122nd ASEE Annual Conference & Exposition
June 14 - 17, 2015
Seattle, WA
Paper ID #12654

Comparison of On-line versus Paper Spatial Testing Methods

Ms. Norma L Veurink, Michigan Technological University

Norma Veurink is a Senior Lecturer in the Engineering Fundamentals Department at Michigan Technological University where she teaches introductory engineering courses and a spatial visualization course designed for engineering students with poorly developed spatial visualization skills. Ms. Veurink manages several summer programs that introduce middle and high school students to engineering. She is active in the Engineering Design Graphics Division of ASEE.

Dr. A. J. Hamlin, Michigan Technological University

AJ Hamlin is a Senior Lecturer in the Department of Engineering Fundamentals at Michigan Technological University, where she teaches first-year engineering courses. Her research interests include engineering ethics, spatial visualization, and educational methods. She is an active member in the Engineering Design Graphics Division of ASEE and is currently serving as the Associate Editor of the Engineering Design Graphics Journal.

Comparison of On-line versus Paper Spatial Testing Methods

Abstract

Spatial visualization skills have been shown to be critical to the success of engineers and engineering students, and the assessment and remediation of those skills is growing in engineering curricula across the country. At Michigan Technological University, spatial skills of freshmen engineering students have been tested with the use of the Purdue Spatial Visualization Test: Rotations (PSVT:R) since 1993. This test has traditionally been administered with paper tests where students record their answers on a scantron form. Due to changes in university computing resources in the 2013-2014 academic year, the opportunity arose to test the freshmen through a Learning Management System (LMS). In the fall of 2014, over 450 of the engineering freshmen were administered the PSVT:R with the traditional paper test method, while approximately the same number took the test through the LMS. Testing half of the students with each method allowed for a comparison to be made between the two testing methods. This paper compares scores from the paper version of the test with scores on the on-line version of the test to determine if students perform equally on the on-line and paper forms of the test.

Background

Online testing has become commonplace in online and blended or hybrid courses. Advances in Learning Management Systems (LMS) and test banks make it easy to create online exams. Even some high-stakes tests, such as the Fundamentals of Engineering Exam, have moved to an online format. Gehringer[1] provides a comparison of some of the benefits and challenges of online versus paper exams. He concludes that online exams, especially open book/open web (OBOW) exams, can provide a more authentic assessment environment. York and Knott[2] compared results from an exam that was given using an LMS with results obtained on a paper exam given the next day. They found that while there were a few individuals that performed much better or much worse on the paper test, the average did not differ dramatically (67% online, 69% paper). Their students preferred the online format and when given the option to take the second exam on paper or online, all students choose to take the online exam. Bayazit and Askar[3] also compared differences between online and paper test performance and duration using an online testing tool they developed. They found there was no significant differences in performance, however students spent more time on the online test.

Other studies have found some differences in on-line versus paper exams. Deutsch, Herrmann, Frese, and Sandholzer[4] found gender differences in students taking online exams. These gender differences were attributed to differences in computer-self efficacy, but they found the differences were reduced considerably after students had a single experience taking an online exam. McDonald[5] considered score equivalence between paper and computer-based assessments and concluded that individual differences in computer experience, computer anxiety, and computer attitudes could impact the potential of some students to perform similarly on paper and computer-based assessments. Shermis and Lombard[6] found that students felt a loss of control and thus increased anxiety when computer-based tests were administered in such a way that each question had to be answered before the subsequent question and previously answered questions could not be reviewed. Mead and Drasgow[7] found differences between computerized and paper-

pencil tests when the tests were based on speed of response. Wilson, Genko, and Yager[8] warned that if scores from a computerized test were meant to be interchangeable with a paper-pencil test, then the mean scores and variances of tests taken with the two different test methods should be the same. However, as early as 1986, the American Psychological Association[9] recognized the differences between paper and computerized tests and advised "When interpreting scores from the computerized versions of conventional tests, the equivalence of scores from computerized versions should be established and documented before using norms or cutting scores obtained from conventional tests."

Description of the study

At Michigan Tech, the Purdue Spatial Visualization Test: Rotations (PSVT:R) has been given as a paper-pencil test to entering engineering students during freshmen orientation since 1993 to identify students that could benefit from a one-credit course designed to improve spatial visualization skills. In the fall of 2013, the university consolidated student computing resources by closing departmental computer labs spread across campus and opening computing labs in a centralized location, the university library. This consolidation allowed for approximately half the entering engineering students in the fall of 2014 (~450) to take the PSVT:R online through an LMS. The other half of the students were administered the test through the paper-pencil method. Both the paper and LMS tests were given at the same time and proctored by engineering faculty and undergraduate assistants. In this study we compare test scores of students using the online versus paper formats.

Most of the first-time freshmen students were assigned to take the LMS version or the paper version of the test according to their orientation team (groups of about 20 students with the same major). The assignment was made by the orientation programs staff based on the size of the teams in an effort to maximize utilization of computing resources. Athletes and students who missed their scheduled test time were given other times and places to take the test. The faculty member administering the athlete and makeup tests chose how to administer the test, with one faculty member choosing to use the LMS method and the other choosing the paper method. Thus placement of students into the different testing methods was not entirely random, but students could not self-select which version of the test they would be given.

Students required to take the PSVT:R at Michigan Tech include engineering students transferring from other colleges and universities, current students who are changing their major to an engineering major, and first-time freshmen. Because transfer students and already enrolled students who are changing their major may have more experience using computers, utilizing an LMS, and taking tests on computers, their scores were analyzed separately from first time freshmen students.

Results

Average scores of first time freshmen taking the PSVT:R through the two different test methods are compared in Table 1 below. In the fall of 2013, all freshmen students were administered the test with the paper-pencil method, so scores from that group are compared as well. As shown in Table 1, the scores of students who took the test through the LMS were lower on average than

students taking the paper version of the test. The scores on the LMS test were highly statistically different than those on both the 2014 and 2013 paper versions of the test. This difference was found for the group as a whole and for the males taking the tests. Although the females taking the LMS version also scored lower than those taking the paper versions, no significant statistical differences were found between the three female groups. This lack of significance may be due to the lower number of females taking the test and the higher standard deviations for the female scores. No statistical differences between scores of students taking the paper test in 2014 with those taking the paper test in 2013 were found.

Table 1: Comparison of average PSVT:R scores for first-time students (maximum score possible = 30)

Type of test and year taken	Average PSVT:R score	Average PSVT:R score of females	Average PSVT:R score of males
LMS in 2014	22.5* (s=4.88, n=430)	20.3 (s=4.74, n=116)	23.4** (s=4.66, n=314)
Paper in 2014	23.8 (s=4.32, n=454)	20.8 (s=4.39, n=90)	24.5 (s=3.96, n=364)
Paper in 2013	23.7 (s=4.48, n=908)	21.2 (s=4.68, n=188)	24.3 (s=4.19, n=720)

* Statistically different than paper test scores in 2014 and 2013 ($p < 0.0001$)
** Statistically different than paper test scores in 2014 and 2013 ($p < 0.0005$)

A prior study of freshmen engineering PSVT:R scores at Michigan Tech (Veurink and Hamlin[10]), revealed that approximately 25% of all students typically score from 28 to 30 on the test, approximately 25% score from 25 to 27, another 25% score from 22 to 24, and the lowest 25% of scores are 21 and below. Other studies related to spatial remediation by the authors (Veurink & Sorby[11] and Hamlin, Veurink, & Sorby[12]) have compared the success of a marginally passing group (PSVT:R score of 19 -21) and a failure group (scores below 19). Because the LMS group performed significantly lower than the paper-pencil groups, the question arose if the difference in test methods impacted students with weaker spatial skills more than those with stronger skills. Tables 2 and 3 show the percentages of scores in the Failing, Marginally Passing, Middle 50%, and High Passing groups by gender for the different testing methods.

Table 2: Comparison of distribution of PSVT:R scores for female first-time students

	LMS in 2014	Paper in 2014	Paper in 2013
Percent failing (PSVT:R < 19)	37.1% (n=43)	28.9% (n=26)	26.1% (n=49)
Percent marginally passing (PSVT:R 19 – 21)	20.7% (n=24)	23.3% (n=21)	20.7% (n=39)
Percent scoring in the historical middle 50% (PSVT:R 22 – 27)	38.8% (n=45)	43.3% (n=39)	46.3% (n=87)
Percent highly passing (PSVT:R 28 – 30)	3.4% (n=4)	4.4% (n=4)	6.9% (n=13)

Table 3: Comparison of distribution of PSVT:R scores for male first-time students

	LMS in 2014	Paper in 2014	Paper in 2013
Percent failing (PSVT:R < 19)	15.0% (n=47)	7.1% (n=26)	10.3% (n=74)
Percent marginally passing (PSVT:R 19 – 21)	17.5% (n=55)	12.6% (n=46)	13.9% (n=100)
Percent scoring in the historical middle 50% (PSVT:R 22 – 27)	47.5% (n=149)	53.0% (n=193)	49.2% (n=354)
Percent highly passing (PSVT:R 28 – 30)	20.1% (n=63)	27.2% (n=99)	26.7% (n=192)

From the distribution of scores shown in Tables 2 and 3 above, it appears taking the LMS version of the test negatively impacted all students taking the test, not just those with weaker spatial skills. The percentages of high pass and middle scores were lower for both males and females in the LMS version of the test compared to the paper versions of the test.

Similar comparisons were made for transfer students and those already at the university but changing majors to engineering. Because some of these students were administered the test through the LMS and some through the paper-pencil method in both 2013 and 2014, scores were grouped by test method and not test date. Table 4 shows the scores of students who took the test through the LMS were statistically different ($p < 0.005$) than those taking the paper version, however, the scores of students taking the LMS version were significantly *higher* than those taking the paper-pencil version. The same was true when comparing the scores of the male students ($p < 0.001$). Females taking the LMS version also scored higher on average than those taking the paper version, however, no significant statistical differences were found between the female groups. The lack of statistical significance could be because of the small difference in scores between the two female groups and the lower number of females taking the test.

Table 4: Comparison of average PSVT:R scores of students with prior college experience

Type of test taken	Average PSVT:R score	Female average PSVT:R score	Male average PSVT:R score
LMS	24.0* (s=4.43, n=89)	21.9 (s=5.78, n=15)	24.4** (s=3.98, n=74)
Paper	22.4 (s=4.86, n=162)	21.5 (s=4.34, n=39)	22.6 (s=4.98, n=123)

* Statistically different than paper test scores of males and females combined ($p < 0.005$)
** Statistically different than paper test scores of males ($p < 0.001$)

It should be noted here that all but nine of the students taking the LMS version of the test did not take it in a proctored setting. In other words, they were allowed to take the test at their convenience and at any computer that had internet access. This computer could be at their home, in a computer lab, at a coffee shop, etc. The test was timed and they were given only one attempt through the LMS, but they did not take the test with a group of students in a formal, proctored setting as the first time freshmen students did. The average score of the nine transfer/changing majors students who took the LMS version with the orientation students in a proctored group setting was 22.7 (most of the nine were males) which is similar to the average

paper score. The paper tests were all given in proctored settings, but many of these students took the test at individual rather than group times. The fact that the proctored LMS scores were similar to the paper scores, but the other LMS scores were not, suggests that the setting of the test plays a role in performance on the test. Students taking the unproctored LMS test likely chose a setting they were comfortable in, and because they could choose the time and place, it is likely they felt less pressure and anxiety when taking the test. Hollister and Berenson[13] found no statistical difference between students taking online exams in a proctored environment or unproctored environment but found greater variability in the scores of the unproctored students. In their study, both the proctored and unproctored groups took the exams at the same time. Hollister and Berenson concluded a structured exam setting which students are familiar with and where questions can be asked of the proctor may provide more "buy in" and a more supportive environment for testing. However, the PSVT:R differs from a typical course exam in that it is shorter in duration (20 minutes) and is a screening tool that requires no preparation nor assistance while taking the test.

Tables 5 and 6 show the percentages of scores in the Failing, Marginally Passing, Middle 50%, and High Passing groups by gender for the different testing methods for the students with prior college experience. From the distribution of scores, it appears taking the LMS version of the test had an opposite effect on those with prior college experience than it did on first time freshmen. For students with prior college experience, the percentage of high pass and middle scores were *higher* for both males and females in the LMS version of the test compared to the paper versions of the test.

Table 5: Comparison of distribution of PSVT:R scores of female students with prior college experience

	LMS	Paper
Percent failing (PSVT:R < 19)	13.3% (n=2)	23.1% (n=9)
Percent marginally passing (PSVT:R 19 – 21)	13.3% (n=2)	20.5% (n=8)
Percent scoring in the historical middle 50% (PSVT:R 22 – 27)	60.0% (n=9)	53.8% (n=21)
Percent highly passing (PSVT:R 28 – 30)	13.3% (n=2)	2.6% (n=1)

Table 6: Comparison of distribution of PSVT:R scores of male students with prior college experience

	LMS	Paper
Percent failing (PSVT:R < 19)	8.1% (n=6)	17.1% (n=21)
Percent marginally passing (PSVT:R 19 – 21)	13.5% (n=10)	18.7% (n=23)
Percent scoring in the historical middle 50% (PSVT:R 22 – 27)	52.7% (n=39)	48.8% (n=60)
Percent highly passing (PSVT:R 28 – 30)	25.7% (n=19)	15.4% (n=19)

Because of the difference found in PSVT:R scores for first-time students taking the LMS versus paper versions of the test, Math ACT scores and performance on a graphics exam taken by many of the students in the study were also compared. Koenig, Frey, and Detterman[14] found the ACT is a measure of cognitive ability and intelligence. Other studies found links between spatial skills and mathematical ability. Burnett, Lane and Dratt[15] and Casey, Nuttal, Pezaris and Benbow[16] found a significant relationship between mental rotation ability and SAT Math scores. Pauley, Kulikowski, Sedransk, and Engel[17] found there is a moderate correlation between the ability to compare and contrast mathematical operations with spatial skills for translating between two- and three-dimensional images. Yue[18] found that for community college students in an engineering graphics course there was a possible relationship between basic mathematical skills (measured using COMPANION placement test) and spatial visualization skills (measured using PSVT:R).

To determine if the students taking the LMS version had lower cognitive ability than students taking the paper versions and because of correlations found between math and spatial ability in other studies, the Math ACT scores of the groups were compared. The graphics exam grades were compared to provide an additional measure of any differences in spatial ability between the two groups.

Average Math ACT scores for first-time freshmen are shown in Table 7 below. Although the LMS group had the lowest average PSVT:R scores they had significantly higher math ACT scores than the 2014 paper group and marginally higher math ACT scores than the 2013 paper group. Statistical differences were more significant for the females than the males. There were no significant differences in math ACTs between the 2013 and 2014 paper groups with the exception of the female 2014 paper group which had marginally lower scores than the 2013 paper group. Since math ability has been positively correlated with spatial ability, the LMS group should have outperformed the paper groups on the PSVT:R.

Table 7: Comparison of average Math ACT scores for first-time students

Type of PSVT:R test and year taken	Average math ACT score	Average math ACT of females	Average math ACT of males
LMS in 2014	28.2* (s=3.48, n=415)	28.1** (s=3.44, n=111)	28.3*** (s=3.49, n=304)
Paper in 2014	27.6 (s=3.27, n=442)	26.8 (s=3.19, n=88)	27.8 (s=3.25, n=354)
Paper in 2013	27.8 (s=3.50, n=890)	27.8** (s=3.42, n=187)	27.7 (s=3.52, n=703)

* Statistically different than 2014 paper test group ($p < 0.005$) and marginally different than 2013 paper test group ($p = 0.012$)

** Statistically different than 2014 female paper test group ($p = 0.002$ for the female LMs and $p = 0.008$ for the female 2013 paper group)

*** Marginally different than the 2014 male paper test group ($p = 0.056$) and 2013 male paper group ($p = 0.016$)

Math ACT scores were also compared for the students with prior college experience. No statistically significant differences in Math ACT were found between groups taking the LMS versus paper version of the PSVT:R.

Student performance on a graphics exam in a first-year engineering course required by all engineering majors at Michigan Tech was compared for students taking the PSVT:R on paper and through the LMS to determine if there was a difference in spatial ability between these groups. In the first-year engineering course, ENG1101, approximately five 1.5 hour class sessions are spent on sketching topics, followed by an exam. The graphics topics covered in this course include isometric and oblique sketching, orthographic projections of normal, inclined, and curved surfaces, rotations, reflections, and planes of symmetry. Students scoring 60% or below on the PSVT paper and LMS versions were excluded from this analysis because they were required to take a supplemental one-credit course to develop their spatial skills and thus had additional training beyond the majority of the ENG1101 students.

The distribution of student scores on the graphics exam as a function of PSVT:R score is shown in Figure 1 for both the LMS and paper versions. For each PSVT:R score, the box contains the central 50% of scores (25^{th} percentile to 75^{th} percentile), with the median identified as the line inside the box. The whiskers extend to the most extreme scores with the exception of the outliers which are plotted individually. As shown in Figure 1, the top scores earned on the graphics exam for all PSVT:R scores and LMS/paper versions are 98% or better, while the minimum scores and the median scores generally increase with increasing PSVT:R score. This trend is also apparent in Table 8, which compares the average score on the graphics exam for students in the marginally passing, middle 50%, and high passing groups.

The Pearson product correlation coefficient between the PSVT:R scores and the graphics exam scores are highly significant for both versions of the PSVT:R (LMS: $r=0.267$, $n=291$, $p=0.000002$; paper: $r = 0.304$, $n=336$, $p<0.000001$). A one-tailed t-test for the graphics exam scores indicated that there was no statistically significant difference on graphics exam scores for those students that took the PSVT:R on paper compared to those that took it through the LMS ($t=0.234$, $n=627$, $p=0.41$). However, for this group of students, there also was no statistically significant difference in PSVT:R scores from those taking the LMS versus paper versions ($t=1.14$, $n= 627$, $p = 0.13$), but there was a significant difference in the Math ACT scores of the two groups, with the LMS group having higher Math ACT scores than the paper group ($t = 2.73$, $n = 592$, $p = 0.003$). Thus, it appears there is no difference in spatial ability as measured by PSVT:R and graphics exam grades for this subset of students in the LMS and paper groups.

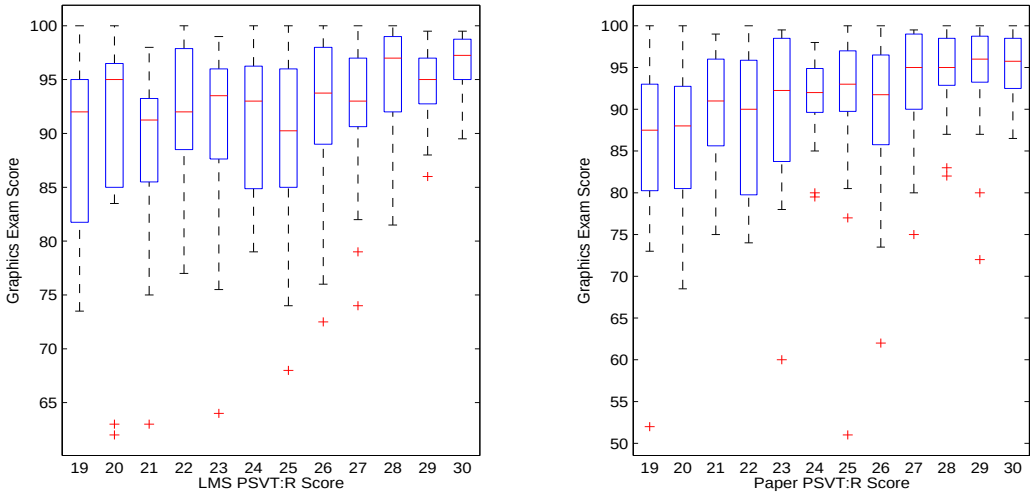

Figure 1: Distribution of graphics exam scores for students scoring 19 or higher on the LMS and paper versions of the PSVT:R

Table 8: Comparison of average graphics exam scores by PSVT:R score group

	LMS	Paper
Graphics Exam Average Marginally Passing (PSVT:R 19-21)	89.1 (n=58, s=8.9)	90.5 (n=40, s=7.1)
Graphics Exam Average Historical Middle 50% (PSVT:R 22-27)	91.0 (n=166, s=7.1)	91.0 (n=195, s=7.9)
Graphics Exam Average Highly Passing (PSVT:R 28-30)	95.1 (n=67, s=4.1)	94.7 (n=86, s=5.1)

Conclusion

In conclusion, it appears that for first-time freshmen, taking the PSVT:R through an LMS is not the same as taking a paper version of the test. In this study, first time freshmen taking the test through an LMS scored on average a full point lower than those taking a paper version of the test, even though the LMS group had higher Math ACT averages than the paper groups. Lower percentages of students using the LMS test method ended up in the high pass and middle groups compared to students taking paper versions of the test. Although students are told not to make marks on the paper version of the test, several students do so. Testing students on-line removes this opportunity to make marks and may also reduce students' use of other tactile motions in solving the rotational problems. These differences may be strong contributors to the differences in test scores. Another reason for the differences in scores may be that first-time freshmen are less accustomed to taking computerized tests than paper tests and the lack of experience with computerized testing and LMS test environments negatively impacts performance.

For students with prior college experience, taking the PSVT:R through an LMS appeared to benefit PSVT:R scores, however, this difference may be due more to students being able to

choose the setting of the test than the difference in test methods. It does appear that at a minimum, administering the PSVT:R through an LMS for this group of students does not negatively impact their performance on the test.

At Michigan Tech, the PSVT:R scores are used to determine which students would benefit from a remedial spatial visualization course. Historically, students scoring 18 (60%) or below on the test were encouraged or required to take the remedial course. If future PSVT:R tests are given using an LMS, consideration should be given of whether the 60% level or another, lower level, should be used to determine which students should take the remedial course.

References

[1]Gehringer, E. (2010, June 20-23). *Online vs. on-paper exams.* Paper presented at the 2010 ASEE Annual Conference and Exposition, Louisville, KY.

[2]York, S., & Knott, T. (2006, June 18-21). *A comparison of on-line and traditional testing methods*. Paper presented at the 2006 ASEE Annual Conference and Exposition, Chicago, IL.

[3]Bayazit, A., & Askar, P. (2012). Performance and duration differences between online and paper-pencil tests. *Asia Pacific Education Review, 12*(2), 219-226.

[4]Deutsch, T., Herrmann, K., Frese, T., & Sandholzer, H. (2012). Implementing computer-based assessment – A web-based mock examination changes attitudes. *Computers & Education, 58*(4), 1068-1075.

[5] McDonald, A.S. (2002). The impact of individual differences on the equivalence of computer-based and paper-and-pencil educational assessments. *Computers & Education, 39*(3), 299-312.

[6]Shermis, M.D., & Lombard, D. (1998). Effects of Computer-Based Test Administrations on Test Anxiety and Performance. *Computers in Human Behavior, 14*(1), 111-123.

[7]Mead, A.D., & Drasgow, F. (1993). Equivalence of Computerized and Paper-and-Pencil Cognitive Ability Tests: A Meta-Analysis. *Psychological Bulletin*, 114(3), 449-458.

[8]Wilson, F.R., Genco, K.T., & Yager, G.G. (1985). Assessing the equivalence of paper-and-pencil vs. computerized tests: demonstration of a promising Methodology. *Computers in Human Behavior, 1*, 265-275.

[9]American Psychological Association Committee on Professional Standards and Committee on Psychological Tests and Assessment, (1986). *Guidelines for computer-based tests and interpretations.* Washington, DC.

[10]Veurink, N., & Hamlin, A.J. (2011). *Spatial Visualization Skills: Impact on Confidence in an Engineering Curriculum.* Paper presented at the 2011 ASEE Annual Conference and Exposition, Vancouver, B.C., Canada.

[11]Veurink, N., and Sorby, S.A. (2011). *Raising the Bar? Longitudinal Study to Determine which Students Would Most Benefit from Spatial Training.* Paper presented at the 2011 ASEE Annual Conference and Exposition, Vancouver, B.C., Canada.

[12]Hamlin, A.J., Veurink, N.L., and Sorby, S.A. (2012). *Importance of Spatial Training for Pre-Calculus Students.* Paper presented at the 67th Midyear Conference of the Engineering Design Graphics Division of ASEE, Limerick, Ireland.

[13] Hollister, K. K., & Berenson, M. L. (2009), Proctored Versus Unproctored Online Exams: Studying the Impact of Exam Environment on Student Performance. *Decision Sciences Journal of Innovative Education, 7*, 271–294.

[14]Koenig, K.A., Frey, M.C., & Detterman, D.K. (2008). ACT and general cognitive ability. *Intelligence, 36*(2), 153-160.

[15]Burnett, S.A., Lane, D.L., & Dratt, L.M. (1979). Spatial visualization and sex differences in quantitative ability. *Intelligence, 3*(4), 345-354.

[16]Casey, M.B., Nuttal, R., Pezaris, E., & Benbow, C.P. (1995). The influence of spatial ability on gender differences in mathematics college entrance test scores across diverse samples. *Developmental Psychology, 31*( ), 697-705.

[17]Pauley, L.L., Kulikowich, J.M., Sedransk, N., & Engel, R.S. (2011). *Studying the reliability and validity of test scores for mathematical and spatial reasoning tasks for engineering students.* Paper presented at the 2011 ASEE Annual Conference and Exposition, Vancouver, B.C., Canada.

[18]Yue, J. (2002). *Do mathematical skills improve spatial visualization abilities?* Paper presented at the 2002 ASEE Annual Conference and Exposition, Montreal, Canada.