INCREASING STUDENT PERSISTENCE IN A SKETCHING APP FOR SPATIAL VISUALIZATION TRAINING

Lelli Van Den Einde[1], Nathan Delson[1], Elizabeth Cowan[2], Daniel Yang[3]

[1]*Assoc. Teaching Professor, Univ. of California, San Diego & eGrove Education (USA)*
[2]*User Experience, eGrove Education (USA),*
[3]*Graduate Student, Mech. and Aerospace Engineering, Univ. of California, San Diego (USA)*

Abstract

Spatial visualization is the mental representation and manipulation of 2D and 3D shapes. Spatial reasoning is a learnable skill that has been correlated with increased GPA's and retention in science, technology, engineering and math (STEM) disciplines. Traditional spatial visualization training includes both multiple choice questions and freehand sketching assignments. To make spatial visualization training easier to teach and more engaging to learn, a touchscreen App was developed where students sketch assignments on the screen. An automatic grading algorithm provides immediate feedback to the user and hints when needed. In spring 2014, the App was used with iPads in a one-unit spatial visualization course, where students took pre- and post-spatial visualization assessment tests. In 2014, 46% of the students who scored low on the pre-test had a significant improvement on the post-test. However, the other 54% of the low pre-test students had virtually no increase in their post-test score. The largest measured difference between these two groups was related to persistence. When students were confronted with a mistake in their initial sketch attempt, they were provided the option to try again on their own, or peek at the solution. Those students who did not improve their post-test scores chose to peak on all but 7% of the assignments and their sketching effort was minimal with the peek being used as a crutch to quickly gain access to the solution. In contrast, students who increased their scores on the post-test resisted peeking and tried to solve the problems more on their own. This study identified a method to measure persistence that could be used as an early indicator of students who are at risk of low performance in the course, allowing for targeted help. But the question remained: could persistence be increased? To increase persistence the App was modified to motivate students to try more on their own before accessing help. Students were provided with a hint as an intermediate step before a peek at the solution. In addition, a star reward system was included to encourage students not to use a hint or peek unless they were stuck. In the last two course offerings (winter 2017 and spring 2017), 82% and 75%, respectively, of the incoming students with low spatial skills showed significant improvement (greater than or equal to 10% on their post-test score). This is an average of 70% increase when compared to the 2014 study results. Additionally, there was a 92% increase in the persistence metric of the overall class, illustrating that User Experience Design can increase student persistence.

Keywords: Sketching, spatial visualization, technology, touchscreens, user experience design.

1 INTRODUCTION

Spatial visualization is the mental representation and manipulation of 2D and 3D shapes. A common way of measuring spatial visualization ability is with the Purdue Spatial Visualization Test: Rotations (PSVT:R) [1]. This is a multiple-choice test with an example shown in Fig. 1. Spatial reasoning is a learnable skill that has been correlated with increased GPA's and retention in science, technology, engineering and math (STEM) disciplines [2]. However, most students do not receive formal training in these skills because the concepts are not a priority of current K-12 curriculum. Research has shown that a single unit spatial visualization course has increased retention in STEM majors, especially for women and other underrepresented minorities [3].

Traditional spatial visualization training includes multiple choice questions and freehand sketching assignments. A prior study showed that sketching assignments were critical to increasing spatial visualization ability [2]. To make spatial visualization training easier to teach and more engaging to learn, the authors developed a touchscreen App where students sketch assignments on the screen [4]. An automatic grading algorithm provides immediate feedback to the user and hints when needed. Starting in spring 2014, the App has been used with iPads in a one-unit spatial visualization course, and more recently a required structural engineering course, where students take pre- and post-spatial visualization assessment tests.

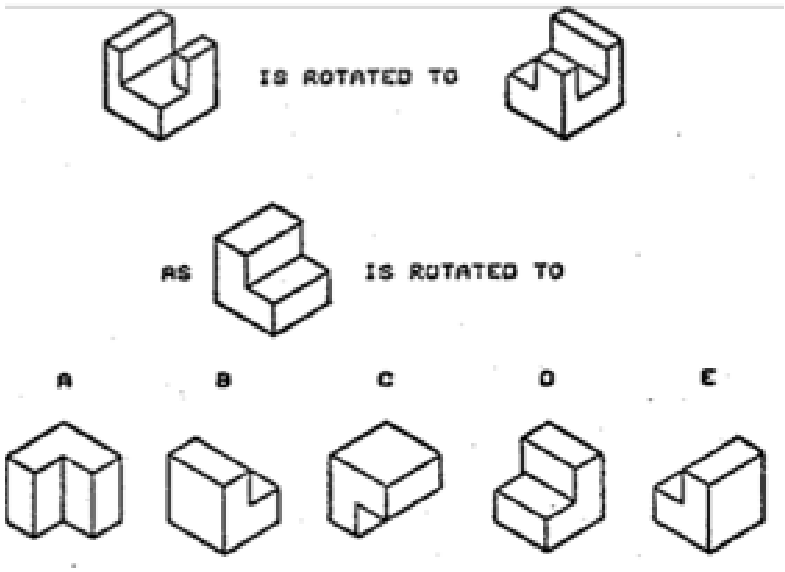


Figure 1. The Purdue Spatial Visualization Test: Rotations (PSVT:R) sample question ***Error! Reference source not found.***

In an early pilot at the University of California, San Diego in 2014 [4], 46% of the students who scored low on the pre-test had a significant improvement on the post-test. However, other students had virtually no increase in their post-test score (specified as an increase greater than or equal to 10%). The largest measured difference between these two groups was related to persistence. When students were confronted with a mistake in their initial sketch attempt, they were provided the option to try again on their own, or peek at the solution. Those students who did not improve their post-test scores chose to peak on all but 7% of the assignments, and often their initial sketching effort was minimal with the peek being used as a crutch to quickly gain access to the solution. In contrast, students who increased their scores on the post-test resisted peeking and tried to solve the problems more on their own. This study identified a method to measure persistence that could be used as an early indicator of students who are at risk of low performance in the course, allowing for targeted help.

But the question remained: could persistence be increased? Since this initial offering of the course, the app was modified to increase student engagement and persistence. Enhancements included user interface modifications, assignment additions and design modifications, and the introduction of gamification **Error! Reference source not found.**. By changing feedback provided to the user, students were now given the option to take a hint as an intermediate step before a peek at the solution. In addition, a star reward system was included to encourage students to not ask for help (hint or peek) unless they were stuck. Two additional pilots were conducted in 2017. In Qinter 2017, the assignment modifications were evaluated in an *elective* 1-credit class dedicated primarily to teaching spatial visualization skills. It was open to mechanical engineering and structural engineering students. Students were given three hours per week in class to do the App with teaching assistant support. In Spring 2017, the flexibility of the App was tested. In this study, students in a *required* engineering graphics course used the App as homework assignments and the teacher had the freedom of choosing which assignments to assign. This paper provides a comparison between results from the 2014 pilot study with the latest two 2017 studies, where improved feedback and the star motivation system were implemented. Preliminary results show that the user experience in the spatial visualization training App can be adapted to increase student persistence.

2 SPATIAL VIS APP OVERVIEW & DEVELOPMENT

The Spatial Vis application was initially developed for use with an iPad, but is currently being ported to iPhone, Android, Chromebook, and Windows platforms. It includes an assignment window that provides an image and instructions for what is to be sketched, the sketching window where the student sketches the solution, the toolbar (that provides some navigation buttons, a pencil, an eraser, and a help button), and finally the Submit button where students submit their sketch for grading (see Fig. 2 Right). The sketching window and the assignment window each have a reference dot so students know the position in the sketching window they need to draw their solution for the grading algorithm to correctly score their submission. When the user selects the submit button at the top right of the toolbar, the grading algorithm is initiated. The grading algorithm produces a pop-up window that gives the student immediate feedback if their solution is correct or not. If the solution is correct, the student moves on to the next assignment. In the 2014 version of the App, an incorrect sketch would allow the students to either retry or peek at the solution. As explained in more detail below, an intermediate option of "hint" was added in 2016.

2.1 Enhancements to the App

Numerous modifications were made to the App between 2014 and 2017. The 2014 App was developed using assignments from a paper textbook [4]. Since then, all the assignments have been recreated and enhancements were made possible due to the move from paper to a touchscreen. The use of color allowed for multiple cues to be added to the sketching window. Fig. 2 highlights the differences in interfaces between the 2014 App and the 2017 app. As shown in Fig. 2, complete gridlines were added to the sketching window, since the color and thickness of the sketched lines were easily differentiated from the grid lines. In other assignments, multiple reference dots of different colors were employed to guide the users in early assignments.

Data queries were done of the student attempts in 2014. Assignments that took many attempts for the class to complete were designated as having a large increase in difficulty, or confusing instructions. The instructions were clarified and animations were added to some of the tutorials. In addition, large increases in difficulty were scaffolded into multiple assignments with smaller steps in difficulty. Overall, care was taken for a gradual increase in difficulty, and for use of cues in early assignments to aid students at the beginning of a lesson [5].

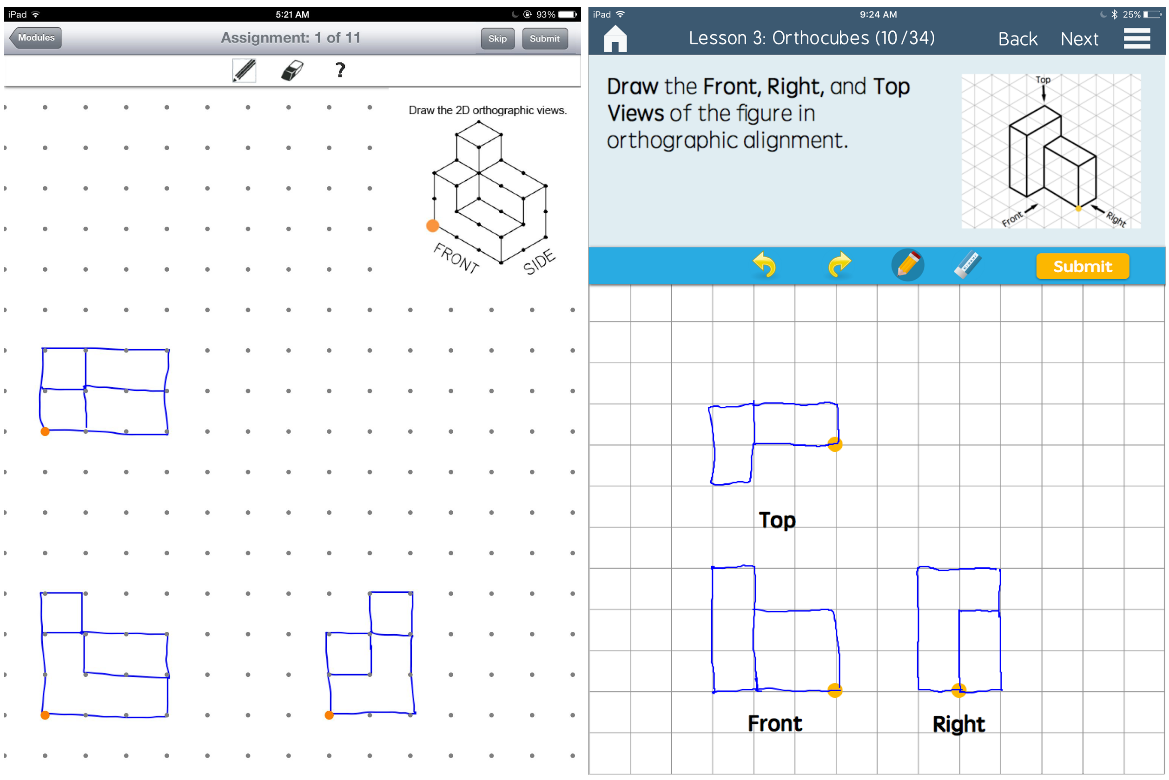


Figure 2. Comparison of Interfaces for the 2014 App version (left) and 2017 App version (right).

Most importantly, an intermediate level of help was added (see Fig. 3). In the 2014 App, when students got an answer incorrect they were given the option to retry the assignment or peek at the solution. In later versions, a hint feature was added to tell the user which parts of their drawing were correct by highlighting them in green and removing the remaining incorrect lines. If a student took a hint and most of their submission remained in the workspace then the student was close to the solution. Alternatively, if the student took a hint and most of their submission disappeared then they must rethink the problem, ask for help, or possibly use a peek.

Rather than relying on the hint and peek features, to provide an additional incentive for students to learn the content/objectives of the modules, the final three assignments in each module were designed as ***In-***

App Test Questions. These assignments were of moderate difficulty and the hint and peek features were disabled.

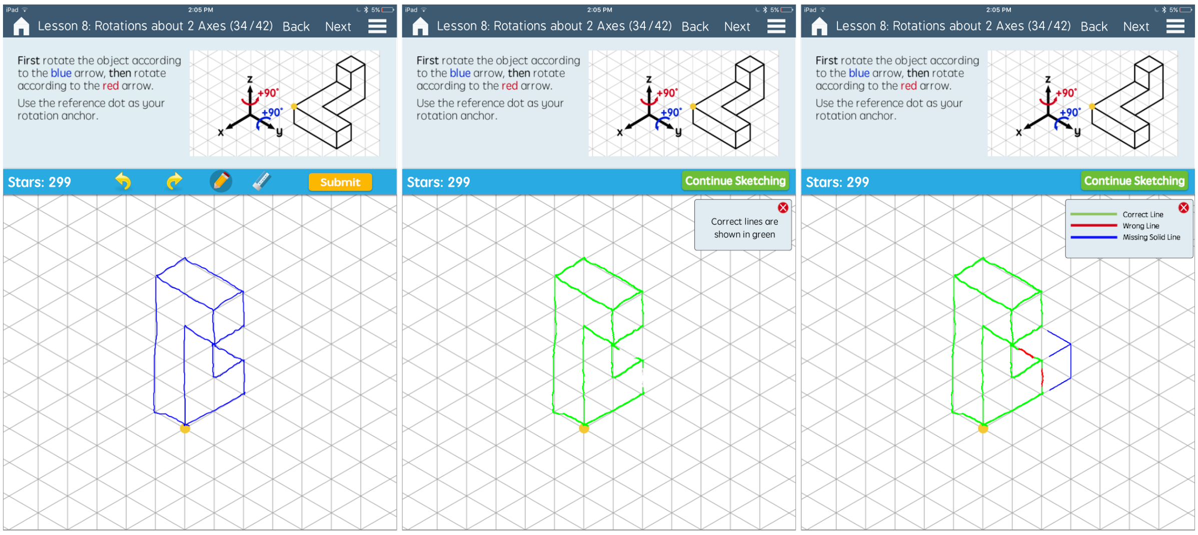
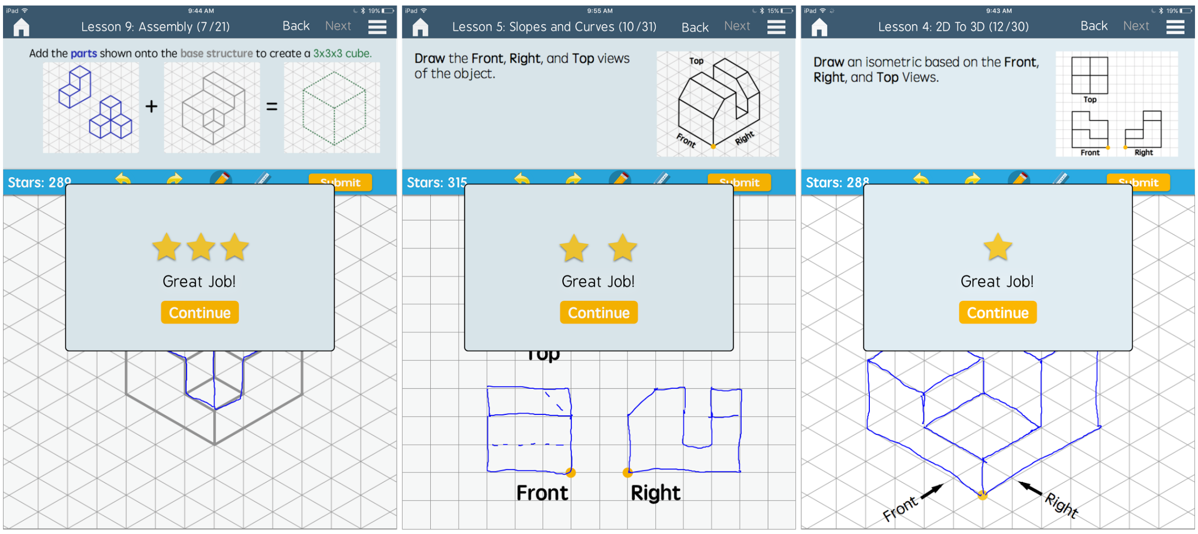


Figure 3. Original Submission (left), hint feedback (middle), and peek feedback (right).

Lastly, to encourage persistence, a gamification feature was added to the application. A goal was to have students retry often and limit their use of the hint and peek, so stars were added as an incentive. The maximum stars a student could receive on an assignment was three. Students could retry as many times as they would like without a penalty (see Fig. 4 left). If a student used the hint feature on an assignment, then the maximum stars they could receive was two (see Fig. 4 middle). If a student used a peek, then the maximum stars they could receive was one (see Fig. 4 right).



No Hint or Peek Used *Hint Used* *Peek Used*

Figure 4. Stars and Help Features

3 PILOT STUDIES IN COLLEGE CLASSROOMS

The Winter 2014 pilot course was an ***elective*** one-unit credit, pass/no pass course at the University of California, San Diego consisting of 52 students who met once a week for ten weeks. Students in the class were given 3 hours per week in class to do the app with teaching assistant support. At the beginning and end of the course the PSVT:R was administered for a timed 20 minute period. Prior analysis focused on students who scored 70% or below on the pre-test because they were considered most at risk for dropping out of a STEM major. The Winter 2017 pilot course was implemented in the same course as the 2014 pilot. 32 students utilized the app with the new modifications to investigate the impact of the user interface changes.

The Spring 2017 pilot course was a ***required*** structural engineering graphics course that assigned the App as homework assignments rather than in-class work. Of the 137 students, only 91 took both the pre- and post-test (some of these students dropped the course). Of these students, only 79 also completed all *the In-App Test Questions*. The students who did not complete all the *In-App Test Questions* were approximately evenly distributed in terms pre-test scores.

3.1 Results

Results from the different pilots were evaluated from pre- and post-test data as well as metrics of student activity within the app. Only students who completed the course and took both the pre- and post-test were included. Students were then grouped by their pre-test score into three groups: Low Pre-test (≤70%), Mid Pre-test (>70% - <90%) and High Pre-test (≥90%). Using a Matlab script, various performance and persistence metrics were pulled out of the database and assessed for the different groups.

3.1.1 Percent of Students with Significant Improvement in their PSVT:R Scores

In the 2014 pilot, out of the 52 students that took the one-unit course, 13 of them scored 70% or below on the pre-test. Those 13 students were the focus of a preliminary analysis [4]. The results showed that the Low Pre-Test Group of students had an average test improvement of 18%. However, there was a noticeable split between the individual students in the Low group. 54% of the students had no or low improvement, while the other 46% of students had significant improvement (taken as 10% test improvement or more). Reasons for why some students improved while others did not were further investigated. The largest difference between the two groups was the number of times they tried again without using the peek feature. The students with significant improvement would retry assignments without peeking 74% more often than those who had low or no improvement [4]. Accordingly, the Percent Tried Again Without Peeking normalized by the Number Wrong on the First Try was identified as the persistence metric.

A similar analysis was done for the 2017 pilot courses to show differences between using the app in-class in an elective class (Winter 2017) and assigning the app for homework in a required class (Spring 2017). Detailed results for the required Spring 2017 class are shown in Table 1. The results showed that the Low Pre-Test Group of students had an average test improvement of 19%. There were only five Low Pre-Test Group students who did not have significant improvement in post-test results. The performance of these five students was further investigated and is described more in Section 3.1.3. Table 1 also shows that the average number of attempts for each assignment as well as the average number of attempts for the In-App Test Questions (where no hint or peek is allowed) was larger for the Low Pre-Test Group and decreased considerably for the Mid and High groups, as expected.

Table 1. Overall Performance Results for Students of All Test Groups in the Spring 2017 Trial.

	Pre Test Group: Low (n=20)	Pre Test Group: Mid (n=44)	Pre Test Group: High (n=15)	All Groups (n=79)
Avg Pre-Test Score	55.8% σ=13.4%	79.6 % σ=7.2%	94.0% σ=3.8%	76.3% σ=15.8%
Avg Post-Test Score	74.2% σ=15.0%	86.1% σ=12.3%	93.3% σ=8.8%	84.5% σ=14.0%
Avg Pre-Post Test Improvement	23.6% σ=19.2%	6.0% σ=13.3%	-1.5% σ=9.9%	9.0% σ=16.9%
Percent Correct First Try (nCFT)	52.5% σ=10.5%	58.1% σ=12.1%	71.1% σ=10.2%	59.2% σ=12.9%
Percent Tried Again Without Peeking (nARNHC/nWFT)	57.7% σ=27.0%	79.5% σ=19.2%	94.7% σ=8.5%	76.8% σ=23.5%
Avg Num Attempts	2.56 σ=.77	2.33 σ=.92	1.67 σ=.40	2.26 σ=.86
Avg Num App Test Attempts	2.95 σ=1.0	2.75 σ=1.4	2.06 σ=1.1	2.67 σ=1.26

A comparison of the results for the students entering with low pre-test scores for all three pilot courses is shown in Table 2. The group of students who had a significant increase in post-test scores grew from

46% using the old app material in the 2014 pilot, to 82% and 75% using the new app material (Winter 2017 and Spring 2017, respectively). This represents an increase of 78% (Winter 2017) and 63% (Spring 2017) in post-test scores using the new app with the user experience enhancements.

Table 2. Test Performance of the Low Pre-Test Group for 2014, Winter 2017, and Spring 2017 Pilots.

Low Pre-Test Students: 70% and below				
Course	Average Pre-Test Score	Average Post-Test Score	Average Test Improvement	Students with Significant Improvement (10%+)
Elective Winter 2014, with Old App (n=13)	55%	64%	18%	46%
Elective Winter 2017, with New App (n=11)	61%	80%	39%	82%
Required Spring 2017, with New App (n=20)	57%	75%	38%	75%

3.1.2 Persistence Metric vs. In-App Test Questions

While the PSVT:R is widely used, it does not measure ability for students to properly draw sketches. To provide an additional assessment metric, three In-App Test Questions, for which there were no hints or peaks, were inserted at the end of each lesson module. The performance was measured by the number of attempts needed to correctly submit these assignments. The lower the number in this metric represented a higher skill. For the Spring 2017 class, the Average Number of Attempts on the In-App Test Questions was plotted against Post-Test scores for both the Low and Mid Groups (Fig. 5). The group of students coming in with high test scores was excluded since their scores were so close to the max that a saturation effect occurred.

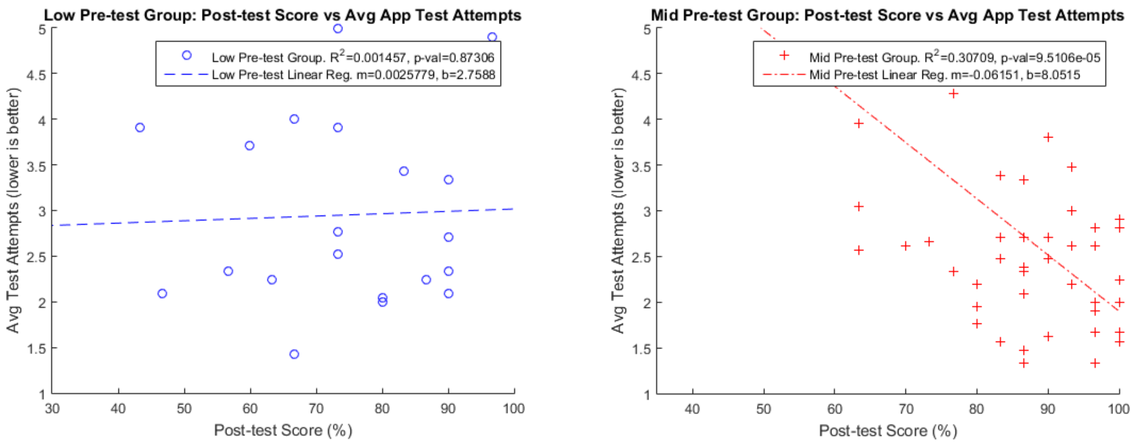


Figure 5. Average Number of in-App Test Question Attempts vs. Post-Test Score for Low Pre-Test Group (left) and Mid Pre-Test Group (right)

The results demonstrate that the Mid Group performance on the In-App Test Questions had a statistically significant correlation with Post-Test score (p=9.5e-5) and a high correlation slope (m=0.06). However, in the Low Pre-Test group, the correlation was not significant (p=0.87) and the correlation slope was low (m=0.003).

The Average Number of In-App Test Question Attempts was also plotted against the Persistence Metric for both Low and Mid Pre-Test Groups (Fig. 6). Here we see a statistically significant correlation for the Mid Group (p=0.004) and a very similar trend (almost statistically significant, p=0.055) for the Low Group. The students who persisted within the App by attempting most of the questions on their own without hinting or peeking required less number of attempts on the In-App Test Questions. This demonstrates that the In-App Test Question is correlated with the PSVT:R Post Score and Persistence metric, at least for the mid group. An advantage of using the In-App Questions is that they provide formative assessment during the course and possibly could identify students that need additional assistance early on.

Overall there was a significant increase in the persistence metric between 2014 and 2017 for all pre-test groups. In 2014, 40% of all assignments were Tried Again Without Peeking. In 2017, this persistence metric increased to 77%; an increase of 93%. We surmise that the implementation of the reward star system had a significant impact on encouraging students not to peek and to continue trying. This illustrates that app design can significantly change student behavior.

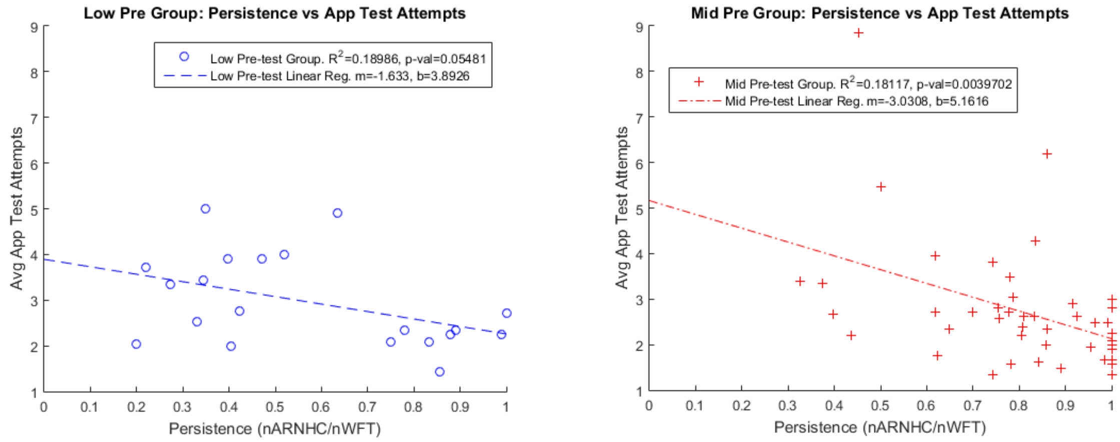


Figure 6. Average Number of in-App Test Question Attempts vs. Persistence Metric for Low Pre-Test Group (left) and Mid Pre-Test Group (right)

To gauge the effects of the Persistence Metric, it was evaluated against the Percent Improvement on the PSVT:R for both the Low and Mid Pre-Test Groups (Fig. 7). While both groups were not statistically significant ($p>0.05$), it is clear for the Mid Pre-Test Group (almost statistically significant) that as the persistence metric increased, the percent improvement on the PSVT:R increased. There was no correlation in the Low group, which is discussed below.

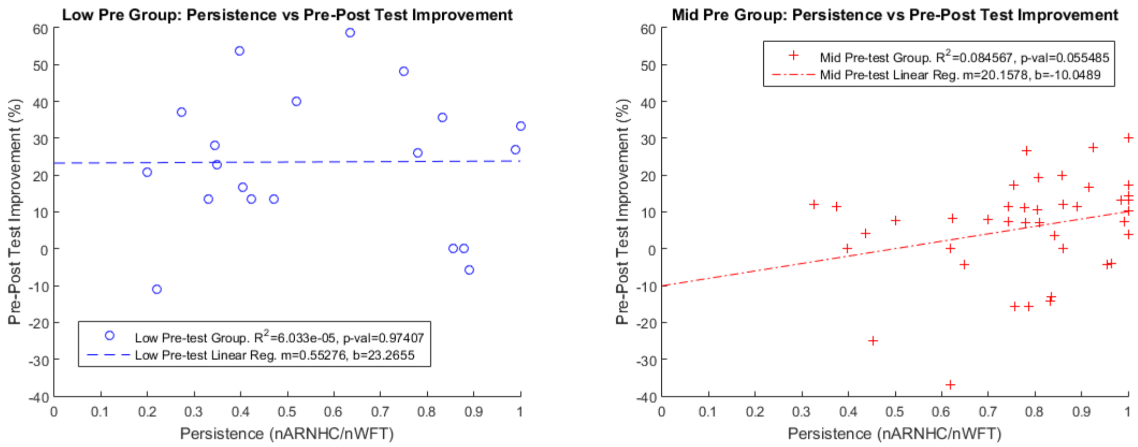


Figure 7. Persistence Metric vs. Percent Improvement on PSVT:R for Spring 2017 for Low Pre-Test Group (left) and Mid Pre-Test Group (right)

3.1.3 Further Evaluation of Low Pre-Test Group

The difference between the Low and Mid Pre-Test groups warranted further investigation. The Low Pre-Test Group was divided into subgroups of those that had a significant increase on their post-test (greater than or equal to 10%), and those that did not have a significant improvement. Plots of these two subgroups are shown in Fig. 8. The Persistence Metric vs. the percentage increase in the post-test score (Pre-Post Test Improvement) is plotted in Fig. 8 (left), while the Persistence Metric vs. the Average Number of In-App Test Attempts is plotted in Fig.8 (right).

There are only 5 students who are in the subgroup that did not have significant improvement in post-test results. We see in the plots that 3 of these students are clustered together (highlighted by a surrounding oval). As shown in Fig. 8 (left), these 3 students had relatively high persistence, but did not have gains on their post-test score. However, as shown in Fig. 8 (right), these 3 students did have strong In-App Test Attempts results (lower number of attempts is better). Accordingly, these 3 students did work hard on the App and did improve in ability to solve App questions. It is hard to hypothesize why these students did not have stronger post-test results, as the Mid group students did. The number of students in this subgroup is too small to draw statistically significant results. Further investigation is warranted for students who did not see increases in post-test score performance.

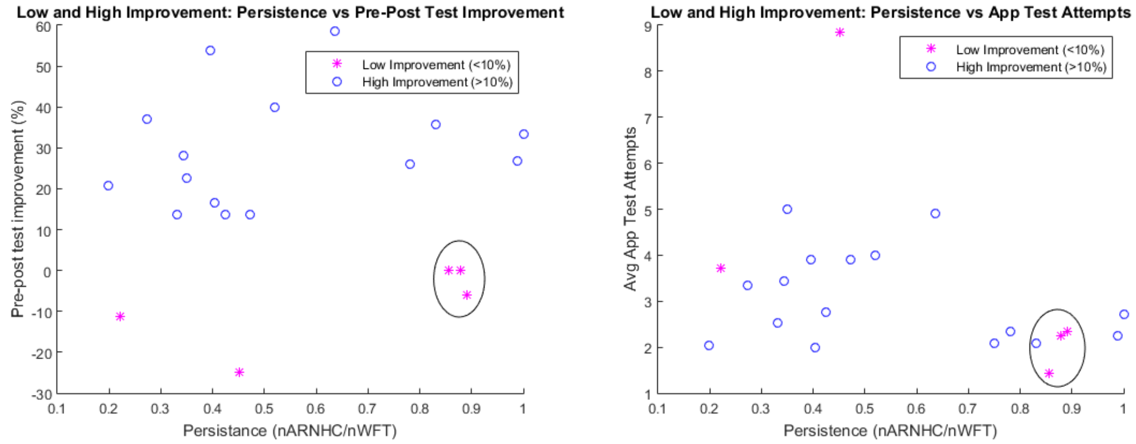


Figure 8. Persistence Metric vs. Percent Improvement on PSVT:R (Left) and Persistence Metric vs. Average Number of In-App Test Attempts (right) for Spring 2017 Individuals in the Low Pre-Test Group.

4 CONCLUSIONS

Changes in the Spatial Visualization Training App between 2014 and 2017 included a rewrite of all the assignments, and taking advantage of the digital interface using color and animations. The difficulty of the original assignments was identified from the student data, and a more gradual increase in assignment difficulty was implemented where warranted. In addition, the user experience was modified to encourage students to be more persistent and to try an assignment again on their own without use of help through use of star accumulation and intermediate levels of help. These enhancements to the app resulted in significant changes in student behavior and achievement. The Persistence Metric increased by 93%. There was also an increase of up to 82% in the number of students who had significant gains in the class, among those that needed it the most (Low Pre-Test students).

A new measure of student performance was evaluated based upon In-App test assignments. This measure was highly correlated to the PSVT:R post-test scores, at least in the Mid group. Variations in the Low Pre-Test group were observed. While many of these students did see significant improvement in their PSVT:R scores, some students did not improve even though they persisted with the app assignments. Further investigation in this area is necessary.

ACKNOWLEDGEMENTS

The authors would like to acknowledge Stephanie Hernandez and Michael Hohl for contributing to the application's graphics, and Justyn Bell and Sean Patno for their programming efforts. Additionally, the authors would also like to thank Jackie First, Stephanie Hernandez, Edward Lan, Samuel Lechtman, and Patricia Tan for administering and coordinating the MAE7 courses.

These studies were supported by the University of California, San Diego, Department of Mechanical and Aerospace Engineering, Department of Structural Engineering, and a grant from the NSF (SBIR Award # 1407123). Nathan Delson has equity interest in eGrove Education, LLC, a company that may potentially benefit from the research results. The terms of this arrangement have been reviewed and approved by the University of California, San Diego in accordance with its conflict of interest policies. In addition, a Small Business Innovation Research (SBIR) grant was awarded to eGrove Education, LLC, by the NSF (Award # 1648534), that also support the research effort of this publication.

REFERENCES

[1] R. Guay, Purdue Spatial Visualization Test-Visualization of Views. Purdue Research Foundation, West Lafayette, IN, 1977.

[2] S.A. Sorby, "Educational research in developing 3-D spatial skills for engineering students", International Journal of Science Education, vol. 31, no. 3, pp. 459-480, 2009.

[3] C. Hill, C. Corbett, A. St Rose, *Why so few? Women in science, technology, engineering, and mathematics*, American Association of University Women, 1111 Sixteenth Street NW, Washington, DC 20036, 2010.

[4] N. Delson, N., & L. Van Den Einde, "Tracking Student Engagement with a Touchscreen App for Spatial Visualization Training and Freehand Sketching", Paper presented at 2015 ASEE Annual Conference & Exposition, Seattle, Washington, 2015. 10.18260/p.24931.

[5] Cowan, N. Delson, B. Mihelich, and L. Van Den Einde, "Improvement in Freehand Sketching Application for Spatial Visualization Training", accepted at the 11th Conference on Pen and Touch Technology in Education, Evanston, IL, 2017.