Spatial Skills Training in Introductory Computing

Stephen Cooper
Stanford University
353 Serra Mall
Stanford, CA 94305
1.650.723.9798
coopers@stanford.edu

Karen Wang, Maya Israni
Stanford University
353 Serra Mall
Stanford, CA 94305
kywang, mayai@stanford.edu

Sheryl Sorby
The Ohio State University
2070 Neil Avenue
Columbus, OH 43210
sorby.1@osu.edu

ABSTRACT

This paper explores the question as to whether there is a relationship between a student's spatial abilities and her achievement in learning to program. After noting that there does seem to be a correlation, the paper explores the impact of trying to improve a student's spatial abilities. The paper reports on a preliminary study involving high school students. The study results suggest a correlation exists between receiving training in spatial skills and improved student performance in introductory computing. While the sample size in the study is small, this improvement appears to occur for students of different races/ ethnicities and across different socio-economic statuses.

Categories and Subject Descriptors

K.3.2 [**Computer and Information Science Education**]: Computer Science Education

General Terms

Measurement, Human Factors.

Keywords

Introductory computer science; spatial skills; training.

1.INTRODUCTION

Strong spatial skills are a predictor of success in many engineering disciplines [38, 29]. Spatial abilities have been linked to higher-level thinking, reasoning, and the creative process [1, 5, 34, 38]. But it is not clear exactly why and how spatial abilities lead to success. In conversations with many leading engineering educators, we have heard that the engineers think that spatial abilities may be linked to students' abilities to think at different levels of abstraction, an essential skill for engineers (and computer scientists) to develop.

Within engineering, studies have shown that women and those from low socio-economic status (SES) backgrounds have lower spatial abilities [4, 25], and, not surprisingly, are under-represented in many engineering disciplines. Fortunately, a good deal of work has been done, and students can be taught how to improve their spatial skills, which has been linked to improved retention within many engineering majors.

ICER '15, August 9--13, 2015, Omaha, Nebraska, USA.

DOI: http://dx.doi.org/10.1145/2766XXX.XXXXXXX

There have been several studies within computer science classrooms suggesting that there is a link between a student's spatial abilities and the student's ability to learn to program [11. 21, 22, 46]. We designed an experiment to see if we could replicate the results from these other studies. Beyond that, if we could replicate the correlation between spatial and programming abilities, we wanted to see what would happen if we tried to improve students' spatial abilities: Would the students have greater programming success?

We ran a pair of two-week summer workshops for rising twelfth grade students, targeting women and under-represented minorities. **Would we see any cognitive or affective differences if we exposed one group to spatial skills training while not teaching spatial skills to the other group?**

2.RELATED WORK

2.1.Spatial Skills Differences Between Sexes

Hyde [20] performed a meta-analysis on studies of males and females that occurred prior to 1974. In identifying 30 studies, she notes small but statistically significant differences between the visual-spatial abilities of males and females. Linn and Peterson [27] performed a meta-analysis of studies occurring from 1974-1982. They found large gender differences (with males scoring much higher) on measures of mental rotation. Many other studies note spatial abilities differences between males and females, with females having lower spatial abilities. While there is a difference of opinion between whether these differences appear prior to or after puberty (for example, Maccoby and Jacklin [28] provide evidence of differences appearing in adolescence while Newcombe et. al. [33] suggest male-female differences exist prior to adolescence), all of these studies do confirm these differences by the time students are adolescents.

2.2.Spatial Skills in Different Disciplines

Many researchers have tried to understand the extent to which differences in spatial abilities impact students across a range of STEM disciplines. Smith [38] conducted research in spatial visualization, identifying numerous careers for which spatial skills play an important role. Norman [34] found that a person's spatial skill level was the most significant predictor of success in his/her ability to interact with and take advantage of the computer interface in performing database manipulations. Barke [1] found that well developed spatial skills are essential for understanding basic and structural chemistry. Sorby [41] found that a person's spatial skills are related to his/her ability to effectively learn to use computer aided design software.

Some people might argue that with the multitude of data visualizations available today, the need for well-developed spatial skills has diminished, letting the computer do the visualizing. However, Cohen and Hegarty [5] found an individual's ability to manipulate/understand computer-based visualization of complex phenomena is correlated with her spatial skill level. In other words, students with poorly developed spatial skills do not understand the visualizations and do not really learn from them.

Gender and SES Differences in Spatial Skills: There is evidence to suggest that the 3-D spatial visualization skills of women lag significantly behind those of males. Theories for the cause of these differences include the belief that spatial ability is related to a male sex hormone [18], or that environmental factors are the primary reasons for male-female differences in spatial skill levels [8]. There are conflicting opinions as to whether differences on spatial performance between genders are linked to differences in mathematics performance. Tartre suggests that this may be the case [45], while Fennema and Sherman found that while there were few sex-related cognitive differences in mathematical abilities between males and females, there were differences in spatial visualization abilities between male and female students [9]. Fennema and Sherman's observations were echoed by Lindberg et al. [26], who did a meta-analysis of studies involving a much larger student population. Gender differences in 3-D spatial skills are likely due to a combination of several factors.

There has been little research on the differences in spatial abilities by race, but what has been done seems to confirm that there is a difference. Study [44] studied minority students at a historically black college. As part of her study on the impact of spatial skills training to improve retention, she noted that minority students' spatial skills were initially significantly behind those of students from a non-minority school.

Levine et al [25] found that the spatial skills for students from low SES groups were significantly lower than the skills for students from middle or high SES groups. Furthermore, there were no gender differences for students in low-SES groups, but there were significant gender differences for students from middle and high SES groups. However, spatial skills for the low SES group males were significantly lower than those for the females from the middle and high SES groups. Casey et al [4] also found significant differences in spatial skills, favoring students from middle or high SES groups. In the US, as under-represented minorities make up a significant portion of the students from the low SES groups, poorly developed spatial skills for these students could have serious implications for broadening participation in STEM.

There is a good deal of evidence to suggest that sketching 3-D objects is a significant factor in the development of these skills [2, 10, 31, 39, 40]. Gimmestad (now Baartmans) [13] conducted a study at Michigan Technological University and found significant gender differences in spatial abilities as measured by the Purdue Spatial Visualization Test: Rotations (PSVT:R) [14]. She also found a person's score on the PSVT:R was the most significant predictor of success in an engineering graphics course, of eleven variables tested. Design graphics courses are often among the first courses in which many first-year engineering students enroll. Students who have poorly developed spatial skills, particularly women, may become discouraged and drop out of engineering altogether if they struggle in their very first "engineering" course. In 1991, Baartmans [39] conducted a pilot study course for improving spatial skills. The results from her pilot study were promising and led to what has turned out to be a sustained journey in improving 3-D spatial skills for engineering students. In a longitudinal study conducted in 2000 [42], Sorby found that for students who initially demonstrated poorly developed spatial skills, enrollment in a spatial skills course improved student success in graphics courses by a half-letter grade (approximately 5 points in a 100 point system), and improved retention in their engineering majors.

2.3.Spatial Skills in Computing

Several researchers have explored the relationship between spatial skills and student programming ability/success. Fincher et al. [11] ran a study with 177 participants from eleven post-secondary educational institutions in Australia, New Zealand, and Scotland. They found a "small positive correlation" between scores in a spatial visualization task and programming marks, though attributed programming success to higher IQ components rather than to spatial skills. Jones and Burnett [21] conducted a study with 24 participants from a Masters course in the UK. They found participants with high spatial abilities completed code comprehension exercises faster than those with lower spatial abilities, along with a strong relation "between spatial ability and results in programming modules." A later study of 49 students [22] found a correlation between mental rotation skills and programming success. Mayer et al. [30] found, running a study with 57 college students in a course in Basic, "success in learning Basic was related to general intellectual ability, especially logical reasoning and spatial ability." Fisher, Cox, and Zhao [12], in a study with 30 undergraduate and graduate students with experience in Java, found "programmers use equivalently risky strategies for program comprehension and spatial cognition." Furthermore, they argued that "similar cognitive skills are used for spatial cognition and program comprehension/development." Webb [46] ran a study with 35 students aged 11 and 14, and found "spatial ability was the best predictor of knowledge of basic commands; and a combination of spatial ability and field independence best predicted scores on generated graphics programs" after learning Logo for one week.

It is important to note that while we were able to find several studies exploring the relationship between spatial and programming abilities (and providing evidence that a correlation exists), we were unable to find any studies where the researchers attempted to improve student spatial abilities as part of the study, as a means to hopefully improve programming ability.

3.METHODOLOGY

3.1.Procedure

In this study, the target students were rising high school seniors who had minimal previous experience in computer science. (Rising high school seniors are those students about to enter 12th grade, the last year of secondary school. Most are 17 years old.) We have been working with several high schools in Oakland and San Jose, as well as throughout the San Francisco Bay Area. High school math and science teachers from select schools were invited to ask their strong math and science students to apply to spend two weeks in an introductory computing workshop at Stanford University. We ran two workshops consecutively in summer 2014. The first workshop was the control group, and the second was the treatment group.

Both groups met from 9 to 5 each day for two weeks. We covered approximately half of Stanford's introductory computing class (for majors). The first two days introduced programming using Alice, while the remaining days covered programming using Java. In addition to computing lectures and labs, each group had 3-4 invited speakers from industry and academia visit (who spoke about different aspects of computing). Each day at lunch, 3-5 college interns, recent alumni working in Silicon Valley, and/or graduate students joined the high school students to discuss different computing career options. Each group had a field trip to a local technology company.

To mitigate covariant factors, both groups were given a 45 minute activity that was separate from the daily planned lecture schedule. For the control group, each day began with a review of the previous day's material led by one of the student teaching assistants for 45 minutes. These sessions consisted of group problem-solving exercises where the teaching assistant guided students through coding tasks targeting specific concepts. This is similar to what is done by Stanford's undergraduate teaching assistants in weekly sessions as part of Stanford's CS1 course.

Similarly, for the treatment group, one of the student teaching assistants led a 45 minute presentation/activity on developing spatial skills. While the content presented was different for each group (computer science material vs. spatial skills material), the activity flow for both was the same. In particular, rather than working on computers, the activities, for both the control and treatment groups, involved presentations on a white board followed by pencil and paper exercises. The rest of the day for both groups was spent with short lectures and longer laboratory periods where the students worked individually and in pairs on various programming activities.

Spatial skills: The spatial skills development sessions were held immediately after breakfast and before starting the day's computer science lesson. During each session, the topic was presented with a slideshow. The 5-7 minute slideshow demonstration introduced the topic by defining terms and showing examples related to the topic. After the demonstration, students completed worksheets that included sketching, matching, and measuring. Students typically completed the exercises individually and confirmed answers with one another upon completion; a few students also assisted their tablemates (students were arranged in pairs, with a table holding three student pairs) in answering the more difficult worksheet questions. Students completed the majority of the exercises on the worksheets. Students used building blocks to assist in completing the worksheets. Often, students built physical models of the worksheets' exercises using the blocks to sketch and measure from. These blocks were simple Lego-like cubes. At the end of each session, the workshop assistants collected all worksheets, pencils, and blocks. A researcher graded all worksheets immediately after the workshop ended. Graded worksheets were not returned to the participants. Students also completed Alice world exercises that taught double rotations and revolutions of objects. These exercises used animations of objects to further reinforce these more difficult topics.

Spatial Skills Curriculum: The curriculum of the daily spatial skills training workshops was adapted from Sorby's workbook [43] and curriculum. It focused on the teaching of mental rotations. Topics covered included:

- Isometric drawings (2 days): These drawings depict a 3-D object on a 2-D sheet of paper. An isometric view is the view looking down a diagonal of a cube that is part of the object.
- Orthographic drawings (1 day): These drawings depict "the faces of the object straight on or parallel to the viewing plane," including top, side, and front views [43].
- Single and double rotations of objects (3 days): This transformation includes turning an object about a straight line, or axis of rotation.
- Reflections and symmetry of objects (1 day): The reflection transformation happens when an object is reflected across an entire plane. An object is symmetrical if a plane can cut the object into two halves that are mirror images of each other.
- Surfaces and solids of revolution (1 day): These shapes are "created by revolving a set of 2-D curves about a coordinate axis" [43].

Other topics included in this workbook but excluded from our curriculum include combining solid objects, inclined and curved surfaces, paper folding, cutting planes, and cross sections. These topics were excluded due primarily to time limitations.

3.2. Data Measures

3.2.1. Assessment Instruments

Four instruments were used. The first instrument was a collection of demographic information about the students. The second instrument included eight Likert scale questions we created concerning student confidence towards learning computing as well as gender roles concerning computing. The Revised Purdue Spatial Visualization Test (Revised PSVT:R) [49] was used to assess the spatial skills ability of participants. This test measures the ability to complete mental rotations, a crucial indicator of spatial ability. The fourth instrument was an adapted version of the AP Computer Science Test administered in 2009 [6]. The AP tests are administered by the College Board, and most colleges in the US give college credit for high performance by high school students who are subsequently admitted to their college. It drew sixteen multiple choice questions from the official AP test.

3.2.2. Data Collection

All appropriate human subject procedures were followed in this study. Both the treatment and control groups completed the demographic instrument (as a pre-test), the attitudes instrument (as both a pre-test and as a post-test), the spatial skills instrument (as a pre- and post-test), and the AP CS test (as a pre- and post-test). The pre-tests were given on the first day of the workshop. The post-tests were given on the last (tenth) day of the workshop. For the control group, we erroneously did not provide a time limit for the Revised PSVT:R exam. (Students should have been given 20 minutes to complete the exam.) To be consistent, we did not provide a time limit for the treatment group. Most students took 20-25 minutes to complete the spatial skills test. Likewise, we did not provide a time limit for completing the 16 AP CS multiple choice questions. Most students completed this exam within 20 minutes. Students completed all instruments online.

3.2.3. Data Analysis

One of the challenges of this project was working with a small sample size (19 students per workshop, 38 students in total) due to the limitations imposed on our resources. A small sample size would be skewed by the presence of outliers, and could invalidate the assumption of a t-test that the variables are normally distributed. Thus, small sample sizes bring up the need for non-parametric tests, which make no assumptions about the probability distributions of variables.

Student's t-tests were used to analyze the aggregate CS, spatial, and attitude data for each session, and non-parametric Mann-Whitney and Kruskal-Wallis tests were used to analyze the categorical data, divided by socio-economic status (SES) and race/ethnicity. We did not divide students by gender; since the workshop was geared towards females, each session only had two males. While our sample size for each session ($n=17$ to 19) was relatively small, the Shapiro-Wilk normality test [37] showed that the aggregate datasets followed normal distributions at an alpha level of 0.01. When each session was further divided into SES and race/ethnicity categories, however, some categories had sample sizes too small to apply reasonable analyses of normality, and thus warranted non-parametric testing. Non-parametric tests such as Mann-Whitney and Kruskal-Wallis are useful for comparing two and three populations, respectively, where the sample sizes are small and the population distributions are unknown.

3.3.Student Population

The targeted students were those under-represented rising seniors who had a strong background in mathematics and/or science, with little to no previous programming experience. For both the control and the treatment groups, twenty students were selected. From partner high schools, we invited math and science teachers to have 2-4 of their students apply. Each workshop had two males, both of whom were of minority groups. Nineteen students completed each workshop. We allowed students to request one workshop (students did not know which was the control session and which was the treatment session), and tried to honor their preferences. This did lead to the situation where the two workshops were not perfectly balanced in terms of demographics or SES.

Table 1. Student demographics

	Hispanic	African American	Asian + Caucasian
Control	9	3	7
Treatment	5	4	10

We decided not to separate by sex, as each group (control and treatment) had only two males. This was somewhat disappointing given that much of the existing spatial skills literature has focused on differences between male and female spatial abilities.

We decided to not separate race from ethnicity. Many of the Hispanic students struggled to identify what their race was, as they identified as being Latina/o. Other studies have also shown that Hispanics are often unsure of what race they are [23]. We grouped Asian and Caucasian together because there was only one non-Hispanic Caucasian student in the control group.

Table 2 shows the socio-economic breakdown of the students. For the treatment group, we collected data on students' mothers' highest level of education, as a person's mother's educational level is a strong predictor of SES [24, 7, 36]. However, this SES data was not directly collected for the control group. We did know what high schools the students attended. The school status level was calculated based on each school's Academic Performance Index (API) score [3]. The API score is a value between 200-1000, with a target of 800 for every high school. We divided high school groups into 200-750 (lower SES schools students attended had APIs ranging from 450-700), 750-850 (middle SES schools students attended had APIs ranging from 780-820), and 850-1000 (upper SES schools students attended had APIs ranging from 880-910). Using this division, we had a close match (all but two students for the treatment group) with the students' self-reported mother's highest level of education. We were thus confident in using the school attended as a proxy for SES level. In our analyses, we combined the middle and upper SES groups, as the treatment group only had two middle class students.

Table 2. Student Socio-Economic Status

	Lower	Middle	Upper
Control	10	4	5
Treatment	8	2	9
Mother's Education Level	**High school, or lower**	**Some college**	**College graduate or Graduate school**
Treatment	6	3	10

4.RESULTS AND DISCUSSIONS

4.1.Spatial Skills/Programming Correlation

Our first step was to verify or refute the previous research on the correlation between spatial and programming skills [12, 21, 22, 30, 11, 46]. Is there a relationship between students having high spatial abilities and better performance in programming?

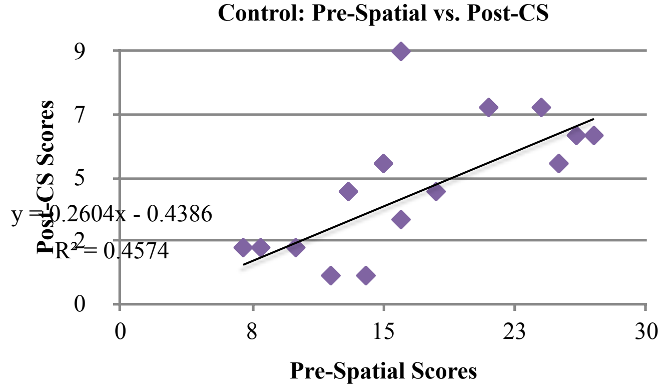

Figure 1. Control group pre-spatial versus post-CS scores

Figure 1 shows the graph of student scores on the spatial exam given at the start of the workshop versus their scores of the AP CS exam given at the end of the workshop. The control group scores confirm a positive correlation (R^2=0.46, R^2=0.65 upon removal of the one outlier who scored high on the post-CS exam, but low on the pre-spatial exam) between pre-spatial and post-CS scores. We can infer the relationship between spatial and programming abilities accounts for sixty-five percent (upon removing the outlier) of the variability in the data set. In the control group, it seems we could have predicted a student's CS performance at the end of the workshop from her original spatial score.

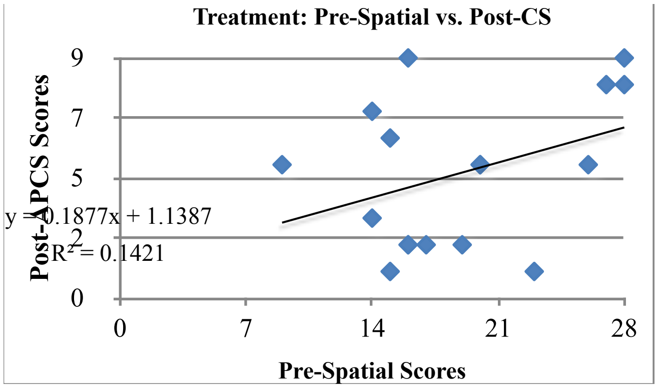

Figure 2. Treatment group pre-spatial versus post-CS scores

We also plotted results from the control group post-spatial versus post-CS scores and the treatment group pre-spatial versus post-CS and post-spatial versus post-CS. The graph of the control group's post-spatial versus post-CS scores was similar to Figure 1 (which is not surprising as we did not teach the control group spatial skills, and their post-spatial scores were similar to their pre-spatial scores). For the treatment group, as shown in Figure 2, the correlation between the pre-spatial and post-CS scores was low. This could suggest that the treatment group's spatial skills training impacted students' programming abilities.

When we graphed the treatment group's post-CS versus post-spatial scores (as shown in Figure 3), the correlation was again low. We are not sure why there was not a stronger correlation, as we had expected a stronger correlation between the students in the treatment group who scored well on the spatial skills exam as well as on the CS exam. One possible reason for this seeming anomaly is stereotype threat and/or testing bias. All six of the students who scored a 1 or a 2 on the CS test (all of whom scored at least a 16 on the spatial exam) were low-income and/or minority (African American or Hispanic). As part of the AP exam, the CS test looked and felt more like a standardized exam than did the spatial skills test. Students with less exposure to and experience with such AP exams might not have developed as efficient test-taking strategies.

Another possible reason has to do with the CS exam itself. The focus of the workshop was on code writing, whereas the multiple choice portion of the AP CS exam was on code reading. The AP CS multiple choice exam thus required a greater amount of inference ability for the students. As this code reading ability was not practiced at all during the workshop, students likely felt less confident concerning the taking of this exam. Certainly, we could have chosen to use a sample free-response question from the AP CS exam instead – this would more directly have measured a student's code writing abilities. However, one of the authors of this paper has previously worked at grading the AP CS exam. (College and high school instructors are hired each summer by the College Board to grade the free-response questions.) Even with a detailed grading rubric, much of the first day of AP CS grading involves calibrating and agreeing upon the correct interpretation of the rubric by the graders. This works well when the graders are able to use the first few hundred exams to agree upon how to equitably grade the remaining 20-30,000 exams. This may well have worked less well for us, when we had fewer than 40 exams to grade, and a lack of inter-grader reliability would likely have left us with an inability to interpret the results of our study. Section 5.2 explores this decision further.

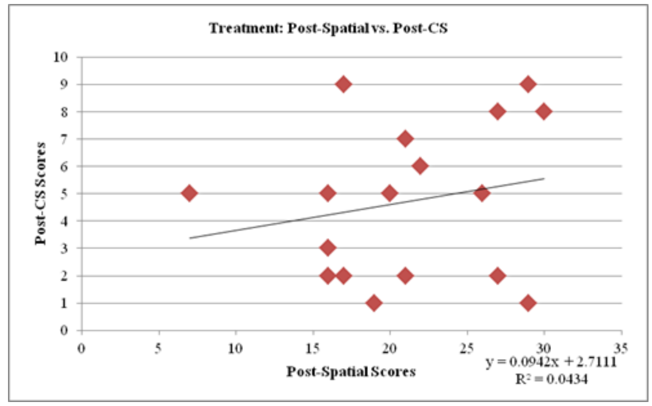

Figure 3. Treatment group post-spatial versus post-CS scores

4.2.Student Performance

We were interested to see how much, on average, students improved in each group, and whether or not this improvement was significant. We looked at student performance in the pre and post-tests for spatial skills and in CS (as measured by the AP exam). We considered the mean CS and spatial deltas of each group, where a delta is the difference between the pre and post-score.

Using paired t-tests (comparing mean pre-test to mean post-test scores), the gains in CS scores were not significant for the control group, but neared significance for the treatment group ($p=0.07$). Table 3 can be read noting that the treatment group improved by an average score of 1.06, and that this was significant at the $p=0.07$ level. Moreover, when the scores for the subset of CS questions with high item discrimination were considered, the gains in CS were significant for the treatment group ($p<0.05$). The gains in spatial scores were significant for the treatment group ($p<0.005$).

Item Discrimination (ID): In order to factor out random guessing, we looked at student CS performance on the subset of six questions that had high item discrimination. Because the AP CS multiple choice exam included questions from the year-long AP class (of which we covered about half), there were questions on the exam on material which the students had either not seen or only covered slightly. ID is an index of a test item's effectiveness at discriminating those who know the content from those who do not. ID values ≥ 0.2 are desired [35]. The label CS ID* denotes this subset of questions on the CS test that had an ID of equal to or greater than 0.3. We chose a value of 0.3 rather than 0.2 as the questions that had a value above 0.2 also had a value above 0.3.

Table 3. Mean deltas (p-values in parentheses)

	Control Δ	**Treatment Δ**
CS	0.50 (0.23)	1.06 (0.07)
CS ID*	0.72 (0.11)	0.82 (p<0.05)
Spatial	0.89 (0.11)	2.63 (p<0.005)

4.3.Relationships between Categories

Next we divided each session into categories based on SES and race/ethnicity, and considered the difference in CS and spatial deltas between these categories. Mann-Whitney U-tests were used for the two SES categories (low and middle/high), and Kruskal-Wallis tests were used for the three race/ethnicity categories (African American, Hispanic and Asian/Caucasian). Generally speaking, Mann-Whitney U-tests are used to compare two populations, and Kruskal-Wallis tests are used to compare three or more populations. Ordinal tests such as Mann-Whitney and Kruskal-Wallis take raw scores and rank them from lowest to highest (in our case from 1 to 19, where 19 is the highest), and are particularly useful when analyzing smaller sample sizes, where there may or may not be a normal distribution of the data. A mean rank is simply the mean of the all the ranks in a group. In Table 4, we report the mean scores for consistency, and include the mean ranks in parentheses.

4.3.1.*SES*

In the control group there was a significant difference in CS deltas (the student's score on the CS post exam minus her score on the CS pre exam) by SES ($p<0.001$). Lower SES students in the control group had a significantly lower CS delta than their middle and upper SES counterparts. This meant that on average, lower SES students in the control group did not improve in their CS content knowledge as much as their higher SES counterparts by the end of the workshop; in fact, the lower SES students' raw scores decreased slightly, which we attribute to random guessing. For the treatment group, however, low-SES students' CS deltas were positive and large enough to create a non-significant difference in CS deltas between them and their higher SES peers. This meant that on average, there were not significant differences between the learning gains from lower SES students in the treatment group and their higher SES counterparts. For spatial deltas, there were no significant differences within either the control or treatment group. In the table below, we show the results of performing a Mann-Whitney U-test between the two categories. The mean scores are shown, and the mean ranks are shown in parentheses. The U-statistics, also shown, are used to compute the p-values.

Table 4. Differences by SES

Control	Lower	Middle + Upper	U	p
CS Δ	-1.56 (5.5)	2.56 (13.5)	4.5	p<0.001
CS ID* Δ	-0.78 (5.3)	3.00 (13.7)	3.0	p<0.001
Spatial Δ	1.20 (10.7)	0.56 (9.2)	38.0	0.28
Treatment	**Lower**	**Middle + Upper**	**U**	**p**
CS Δ	0.57 (8.4)	1.40 (9.5)	30.5	0.33
CS ID* Δ	0.57 (8.1)	1.09 (9.60)	29.0	0.28
Spatial Δ	2.38 (9.6)	2.82 (10.3)	41.0	0.40

This data analysis suggests that the spatial skills training helped the low-SES students, so that their CS knowledge gain was not distinguishable from the middle/high-SES students. In other words, in the group given spatial skills training, the lower SES students improved just as much as the higher SES students did in CS. This was not true in the control group, where there was a statistically significant distinction between the high-SES and low-SES students in terms of improvement in CS. Coupled with the results in Table 3, which shows that there was a statistically significant CS learning gain for the treatment group, this suggests that spatial skills training benefitted students from all SES levels.

4.3.2.*Race/Ethnicity*

In the control group we found a significant difference in CS deltas, with Hispanic students having the lowest CS delta ($p<0.01$). Since a Kruskal-Wallis test does not say anything about where the difference occurs, in order to determine which pairs of groups differ, we performed a Mann-Whitney U-test and then used a Bonferroni correction. Running a Mann-Whitney U-test and then using a Bonferroni correction showed that the significant differences were between Hispanic and Asian/Caucasian students.

But, in the treatment group, there was no significant difference. For spatial deltas, there was no significant difference within either the control or treatment group. We show the results of performing a Kruskal-Wallis test between the three categories, showing the mean scores, mean ranks in parentheses, and p-values.

Table 5. Deltas by race/ethnicity

Control	Hispanic	African American	Asian + Caucasian	p
CS Δ	-1.63 (5.3)	0.33 (9.8)	3.00 (14.2)	p<0.01
CS ID* Δ	-0.88 (5.8)	0.00 (8.2)	2.86 (14.4)	p<0.01
Spatial Δ	1.00 (10.2)	1.67 (11.7)	0.43 (9.0)	0.78
Treatment	**Hispanic**	**African American**	**Asian + Caucasian**	**p**
CS Δ	1.20 (9.1)	-0.67 (6.5)	1.56 (9.5)	0.62
CS ID* Δ	0.80 (8.7)	0.00 (7.0)	0.80 (9.2)	0.69
Spatial Δ	1.80 (8.9)	4.00 (12.5)	2.50 (8.7)	0.59

One interpretation of this data analysis suggests that the spatial skills training especially helped the Hispanic students, so that their CS knowledge gain was not distinguishable from the other students. An alternate interpretation is that Hispanic students did better at the same time that Asian and Caucasian students did comparatively (as compared to the control group) less well. At a minimum, a larger study seems warranted.

4.4.Attitudes

Students rated their attitudes before and after the workshop on three categories: perceived programming experience, confidence in their ability to learn programming, and notions about gender performance in CS. Gains in perceived programming experience and confidence in their ability to learn programming were significant for both groups ($p<0.05$), which is not often the case for students receiving their first programming course [32]. In both control and treatment groups, students scored high (24.1 and 23.2 out of 25, respectively) on notions about gender performance to begin with, so there was no significant change in gender scores.

Lower SES students reported lower confidence before and after the workshop than middle and upper SES students in the treatment group. In the control group, lower SES students reported higher perceived programming skills at the end of the workshop ($p<0.05$). African American and Hispanic students reported higher perceived programming skills than Asian/Caucasian students at the end of the workshop in the treatment group ($p<0.05$).

5. ANALYSIS AND CONCLUSION

5.1. Student Learning

Despite the fact that the treatment workshop contained less CS content (replacing the first 45 minutes of CS review with spatial skills training for eight of the ten days of the workshop), students in the treatment section did better on the CS instrument as well as performing better on the spatial skills instrument at the end of the workshop. The students also had greater gains with respect to their confidence and their perceived programming experience. Note that we are not claiming causation, but rather an interesting correlation.

5.2. Covariant Factors/Issues

Several other factors may have accounted for the changes that we saw. The first was that the instructors for the two workshops were different. It may be the case that the treatment instructor explained the material more clearly to the students. That said, the lead author of this paper taught the control section. He has nearly 20 years teaching experience at the college level, and has taught more than 1000 high and middle school students as part of more than 20 previous outreach efforts. The instructor for the treatment group has been teaching undergraduates for two years. This was his first time teaching a high school student group. It would have been better to have the same instructor teach both groups, but there is a non-trivial physical toll taken teaching a summer workshop for two weeks, eight hours a day.

Another possible factor was that the demographics of the student groups varied across the two workshops. The control group consisted of more students from lower SES backgrounds, and had more African American and Hispanic students.

Additional concerns lie in only using standardized tests as instruments of measure. Standardized tests may favor certain income or race/ethnicity groups while negatively impacting others due to a variety of factors including stereotype threat, difference in standardized testing experience and coaching, among others.

We did not use a validated (or even a well-established) attitudes instrument, such as Hoegh [19] or Wiebe [48] or McKlin [16]. This was done because the validated instruments are longer, and we were worried about "survey overload." Using a validated instrument would have given us greater confidence regarding the student attitude results.

We did not set a time limit on the PSVT:R exam. It is not clear to us what the impact of a 20 minute time limit would have had.

Using the multiple choice section of the AP CS exam tested student ability to interpret and trace code, rather than the ability to write code. The focus of the workshop was on having the students learn to write code. As noted above (in section 4.1), we could have chosen one or two of the free response questions from the AP CS exam. Asking students to write a method or program would have more closely matched what we were asking them to do during the workshop. However, it would have been harder to fairly grade those students whose responses were not perfectly correct, as the sample AP CS rubrics (or any rubric we could have developed to grade the free response questions) would have been dependent on having multiple graders agree upon interpretations. And we were concerned about this given our small sample size.

Hattie [17] conducted a meta-analysis of more than 800 studies. He found several factors that influenced student achievement in schools. These included the home (and in particular, the role parents play), the school environment including the classroom climate and peer influences, the teacher, the curriculum, and the pedagogy used. Any of these factors could be at play in our setting.

There has been a good deal of study as to how student attitudes are related to their successes and lack of successes in STEM. For example, in her meta-analysis, Weinburgh [47] found significant differences in attitudes towards science, with males having a significantly more positive view of science than females. And Gunderson et al. [15] explored the role that adults play in trying to understand why females have more negative attitudes towards math than males. We did not see particularly interesting attitudes differences (as noted above in section 4.4). However, we did not study this possible covariant factor at depth.

Finally, we administered all exams online. Due to our error, students were not provided with scratch paper. This would likely have been useful for the AP CS exam, to allow students to trace on paper the sample code they were reading.

5.3. Concluding Remarks

We have much to learn about how best to teach students to improve spatial skills, and how best to incorporate spatial skills training into existing curricula/courses. The results from our pilot study are positive. Our results indicate that it may be possible to improve spatial skills in a short amount of time, which can be promising for students with low spatial skills. We recommend further exploring the relationship between spatial skills and programming, and how best to teach spatial skills to improve student programming abilities. If results of our pilot study can be replicated and expanded upon by others, it would seem worthwhile to investigate why this relationship seems to exist, to better understand the nature of students learning to program and to develop their spatial abilities.

One of the most important areas for improvement would be to acquire a larger sample size of students, which would allow for more robust statistical analyses and avoid some of the roadblocks that we encountered. We also intend, in future interventions, to use a free response question from a previous AP CS exam to measure student CS content knowledge. We believe a free response question will better reflect the content covered during the workshop, where the focus is more on code writing than code reading.

6. ACKNOWLEDGMENTS

Thanks to S-Y Yoon who allowed us to use the Revised PSVT. Thanks to M. Sahami for his insightful comments on an earlier draft of this paper.

7. REFERENCES

1. Barke, H.-D. & Engida, T. 2001. *Structural Chemistry and Spatial Ability in Different Cultures*. Chemistry Education: Research and Practice in Europe, 2(3): p. 227-239.
2. Bowers, D. & Evans, D. 1990. The role of visualization in engineering design. *Proceedings of the NSF Symposium on Modernization of the Engineering Design Graphics Curriculum*, Austin, TX, 89-94.
3. California Department of Education. Academic performance index. http://www.cde.ca.goc/ta/ac/ap/ Accessed 8/28/2014
4. Casey, B. M., Dearing, E., Vasilyeva, M., Ganley, C. and Tine, M. 2011. Spatial and numerical predictors of measurement performance: The moderating effects of community poverty and gender. *Journal of Educational Psychology*, 103: 296–311.

5. Cohen, C. & Hegarty, M. 2007. Individual differences in use of external visualization to perform an internal visualization task. *Applied Cognitive Psychology*, 21, 701-711.
6. College Board. (2010). *2009 ap computer science a released exam*. New York: The College Board.
7. Fahey, T., Keilthy, P. & Polek, E. 2012. *Family relationships and family well-being: A study of the families of nine year-olds in ireland*. Dublin: University College Dublin.
8. Fennema, E. & Sherman, J. 1977. Sexual stereo-typing and mathematics learning. *Arithmetic Teacher*, 24(5), 369-372.
9. Fennema, E., & Sherman, J. 1977. Sex-related differences in mathematics achievement, spatial visualization and affective factors. *American educational research journal*, 14(1), 51-71.
10. Field, B. W. 1994. A course in spatial visualization. *Proceedings of the 6th International Conference on Engineering Design Graphics and Descriptive Geometry*, Tokyo, Japan, 257-261.
11. Fincher, S., Baker, B., Box, I., Cutts, Q., de Raadt, M., Haden, P., Hamer, J., Hamilton, M., Lister, R. & Petre, M. 2005. *Programmed to succeed? A multi-national, multi-institutional study of introductory programming courses* (University of Kent, Computing Laboratory Technical Report 1-05). http://www.cs.kent.ac.uk/pubs/2005/2157/ Accessed 8/25/2014
12. Fisher, M., Cox, A. & Zhao, L. 2006. Using sex differences to link spatial cognition and program comprehension. In *Proceedings of the 22nd IEEE International Conference on Software Maintenance (ICSM)*, 289 - 298.
13. Gimmestad, B. J. 1989. Gender differences in spatial visualization and predictors of success in an engineering design course. *Proceedings of the National Conference on Women in Mathematics and the Sciences*, St. Cloud, MN, 133-136.
14. Guay, R. B. 1976. *Purdue spatial visualization test: Rotations*. West Lafayette, IN, Purdue Research Foundation.
15. Gunderson, E. A., Ramirez, G., Levine, S. C., & Beilock, S. L. 2012. The role of parents and teachers in the development of gender-related math attitudes. *Sex Roles*, 66(3-4), 153-166.
16. Guzdial, M., Ericson, B., McKlin, T. & Engelman, S. 2012. A statewide survey on computing education pathways and influences: factors in broadening participation in computing. In *Proceedings of the ninth annual international conference on International computing education research (ICER '12)*, 143-150.
17. Hattie, J. 2013. *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. Routledge.
18. Hier, D.B. & Crowley, Jr., W.F. 1982. Spatial ability in androgen-deficient men. *New England Journal of Medicine*, 306(20), 1202-1205.
19. Hoegh, A. & Moskal, B. 2009. Examining science and engineering students' attitudes toward computer science. In *Proceedings of the Frontiers in Education Conference*, San Antonio, Texas (6 pages).
20. Hyde, J. S. 1981. How large are cognitive gender differences? A meta-analysis using ω^2 and d. *American Psychologist*, 36(8), 892-901.
21. Jones, S. & Burnett, G. 2007. Spatial skills and navigation of source code. In *Proceedings of ITiCSE 2007*, 231-235.
22. Jones, S. & Burnett, G. 2008. Spatial ability and learning to program. *Human Technology*, 4(1), 47-61.
23. Krogstad, J. & Cohn, D. 2014. U.S. census looking at big changes in how it asks about race and ethnicity. Available from: http://www.pewresearch.org/fact-tank/2014/03/14/u-s-census-looking-at-big-changes-in-how-it-asks-about-race-and-ethnicity/ Accessed 8/29/2014
24. Liberatos, P., Link, B. & Kelsey, J. 1988. The measurement of social class in epidemiology. *Epidemiologic Reviews*, 10, 87-121.
25. Levine, S. C., Vasilyeva, M., Lourenco, S. F., Newcombe, N. S. & Huttenlocher, J. 2005. Socioeconomic status modifies the sex difference in spatial skill. *Psychological Science. American Psychological Society*, 16(11) 841-845.
26. Lindberg, S. M., Hyde, J. S., Petersen, J. L., & Linn, M. C. 2010. New trends in gender and mathematics performance: a meta-analysis. *Psychological Bulletin*, 136(6), 1123-1135.
27. Linn, M. C. & Petersen, A. C. 1985. Emergence and characterization of sex differences in spatial ability: A meta-analysis. *Child Development*, 56, 1479-1498.
28. Maccoby, E. E. & Jacklin, C. N. (Eds.). 1974. *The psychology of sex differences (Vol. 1)*. Stanford, CA: Stanford University Press.
29. Maier, P. H. (1994). *Raeumliches vorstellungsvermoegen.* Frankfurt a.M., Berlin, Bern, New York, Paris, Wien: Lang.
30. Mayer, R. E., Dyck, J. L. & Vilberg, W. 1986. Learning to program and learning to think: What's the connection? *Communications of the ACM*, 29, 605–610.
31. McKim, R. H. 1980. *Experiences in visual thinking*. Boston, MA: PWS Publishers.
32. Moskal, B., Lurie, D. & Cooper, S. 2004. Evaluating the effectiveness of a new instructional approach. *SIGCSE Bull.* 36, 1 (March 2004), 75-79.
33. Newcombe, N., Bandura, M. M., & Taylor, D. G. 1983. Sex differences in spatial ability and spatial activities. *Sex Roles*, 9(3), 377-386.
34. Norman, K.L. 1994. Spatial visualization-A gateway to computer-based technology. *Journal of Special Educational Technology*, XII(3), 195-206.
35. Office of Measurement and Evaluation of Teaching, University of Pittsburgh. Item difficulty item discrimination. http://www.omet.pitt.edu/docs/OMET%20Test%20and%20Item%20Analysis.pdf Accessed 8/29/2014
36. Ou, S-R. & Reynolds, A. 2008. Predictors of educational attainment in the chicago longitudinal study. *School Psychology Quarterly*, 23(2), 199–229.
37. Shapiro, S. S., & Wilk, M. B. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 591-611.
38. Smith, I. M. 1964. *Spatial ability-Its educational and social significance*. London: University of London.
39. Sorby, S. A. & Baartmans, B. J. 1996. A course for the development of 3-D spatial visualization skills. *Engineering Design Graphics Journal*, 60(1), 13-20.
40. Sorby, S. A. & Gorska, R. A. 1998. The effect of various courses and teaching methods on the improvement of spatial ability. *Proceedings of the 8th International Conference on Engineering Design Graphics and Descriptive Geometry*, Austin, TX, 252-256.
41. Sorby, S. A. 2000. Spatial abilities and their relationship to effective learning of 3-d modeling software. *Engineering Design Graphics Journal*, 64(3), 30-35.

42. Sorby, S. A. 2001. A course in spatial visualization and its impact on the retention of women engineering students. *Journal of Women and Minorities in Science and Engineering*, 7(2), 153-172.
43. Sorby, S. A. 2011. *Developing Spatial Thinking Workbook*. Boston, MA: Cengage Learning.
44. Study, N. E. 2006. Assessing and improving the below average visualization abilities of a group of minority engineering and technology students. *Journal of Women and Minorities in Science and Engineering*, 12 (4) 363-374.
45. Tartre, L. 1990. Spatial skills, gender, and mathematics. In E. H. Fennema & G. C. Leder (Eds.), *Mathematics and Gender*, (pp. 27-59). New York, NY: Teachers College Press.
46. Webb, N. 1984. Microcomputer learning in small groups: Cognitive requirements and group processes. *Journal of Educational Psychology*, 76, 1076–1088.
47. Weinburgh, M. 1995. Gender differences in student attitudes toward science: A meta-analysis of the literature from 1970 to 1991. *Journal of Research in Science Teaching*, 32(4), 387-398.
48. Wiebe, E. N., Williams, L., Yang, K. & Miller, C. 2003. Computer Science Attitude Survey. (Report No.: NCSU CSC TR-2003-1) NC State University, Raleigh, NC.
49. Yoon, S. Y. (2011). *Revised purdue spatial visualization test: Visualization of rotations (Revised PSVT:R)* [Psychometric Instrument]