**	.42
Mrs. Rong females	42	55 (36)	83 (29)	17.42	.00**	.3	15 (21)	40 (28)	33.36	.00**	.45	17 (29)	35 (37)	7.41	.01*	.15
Ms. Apple males	37	46 (36)	88 (25)	44.47	.00**	.55	22 (26)	61 (35)	35.09	.00**	.49	18 (24)	46 (43)	13.81	.00**	.28
Ms. Apple females	54	46 (37)	89 (27)	53.51	.00**	.50	21 (25)	61 (33)	52.65	.00**	.50	23 (30)	39 (36)	7.53	.01**	.12
All control	28	48 (40)	71 (37)	6.53	.02*	.20	29 (27)	40 (38)	2.17	.15	.07	20 (34)	38 (35)	3.95	.06	.13
Control males	15	57 (42)	67 (41)	.68	.42	.05	27 (29)	42 (43)	2.58	.13	.16	27 (37)	47 (30)	1.91	.19	.12
Control females	13	38 (36)	77 (33)	8.96	.01*	.43	31 (25)	38 (33)	.32	.58	.03	12 (30)	27 (39)	2.18	.17	.15

$^{**}p < .01$; $^{*}p < .05$.

color (35.16%) average pre were slightly higher than whites (28.57%), but the whites average post (60.95%) surpassed the students of color (52.75%).

Juniper's LPCI Results

An RM-ANOVA revealed a significant increase in scores with a small effect size from pre to post on Domain C for experimental overall, $F(1, 140) = 4.65$, $p < .05$, partial $\eta^2 = .03$ and experimental males, $F(1, 76) = 4.33$, $p < .05$, partial $\eta^2 = .05$, and on Domain E for experimental

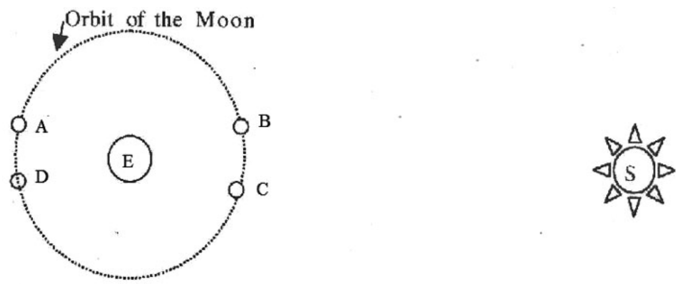

Figure 1. Earth/moon/sun alignment (Lindell & Olsen, 2002).

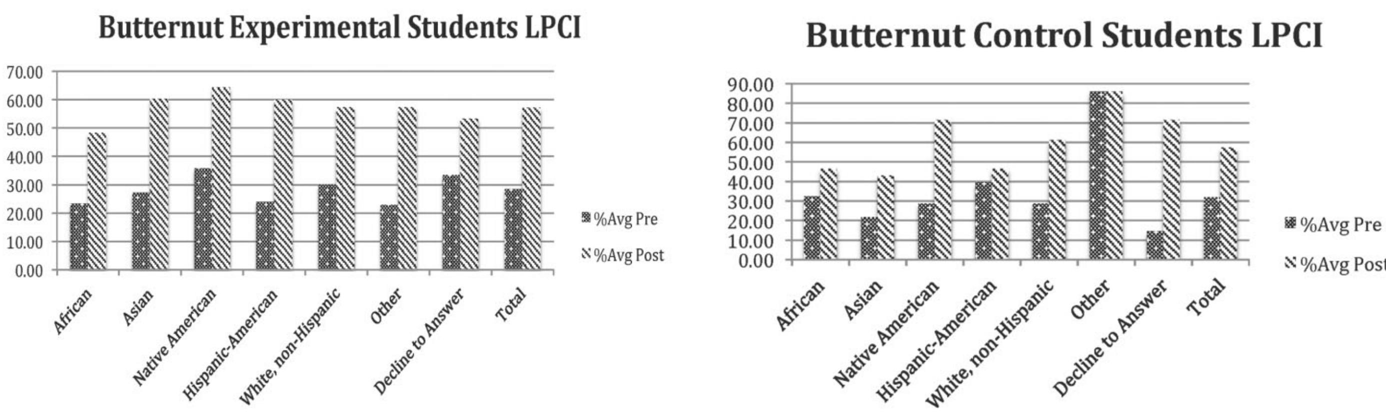

Figure 2. LPCI results by race for Butternut's experimental and control groups.

Table 4

Butternut's Percentage Correct on LPCI GSV Domain From Pre to Post

Group	*N*	Pre (*SD*)	Post (*SD*)	F	*p* Value	Partial η^2
All experimental	221	28 (17)	58 (23)	328.86	.00**	.60
Experimental males	98	28 (17)	60 (22)	162.80	.00**	.63
Experimental females	123	28 (18)	56 (23)	167.08	.00**	.58
Mr. Land males	21	28 (18)	50 (23)	19.52	.00**	.49
Mr. Land females	27	30 (20)	53 (24)	21.07	.00**	.45
Mrs. Rong males	40	30 (18)	62 (21)	81.00	.00**	.68
Mrs. Rong females	42	27 (17)	51 (21)	50.18	.00**	.55
Ms. Apple males	37	27 (16)	64 (22)	70.65	.00**	.66
Ms. Apple females	54	29 (17)	63 (23)	108.22	.00**	.67
All control	28	32 (19)	48 (25)	10.28	.00**	.28
Control males	15	35 (23)	50 (30)	3.80	.07	.21
Control females	13	27 (12)	46 (19)	6.88	.02*	.36

**p < .01; *p < .05.

Table 5
Juniper's Percentage Correct on Science Domain E From Pre to Post

Group	*N*	E-Phase and Sun/Earth/Moon Positions				
		Pre (*SD*)	Post (*SD*)	F	*p* Value	Partial η^2
All experimental	141	19 (23)	32 (29)	25.57	.00**	.15
Experimental males	77	23 (25)	35 (30)	12.56	.00**	.14
Experimental females	64	14 (19)	28 (28)	12.90	.00**	.17
Mrs. Castle males	55	25 (26)	38 (32)	7.88	.01*	.13
Mrs. Castle females	41	16 (18)	33 (28)	13.34	.00**	.25
Mrs. Stevens males	22	17 (20)	29 (24)	5.51	.03*	.21
Mrs. Stevens females	23	10 (19)	19 (26)	1.54	.23	.07
All control	65	19 (24)	37 (30)	15.55	.00**	.20
Control males	33	15 (19)	40 (30)	21.60	.00**	.40
Control Females	32	23 (27)	34 (31)	2.29	.14	.07

**p < .01, *p < .05.

overall, experimental males, and experimental females (see Table 5).

The experimental students of color (n = 50) showed significant gains with medium effect sizes on Domain E, F(1, 49) = 12.644, p < .05, partial η^2 = .21 and Domain G, F(1, 49) = 4.26, p < .05, partial η^2 = .08. The white students (n = 73) revealed a significant increase with a large effect size on Domain E, F(1, 72) = 11.73, p < .05, partial η^2 = .14. As we examine Domain E by experimental teacher, all groups except Mrs. Stevens' females showed significant increases. Mrs. Castle's experimental females showed a significant increase with a medium effect size on Domain G, F(1, 40) = 4.14, p < .05, partial η^2 = .09. The control group overall and control males showed significant gains on Domain E. Moreover, only the white students in the control group (n = 42) showed a significant increase from pre to post on Domain E, F(1, 41) = 12.70, p < .05, partial η^2 = .24 and Domain G, F(1, 41) = 4.32, p < .05, partial η^2 = .10. The control group did not show any other significant increases from pre to post on any other domains.

Similar to Butternut, 34% of experimental males and 45% females selected the correct alignment for the Earth, moon, and sun to produce a full moon on the post-assessment. There was also variation in responses on the post-assessment among the females when choosing the correct alignment to have a waxing crescent, with 11% choosing A, 20% B, 31% C, 16% D, 9% E, and 12.5% F. Although more males (32%) selected the correct answer, there was also variation in their responses: 14% A, 19% C, 13% D, 5% E, and 16% F.

The experimental group overall, experimental males, experimental females, control group overall, and the control males showed a significant increase in their scores from pre to post on the GSV domain of the LPCI. Table 6 displays the significant increase percentages of all gender groups by teacher. There was not a significant increase from pre to post for Mrs. Stevens' males and females, and the control females.

In order to examine if students had an understanding on the direction the moon orbits around the earth, two questions were asked on the LPCI related to this topic. Of the experimental group, 52% correctly responded to the first question, and 57% correctly responded on the second question. When examining the males' and females' responses from pre to post, the gains were similar: males 13% and females 16%, with females having a slightly higher pre- and post-percentage. However, on the last question, the females (61%) started higher than the males (51%) on the pre, similar to Butternut, but slightly decreased (57%), whereas the males increased (56%) on the post. The control group males had a gain of 9% and the females had a gain of 10% from pre to post on the first

Table 6
Juniper's Percentage Correct on LPCIGSV Domain From Pre to Post

Group	*N*	Pre (*SD*)	Post (*SD*)	F	*p* Value	Partial η^2
All experimental	141	26 (17)	36 (21)	25.78	.00**	.16
Experimental males	77	28 (18)	38 (22)	14.38	.00**	.16
Experimental females	64	25 (16)	34 (21)	11.23	.00**	.15
Mrs. Castle males	55	29 (20)	40 (22)	10.70	.00**	.17
Mrs. Castle females	41	27 (16)	39 (22)	11.59	.00**	.23
Mrs. Stevens males	22	25 (14)	34 (19)	3.52	.07	.14
Mrs. Stevens females	23	20 (15)	25 (16)	1.11	.30	.05
All control	65	25 (18)	38 (22)	12.73	.00**	.17
Control males	33	24 (17)	42 (22)	12.56	.00**	.28
Control females	32	26 (19)	34 (21)	2.42	.13	.07

**p < .01.

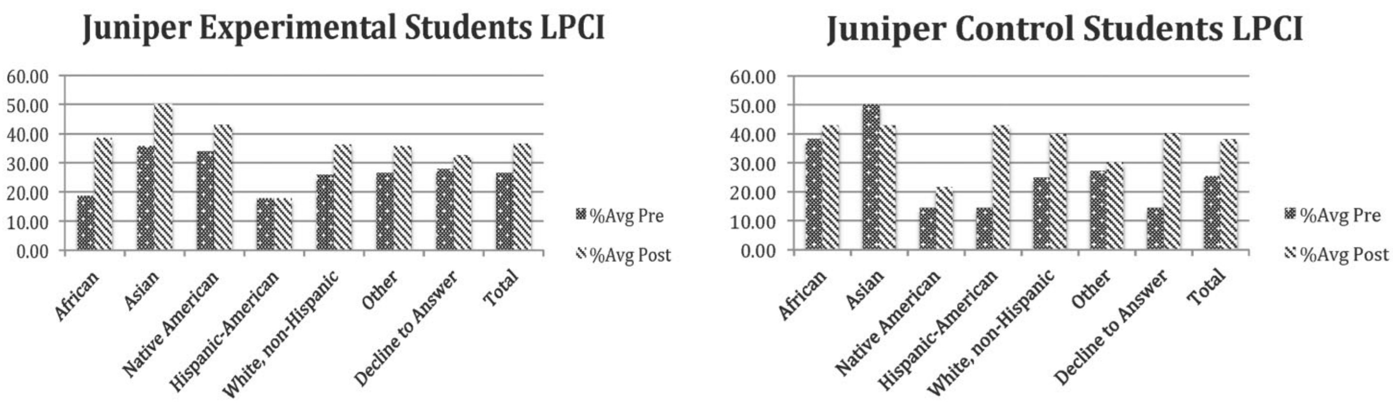

Figure 3. LPCI results by race for Juniper's experimental and control groups.

question. On the last question, the males had a gain of 18%, whereas the females decreased 3% from pre to post.

As we examine students' average scores from pre to post by race, we see an increase in all demographics from pre to post for the experimental group. Yet, we see a smaller gain from pre to post for the control group (see Figure 3).

As we look at the experimental group more generally by whites and students of color, we find whites average pre (25.83%) increased to 36.20%, and students of color average pre (27.10%) increased to 36.76%. However, for the control group, the students of color average pre (26.09%) were slightly higher than whites (24.83%), but the whites average post (39.80%) slightly surpassed the students of color (34.78%).

LPCI: Common Student Misconception

At both middle schools (Butternut and Juniper), the most commonly missed question on the post-assessment by both the experimental and control groups was explaining how a new moon occurs when no lighted portion of the moon is visible to an observer on Earth. At Butternut, on the post-assessment, only 33% of the experimental group correctly responded that the moon is between the Earth and the Sun, whereas 20% stated that the moon is completely covered by the shadow of the Earth, and 13% mentioned the moon is completely covered by the shadow of the sun. However, 43% of the control group at Butternut correctly answered, which was a 25% gain from the pre. Similar to the experimental group, 14% selected the other two choices. For Juniper's experimental group, 19% responded correctly, and 25% selected the option of the moon is completely covered by the Earth's shadow and 16% said the moon is completely covered by the sun's shadow. Only 17% of the control group at Juniper selected the correct answer on the post-assessment, whereas 29% stated that the moon is covered by the earth's shadow and 17% stated the moon is covered by the sun's shadow.

Butternut's GSA Results

Four of the 16 questions on the GSA assessed students' GSV. The first question showed students the top, side, and front view of a figure and asked them to select the 3D figure that was represented from the given views. The second question required students to select the front view of a given figure. The third question asked students to select the correct response that would represent a cone sliced by a wooden board when viewed from above. The final question asked students to select the 3D figure represented from a given mat plan. An RM-ANOVA revealed a significant increase with a small effect size on the Geometric Spatial Visualization Domain for the experimental females $F(1, 119) = 5.174$, $p < .05$, partial $\eta^2 = .042$ and Mr. Land's females, $F(1, 26) = 9.194$, $p < .01$, partial $\eta^2 = .26$. The partial η^2 value of .26 indicated that approximately 26% of the gain in understanding GSV can be directly attributed to the NASA-based curriculum. The other 73% could be attributed to differential maturation, differential motivation, and so forth. As we analyzed student responses from pre to post, we noticed that some students correctly answered the questions on both pre- and post-assessments (correct–correct), whereas others incorrectly answered the questions on the pre-assessment but correctly answered the questions on the postassessment (incorrect–correct). For example, on question 1 (Table 7), 15% of Mr. Land's females correctly responded to the question on both the pre- and post-assessment. Through participating in the REAL curriculum, an additional 39% of the females had correct responses on the postassessment when they initially responded incorrectly on the pre. Although Mrs. Roling's males had a high percentage (40%) of correct–correct responses on this question, an additional 17.5% achieved the correct answer, whereas 40% maintained their knowledge after instruction. Question two was the most challenging for Mr. Land's males, with only 10% answering it correctly, whereas 58% of the females correctly responded. Overall, the experimental and control groups performed the best on question 3. From the data, we conclude that at the end of the Earth/Space unit, the experimental groups were more knowledgeable on GSV than the control group.

As we examine the data by race, the largest average gain from pre to post of the experimental group was by the African American students (21.87%), whereas the lowest average gain were Hispanics (2.44%), and whites had a gain of 5.44%. For the control group, the largest average gain from pre to post was by the African American students (25%), and the lowest average was a decrease of 25% from pre to post by the Asian and Native American students, whereas the white students had a slight gain of 8.33%. As we look at the average gain in general by whites and students of color, for the experimental group, the whites ($n = 124$) had a gain of 5.44% and the students of color ($n = 69$) had a gain of 3.23%. The average gain for the students of color ($n = 9$) and whites ($n = 15$) was the same for the control group (8.33%).

Juniper's GSA Results

The same four GSA questions were asked to students at Juniper Middle School. An RM-ANOVA revealed a

Table 7
Butternut's Percentage Correct–Correct (C–C) and Incorrect–Correct (I–C) From Pre to Post

Group	Q1 C–C (%)	Q1 I–C (%)	Q2 C–C (%)	Q2 I–C (%)	Q3 C–C (%)	Q3 I–C (%)	Q4 C–C (%)	Q4 I–C (%)
All experimental	27	24	35	15	49.5	13	31	24
Experimental males	28	23	35	16	50	13	35	20
Experimental females	26	26	34	14	49	13	27	28
Mr. Land males	10.5	37	5	5	26	16	5	26
Mr. Land females	15	39	31	27	38	27	11.5	35
Ms. Apple males	24	21	38	15	56	12	26	15
Ms. Apple females	26	23	36	15	51	5	26	30
Mrs. Rong males	40	17.5	47.5	22.5	59	13	59	23
Mrs. Rong females	32.5	22.5	35	5	52.5	15	40	20
All control	17	27.5	14	21	38	17	10	20
Control males	25	25	12.5	19	37.5	25	12.5	25
Control females	8	31	15	23	38	8	8	15

Table 8
Juniper's Percentage Correct–Correct (C–C) and Incorrect–Correct (I–C) From Pre to Post

GSA Group	Q1 C–C (%)	Q1 I–C (%)	Q2 C–C (%)	Q2 I–C (%)	Q3 C–C (%)	Q3 I–C (%)	Q4 C–C (%)	Q4 I–C (%)
All experimental	15	23	17	18	35	24	16	11
Experimental males	17	19	19	17	36	20	19	10
Experimental females	13	27	14	21	33	29	13	13
Mrs. Stevens males	9	27	18	14	45	14	14	9
Mrs. Stevens females	13	35	9	17	22	18	4	4
Mrs. Castle males	19	16	20	18	32	23	21	11
Mrs. Castle females	12.5	22.5	17.5	22.5	40	35	17.5	17.5
All control	21.5	26	21.5	11	58.5	18.5	32	11
Control males	30	18	21	12	61	18	36	12
Control females	12.5	34	22	9	56	19	28	9

significant increase with a medium effect size from pre to post for Mrs. Castle's experimental females, $F(1, 39) = 4.16$, $p < .05$, partial $\eta^2 = .096$, control females, $F(1, 31) = 4.59$, $p < .05$, partial $\eta^2 = .129$, and the control students of color, $F(1, 17) = 4.99$, $p < .05$, partial $\eta^2 = .23$. There were not any significant increases among the other subgroups. Unlike Butternut's results, the data for correct–correct and incorrect–correct responses revealed the control group had a stronger knowledge of GSV than the experimental group (Table 8). Question 4 was the most challenging for Mrs. Steven's experimental females with only 8% responding correctly, whereas 23% of her males correctly responded. Overall, the experimental and control groups performed the best on question 3, which is similar to the results found for Butternut.

As we examine the data by race, the largest average gain from pre to post of the experimental group was the Hispanic students (31.25%), whereas the lowest average gain showed a decrease for African Americans (5%) and Asians (6.26%), and whites had a slight gain of 2.40%. For the control group, the largest average gain from pre to post was the Hispanics (25%), and the lowest average was a decrease of 25% from pre to post by the Asians and a 16.66% decrease for African Americans, whereas the whites had a slight gain of 6.55%. As we look at the average gain in general by whites and students of color, for the experimental group, whites had a gain of 2.40% and the students of color had a gain of 1.12%. The average gain for the students of color for the control group was 0%, and the whites average gain was 6.55%.

Butternut's PSVT-Rot Results

PSVT-Rot was given to 213 experimental and 27 control students. A RM-ANOVA revealed a significant increase in the mean values with medium effect sizes from pre to post for the experimental group overall, $F(1, 212) = 15.855$, $p < .00$, partial $\eta^2 = .070$, and experimental females, $F(1, 119) = 15.390$, $p < .00$, partial $\eta^2 = .115$. Both the students of color, $F(1, 72) = 15.77$, $p < .001$, partial $\eta^2 = .18$, and white students, $F(1, 124) = 6.88$, $p < .05$, partial $\eta^2 = .05$, in the experimental group showed significant increase from pre to post. Significant increases were also achieved by Mr. Land's females, $F(1, 25) = 10.028$, $p = .004$, partial $\eta^2 = .286$, and Ms. Apple's males, $F(1, 36) = 14.75$, $p < .00$, partial $\eta^2 = .291$, and females, $F(1, 53) = 7.659$, $p = .008$, partial $\eta^2 = .126$. The experimental group's females outscored the males on both the pre and post on 65% of the assessment. The control group showed no significant increases.

As we examine the data by race, the largest average gain from pre to post of the experimental group was the African American students (10.56%) and Hispanic students (7.27%), whereas the lowest average was a decrease for Native Americans (2.50%), and whites had a gain of 3.84%. For the control group, the largest average gain from pre to post was the Native Americans (20%) and Hispanic students (17.5%), and the lowest average was a decrease of 5% from pre to post by the Asian students, whereas the white students had a slight gain of 1.07%. As we look at the average gain in general by whites and students of color, for the experimental group, the whites had a gain of 3.84% and the students of color had a gain of 4.54%. The average gain for the students of color for the control group was 16.54%, and the whites' average gain was 1.07%.

Juniper's PSVT-Rot Results

The PSVT-Rot was given to 134 experimental and 63 control students. The experimental and control groups showed no significant increases on the PSVT-Rot from pre to post. However, if we examine the data by race, Hispanic students had the largest average from pre to post for both the experimental (pre: 17.50% and post: 26.25%) and control groups (pre: 60% and post: 80%).

Conclusion

As we compared males and females between and within the experimental and control groups at Butternut Middle School, we found that experimental males and females achieved more significant gains from pre to post on GSV when compared with the control group. In fact, both the experimental males and females showed a significant increase from pre to post on all the science domains (i.e., Domains C, E, and G) and the Geometric Spatial Visualization Domain on the LPCI. This can be attributed to teacher instruction and the implementation of the NASA-based REAL curriculum. The REAL curriculum is designed to develop students' GSV through the use of 2D and 3D models, Stellarium (free planetarium software licensed under the GNU General Public License; www.stellarium.org), and moon journaling. Only the control females showed a significant gain from pre to post on Domain C and Geometric Spatial Visualization. The control females did not show any significant gains on Domain E and G, and the control males did not reveal any significant gains on any domains on the LPCI. At Juniper Middle School, the experimental males and females made significant gains on Domain E and Geometric Spatial Visualization, whereas the experimental males and the experimental group overall showed significant gains on Domain C. However, the control group overall and the control males only had significant increase from pre to post on Domain E and Geometric Spatial Visualization. Like Butternut, the gains for the experimental group can be attributed to teacher instruction and the NASA-based curriculum. Although Ms. Branson's (control group's teacher) instruction included more spatial modeling (2D and 3D) than we would expect from a control teacher using a somewhat traditional curriculum, her students' gains were limited only to the male students.

Unlike the results of Gallagher et al. (2000) and Casey et al. (2001), "the items that showed the largest gender differences favoring males were items requiring the creation of a spatial image or representation as well as the retention and manipulation of a given image" (Casey et al., 2001, p. 47), our data revealed the opposite. For example, on question 1 of the GSA, the students were given the top, side, and front views of a solid figure, and they had to determine which solid figure best represents the given views. On this question, the experimental males outscored the females on the pre assessment, but on the post, the females outscored the males at both Butternut and Juniper. However, the control males at Butternut and Juniper outscored the females on both the pre and post. Studies have shown that when females are given the opportunity to develop their spatial reasoning skills through purposeful 2D and 3D modeling experiences, gender differences are reduced or eliminated (Tzuriel & Egozi, 2010; Vasta, Knott, & Gaze, 1996; Wilhelm, 2009). It is interesting to note that when examining LPCI by race, the experimental groups' students of color had larger gains than the control groups' students of color at both Butternut and Juniper. Moreover, both the students of color and the white students in the experimental group revealed significant gains at both schools. At Butternut, the students of color and white students in the experimental group showed a significant increase from pre to post on all of the science domains. At Juniper, the students of color in the experimental group revealed significant gains on Domains E and G, and the white students in the experimental group showed significant increases on Domain E. But, only the white students in the control group at both schools showed significant increases. This could possibly be attributed to the NASA-based curriculum because it was designed to engage all students, whereas the regular Earth/Space unit taught in the control classrooms was based on a Eurocentric curriculum.

Another interesting finding from our study revealed the ways in which the students of color (whether experimental or control students) made great gains in their development of mental rotation reasoning, as evidenced in the GSA and PSVT-Rot results. The students of color, in both experimental and control groups, achieved higher overall gain scores on the PSVT-Rot than white students. This achievement has not been illustrated in the literature prior to this research study.

Limitations of the Study

We acknowledge that the control numbers in our quasi-experimental study were small. However, this is the first study that compares both race and gender differences on astronomical spatial content. Findings from our study warrant more focus on better ways to advance Earth space reasoning and how such development occurs by gender and race.

It is imperative that more spatial tasks are implemented in the curricula of both science and mathematics (Casey et al., 2001) that engages all students' geometric spatial development. Similar to Reynolds and Wheatley's (1997) results, we argue that students who have a well-developed spatial sense generally solve problems in more meaningful ways. Understanding how males, females, and racial groups' developmental knowledge of spatial content differs may lead to knowing how to better support spatial reasoning for both genders that will increase all students' scientific understanding.

Our study is significant not only for shedding more light on students' GSV development in the area of astronomy, but also providing evidence that mental rotation advancement (PSVT-Rot) can occur after participating in a lunar phases Earth/Space unit. This advancement occurred within both the control and experimental groups with students of color.

References

Baki, A., Kosa, T., & Guven, B. (2011). A comparative study of the effects of using dynamic geometry software and physical manipulatives on the spatial visualisation skills of pre-service mathematics teachers. *British Journal of Educational Technology*, *42*(2), 291–310. doi: 10.1111/j.1467-8535.2009.01012.x

Baxter, J. (1989). Children's understanding of familiar astronomical events. *International Journal of Science Education*, *11*, 502–513.

Bayraktar, S. (2009). Pre-service primary teachers' ideas about lunar phases. *Journal of Turkish Science Education*, *6*(2), 12–23.

Black, A. (2005). Spatial ability and earth science conceptual understanding. *Journal of Geoscience Education*, *53*(4), 402–414.

Black, A. A. (2004). Relationship of earth science misconceptions and conceptual understandings with three types of spatial abilities in university non-science majors. Paper presented at the National Association for Research in Science Teaching, Vancouver, Canada.

Bodner, G., & Guay, R. (1997). The Purdue visualization of rotations test. *The Chemical Educator*, *2*(4), 1–17.

Broadfoot, J. M. (1995). *Development of visuospatial abilities among undergraduate astronomy students* (Unpublished master's thesis). Curtin University of Technology, Perth, WA.

Campbell, P. B. (1989). So what do we do with the poor, non-white female? Issues of gender, race, and social class in mathematics and equity. *Peabody Journal of Education*, *66*(2), 95–112.

Casey, M. B., Colón, D., & Goris, Y. (1992). Family handedness as a predictor of mental rotation ability among minority girls in math-science training program. *Brain and Cognition*, *18*, 88–96.

Casey, M. B., Nuttall, R. L., & Pezaris, E. (2001). Spatial-Mechanical reasoning skills versus mathematics self-confidence as mediators of gender differences on mathematics subtests using cross-national gender-based items. *Journal for Research in Mathematics Education*, *32*(1), 28–57.

Coluccia, E., & Louse, G. (2004). Gender differences in spatial orientation: A review. *Journal of Environmental Psychology*, *24*, 329–340.

Dai, M. F. W. (1991). *Identification of misconceptions about the moon held by fifth and sixth-graders in Taiwan and an application of teaching strategies for conceptual change* (Doctoral dissertation). University Microfilms International, Athens, GA. (Order Number 9124300).

Danaia, L., & McKinnon, D. H. (2007). Common alternative astronomical conceptions encountered in junior secondary science classes: Why is this so? *Astronomy Education Review*, *6*(2), 32–53.

Delgado, A. R., & Prieto, G. (2004). Cognitive mediators and sex related differences in mathematics. *Intelligence*, *32*, 25–32.

Eals, M., & Silverman, I. (1994). The Hunter-Gatherer theory of sex differences: Proximate factors mediating the female advantage in recall of object arrays. *Ethology and Sociobiology*, *15*, 95–105.

Gallagher, A. M., De Lisi, R., Holst, P. C., McGillicuddy-DeLisi, A. V., Morely, M., & Calahan, C. (2000). Gender differences in advanced mathematical problem solving. *Journal of Experimental Child Psychology*, *75*, 165–191.

Halpern, D. F. (2007). Science sex and good sense: Why women are underrepresented in some areas of science and math. In S. J. Ceci & W. M. Williams (Eds.), *Why aren't more women in science? Top researchers debate the evidence* (pp. 121–130). Washington, DC: American Psychological Association.

Halpern, D. F., Benbow, C. P., Geary, D. C., Gur, R. C., Hyde, J. S., & Gernsbacher, M. A. (2007). The science of sex differences in science and mathematics. *Psychological Science in the Public Interest*, *8*(1), 1–51.

Huck, S. W. (2012). *Reading statistics and research* (6th ed.). Boston: Pearson.

Levine, S. C., Vasilyeva, M., Lourenco, F., Newcombe, N. S., & Huttenlocher, J. (2005). Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science*, *16*, 841–845.

Lindell, R., & Olsen, J. P. (2002). Developing the Lunar Phases Concept Inventory. American Association of Physics Teachers Summer Meeting.

Linn, M., & Petersen, A. (1985). Emergence and characterization of sex differences in spatial ability; meta-analysis. *Child Development*, *56*, 1479–1498.

Lippa, R. A., Collaer, M. L., & Peters, M. (2010). Sex differences in mental rotation and line angle judgments are positively associated with gender equality and economic development across 53 nations. *Archives of Sexual Behavior*, *39*(4), 990–997.

Manning, M. L. (1998). Gender differences in young adolescents' mathematics and science achievement. *Childhood Education*, *74*(3), 168–171.

McGee, M. (1979). Human spatial abilities: Psychometric studies and environmental, genetic, hormonal and neurological influences. *Psychological Bulletin*, *5*, 889–918.

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.

Mulholland, J., & Ginns, I. (2008). College moon project Australia: Preservice teachers learning about the moon's phases. *Research in Science Education*, *38*, 385–399.

NCTM. (2000). *Principles and standards for school mathematics*. Reston, VA: Author.

NCTM. (2002). *Navigating through geometry in grades 6–8*. Reston, VA: Author.

Parker, J., & Heywood, D. (2007). The earth and beyond: Developing primary teachers' understanding of basic astronomical events. *International Journal of Science Education*, *20*(5), 503–520.

Plummer, J., Wasko, K., & Slagle, C. (2011). Children learning to explain daily celestial motion. *International Journal of Science Education*, *33*(14), 1963–1992.

Reynolds, A. M., & Wheatley, G. H. (1997). A student's imaging in solving a nonroutine task. *Teaching Children Mathematics*, *4*(2), 100–104.

Sorby, S. A. (1999). Educational research in developing 3-D spatial skills for engineering students. *International Journal of Science Education*, *31*(3), 459–480. doi: 10.1080/09500690802595839

Tzuriel, D., & Egozi, G. (2010). Gender differences in spatial ability of young children: The effects of training and processing strategies. *Child Development*, *81*(5), 1417–1430.

Vasta, R., Knott, J. A., & Gaze, C. E. (1996). Can spatial training erase the gender differences on the water-level task? *Psychology of Women Quarterly*, *20*, 549–568.

Voyer, D., Voyer, S., & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin*, *117*(2), 250–270.

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., Toland, M. D., Cole, M., & Wilhelm, R. (2013). Evaluating middle school students' spatial-scientific performance in Earth-space science. Paper presented at the 222nd meeting of the American Astronomical Society in Indianapolis, Indiana (June 2–6, 2013).

Wilhelm, R., & Wilhelm, J. (2007). Realistic Explorations in Astronomical Learning. The REAL Curriculum.

Yair, Y., Schur, Y., & Mintz, R. (2003). A "thinking journey" to the planets using scientific visualization technologies: Implications. *Journal of Science Education and Technology*, *12*(1), 43–49.

Authors' Notes

Correspondence concerning this article should be addressed to Christa Jackson, School of Education, Iowa State University, 1688 Lagomarcino Hall, Ames, IA 50011, USA. E-mail: jacksonc@iastate.edu

Keywords: geometric spatial visualization, gender, middle grades.