Allison Liu, Graduate Student, the University of Pittsburgh, asl36@pitt.edu
Chris Schunn, Cognitive Scientist, the University of Pittsburgh, schunn@pitt.edu
Robin Shoop, Director, Carnegie Mellon Robotics Academy, rshoop@cmu.edu

Robot Virtual Worlds: Physical vs. Virtual Programming Fall 2012 study results

Public high-school students from two elective programming classes participated in the study. The same teacher taught both classes. One class completed a ROBOTC programming course using physical VEX robots (the Physical class), while the other class completed a ROBOTC programming course using virtual VEX robots (the Virtual class). Thirteen students were in the Physical class, and 17 students were in the Virtual class. Both classes consisted primarily of freshmen and sophomore students with little to no prior programming experience.

Both the Physical class and the Virtual class completed the same pre-test and post-test online. The pre-test and post-test contained the same 50 items, and both classes completed the pre-test around the same date. Eleven students in the Physical class and 15 students in the Virtual class completed both the pre-test and post-test, and were included in the analyses. Analyses investigated whether there were learning differences between students who interacted with physical robots versus students who interacted with virtual robots.

Analyses

Three analyses were performed on the data. First, students' total scores on the pre-test were compared to their total scores on the post-test. To control for students' differing pre-test scores, an ANCOVA was run using condition (Physical, Virtual) as the independent variable, post-test score as the dependent variable, and pre-test score as the covariate.

Second, we examined whether learning differed across topic sub-categories. Four sub-categories were defined, into which all problems on the pre-test and post-test could be placed. These sub-categories were:

> *Algorithmic thinking*: Problems that involved thinking through the process of the programming problem (e.g., planning the program, using pseudocode, predicting what a program would do) or more abstract concepts of programming. Example: "Simple behaviors are made up of complex behaviors: True/False"
>
> *General programming*: Problems that involved syntax or concepts that are applicable to multiple programming languages. Example: "If the condition of an If statement is true, then all of the code inside of its curly braces will run: True/False"
>
> *ROBOTC Syntax*: Problems that involved ROBOTC syntax or the ROBOTC program (e.g., how to use menus in the ROBOTC application). Example: "To make the robot stop, you set its motor values equal to __"

Physical Robot: Problems that involved the physical VEX robot's functioning. Example: "The VEX Ultrasonic Rangefinder (sonar sensor) measures distance using __"

The number of problems in each sub-category and the Cronbach's alpha (α; calculated using both conditions' post-test scores) for each category are shown in Table 1. Note that the sub-categories were not mutually exclusive; that is, the same problem could fit into multiple sub-categories.

Problem Sub-Category	**Number of Problems**	**α**
Algorithmic Thinking	4	.54
General Programming	13	.56
ROBOTC Syntax	37	.84
Physical Robot	19	.81

Table 1. Number of sub-category problems and sub-category alphas.

Due to the uneven number of problems in each category, we used the proportion of correct answers within each category as a measure of accuracy. An ANCOVA was performed to control for pre-test scores, using condition as the independent variable, post-test score as the dependent variable, and pre-test score as the covariate.

Thirdly, we looked at the amount of days between participants' pre-test attempt and post-test attempt. This was used as a measure of the time needed to complete the course, to see whether one condition required less time than the other to learn the same amount of information. A one-way ANOVA was performed, using condition as the independent variable and number of days as the dependent variable.

Results

Overall Scores

No differences were found between the Physical and Virtual class in their overall post-test scores (with pre-test score controlled) [$F(2, 23)=0.19$, $p=0.67$] when pre-test scores were controlled. Both classes began with similar pre-test scores and ended with similar post-test scores. Figure 1 shows that overall learning gain did not differ by pre-test score, as almost all participants improved regardless of their pre-test score. The average pre-test and post-test scores for both classes can be found in Table 2.

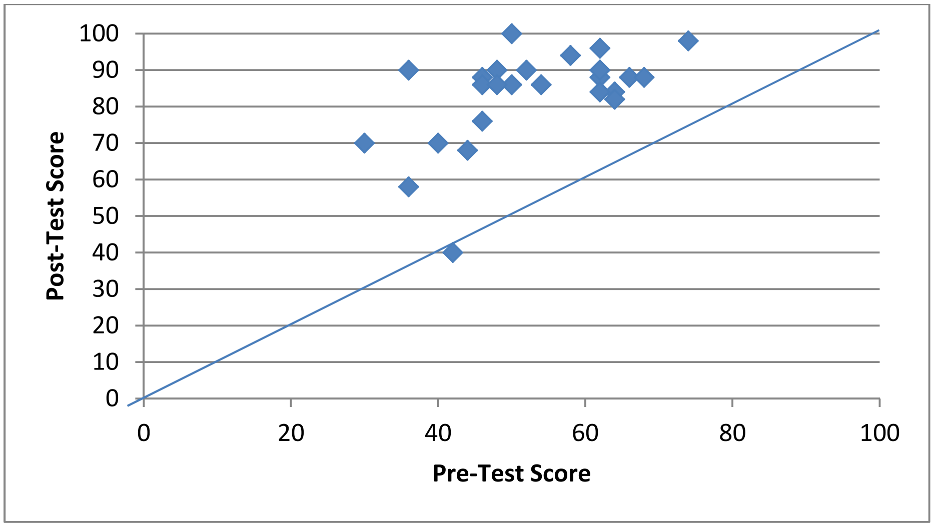

Figure 1. Pre-test score vs. post-test score. Points above the line improved on the post-test compared to pre-test.

Condition	Pre-Test Average	Post-Test Average	Average Time Taken
Physical	50.2 (*SD*=11.2)	82 (*SD*=10.6)	85.0 (*SD*=0.0)
Virtual	55.9 (*SD*=11.5)	84.5 (*SD*=14.6)	54.7 (*SD*=18.2)

Table 2. Averages (and standard deviations) of pre-test score, post-test score, and time taken, separated by condition.

Sub-Category Scores

With pre-test score controlled, he two classes did not show any learning differences across the four sub-categories of algorithmic thinking [$F(2, 23)=0.061$, $p=0.81$], general programming [$F(2, 23)=1.3$, $p=0.27$], ROBOTC syntax [$F(2, 23)=0.079$, $p=0.78$], or physical robots [$F(2, 23)=0.11$, $p=0.74$]. The average pre-test and post-test scores (and standard deviations), measured as proportion correct, for each sub-category for the two classes can be found in Table 3.

Condition	Alg. Thinking Pre-Test	Alg. Thinking Post-Test	Gen. Prog. Pre-Test	Gen. Prog. Post-Test	ROBOTC Syntax Pre-Test	ROBOTC Syntax Post-Test	Phys. Robot Pre-Test	Phys. Robot Post-Test
Physical	0.67 (0.27)	0.88 (0.20)	0.54 (0.12)	0.79 (0.11)	0.49 (0.10)	0.81 (0.11)	0.42 (0.12)	0.82 (0.16)
Virtual	0.80 (0.29)	0.95 (0.14)	0.59 (0.12)	0.86 (0.15)	0.51 (0.11)	0.83 (0.16)	0.51 (0.14)	0.86 (0.17)

Table 3. Average proportion correct (and standard deviations) of each sub-category, separated by condition.

Time Taken

The average time taken for both classes to complete the programming course can be seen in Table 2. The Physical class took significantly more time than the Virtual class [$F(1, 24)=30.3$, $p<0.001$]. All students in the Physical class completed the course in the same amount of time, as working with Physical robots did not afford them the same freedom of students in the Virtual class of working independently through the course. Thus, due to heterogeneity of variance, we also ran a t-test with equal variances not assumed, which was significant [$t(24)=6.5$, $p<0.001$, $d=2.2$]. Overall, the Physical class took an extra 30.3 days (approximately one month) to complete the course than the Virtual class (see Figure 2).

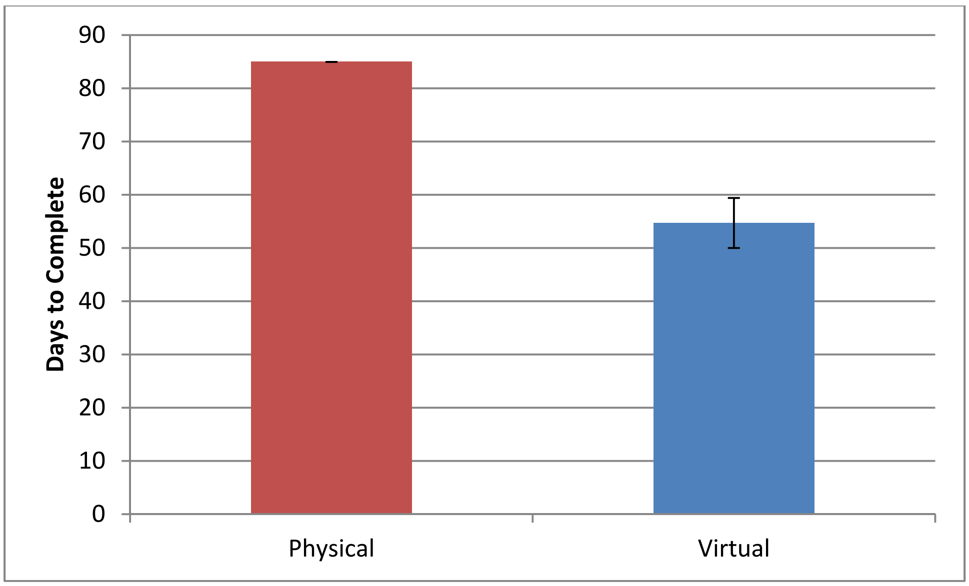

Figure 2. Days taken to complete the course, separated by condition.

Summary

Both the Physical class and the Virtual class showed equal learning gains, as their overall post-test scores were the same (controlling for pre-test scores). The type of learning did not differ between the two classes either, as evidenced by the equal learning gains seen across all four sub-categories. However, the Virtual class did show a time reduction benefit, as they completed the course about a month earlier than the Physical class, with no effect on their overall learning. This suggests that working with the virtual robots allowed students to learn more efficiently in this context when compared to the physical robots.

The teacher's informal observations support this conclusion. The teacher noted that students in the Physical class had to deal with the additional mechanical issues that came from working with a physical robot. Consequently, the teacher spent most of his time in the Physical class helping students with communication problems between the robot and computer. In the Virtual class, he and his students were able to focus their time on programming instead of the mechanical side.

Comparison with Other Virtual Classes

To confirm that the learning gains and time savings seen in the Virtual class were consistent, we also looked at two additional classes who completed the same programming course with virtual VEX robots. A graph comparing the three courses' pre-test scores, post-test scores, and days to complete the course can be seen in Figure 3. One class (Class 2) had 23 students who completed both the pre-test and post-test, and the other class (Class 3) had 13 students who completed both the pre-test and post-test.

We ran a paired t-test for each class to compare total pre-test scores to total post-test scores. Average pre-test score and post-test score for both classes can be seen in Table 4. Both Class 2 [$t(22)=-14.4$, $p<0.001$] and Class 3 [$t(12)=-7.1$, $p<0.001$] significantly improved their scores on the post-test, suggesting that the course led to comparable programming learning gains across the three Virtual robot classes.

To compare the time taken by each class to complete the course, we ran a one-way ANOVA, using Class (Class 1 being the Virtual class in the study above, Class 2 being the class of 23 students, and Class 3 being the class of 13 students) as the independent variable and number of days as the dependent variable. Average number of days taken to complete the course can be seen in Table 4. The ANOVA was significant [$F(2, 48)=31.6$, $p<0.001$], and a post-hoc contrast showed that Class 3 took significantly less time than the other two classes. Again, due to heterogeneity of variance between Class 3 and the other two classes, we also ran an independent t-test comparing the amount of time taken by Classes 1 and 2 (combined) and Class 3, which was significant [$t(49)=13.2$, $p<0.001$]. This suggests that Virtual robots in all three Virtual classes allowed students to complete the course in significantly less time than the Physical class in the study above.

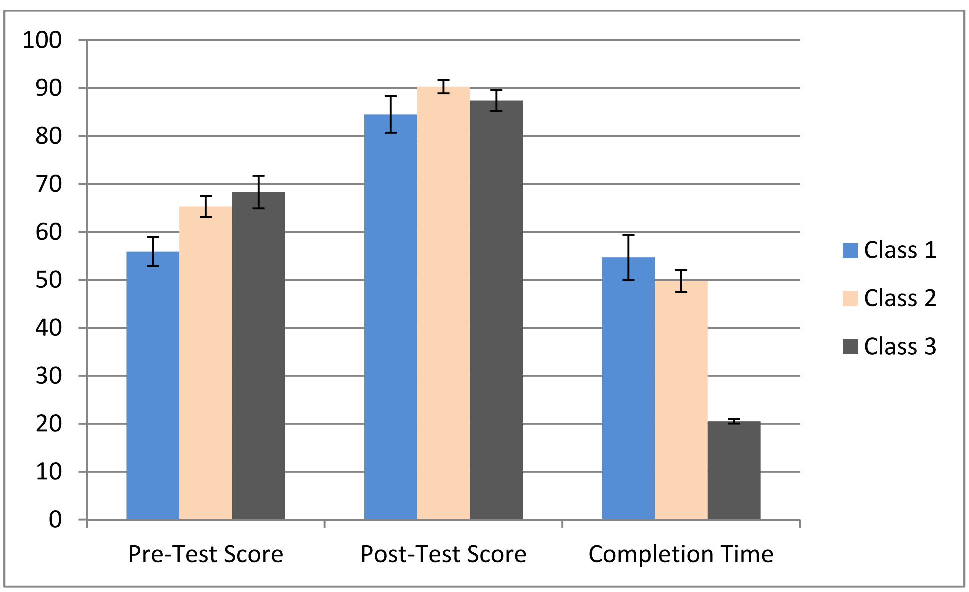

Figure 4. Average pre-test score, post-test score, and days to complete the course, separated by class.

Class	Average Pre-Test Score	Average Post-Test Score	Average Days Taken
Class 2 (N=23)	65.3 (*SD*=10.5)	90.3 (*SD*=6.7)	49.8 (*SD*=12.3)
Class 3 (N=13)	68.3 (*SD*=12.3)	87.4 (*SD*=8.1)	20.5 (*SD*=1.7)

Table 4. Average pre-test score, post-test score, and days taken to complete the course (and their standard deviations), separated by class.