# Tug of War: What is it Good For?
# Measuring Student Inquiry Choices in an Online Science Game

Nicole R. Hallinen[1], Julius Cheng[1], Min Chi[2], Daniel L. Schwartz[1]
[1]Stanford Graduate School of Education, 485 Lasuen Mall, Stanford, CA 94305
[2]North Carolina State University, Computer Science Department, Raleigh, NC 27695
hallinen@stanford.edu, juliusc@stanford.edu, mchi@ncsu.edu, danls@stanford.edu

**Abstract:** We designed and tested a computer game to measure middle-schoolers' science inquiry. In the game, students can run experiments or answer challenge questions. Students who use more of their chances for experimentation performed better on challenge questions and a posttest. Shorter times per experiment were associated with higher science grades. Choices to engage in inquiry predicted academic achievement better than accuracy. We conclude that science-learning assessments should measure inquiry choices in addition to knowledge.

## Measuring Students' Inquiry in Scientific Phenomena

Exploration allows scientists to test and form hypotheses about new domains. For students to become literate in the practices of science, they should engage in exploration and hypothesis testing, hallmarks of inquiry approaches to science instruction (e.g. National Research Council, 2007; Dunbar, 1993). However, current knowledge-based assessment strategies do not measure inquiry and exploration. We propose that using educational software to record students' choices can provide more information about their learning than typical assessments (Schwartz & Arena, 2013). Choice-based assessments are explicitly designed for open-ended exploration and are fundamentally different from technologies that make choices for the student.

We created a game to examine students' inquiry choices when learning a new idea in science. To win the game, students need to correctly complete eight questions in a row. They have repeated chances to run their own brief experiments about the topic or simply answer the eight challenge questions. We tested the game with eighth grade students and present preliminary findings relating their inquiry choices and academic achievement. We show that measuring inquiry, even with simple data-mining techniques, is useful for assessing students.

## Method

Eighth grade students ($n$ = 136) from a suburban middle school played an interactive computer game during class time late in the school year. The school was comprised primarily of Asian, Filipino, and Hispanic or Latino children; approximately 40% were socioeconomically disadvantaged. Thirty-six participants were excluded because of missing permissions and an additional nine were removed due to computer error. Our final dataset consists of 91 students. Participation was voluntary and did not affect students' class grades.

### Computer Game Environment

In the Tug-of-War science computer game, modeled after a PhET simulation on forces and motion (http://PhET.colorado.edu), students learn to predict the outcome of a simulated tug-of-war between two teams of up to four characters. The winning team is determined by summing the strength values of characters on the team, where the strength values are 1, 2, and 3 for the small, medium, and big characters, respectively. The character's position along the rope does not affect the result. Students were not given knowledge of either of these determinants and this topic was not previously covered in their science curriculum.

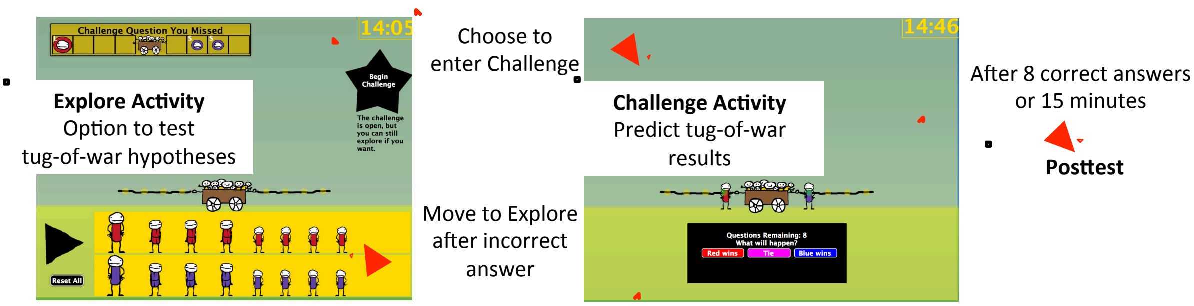


Figure 1. Paths through the Tug-of-War Explore and Challenge Activities

Students begin by manipulating a subset of the characters on a brief introduction screen designed to familiarize them with the Tug-of-War interface. After 60 seconds, students may enter the Challenge Activity, which involves predicting which team will win a tug-of-war (red team, blue team, or tie). Questions become progressively more complex and are randomly drawn from a bank of possible configurations. Students must correctly answer eight consecutive questions to succeed in the challenge.

When students answer a challenge question incorrectly, they are sent to the Explore Activity. There, they can place characters on the tug-of-war teams and view the outcome as often as they want. This enables students to test hypotheses about the tug-of-war motion. As such, we call each "set-up and view" sequence a hypothesis test. Students are free to reenter the Challenge Activity at any time, even without testing any hypotheses. The Tug-of-War game ends either when a student succeeds in the challenge by answering eight questions or after 15 minutes have elapsed. Next, all students completed a brief computerized posttest that showed one side of a tug-of-war configuration and displayed an array of ten possible opposing teams. Students were asked to select all of the teams that would tie against the example team.

## Results

We report two measures of students' exploration choices. First, we computed a *proportion of exploring* variable. At two points in the game, students have the choice of performing a hypothesis test or entering the Challenge Activity. These points occur when students enter the Explore Activity and after each hypothesis test. We call these decision points *opportunities to explore*. *Proportion of exploring* is computed by dividing each student's number of hypothesis tests by his or her total opportunities to explore. The average value for this variable was 0.44 (SD = 0.23), indicating that students took advantage of 44% of the opportunities to explore. On the remaining opportunities, students returned to the Challenge Activity. Next, each student's *mean explore time* was computed by finding the average time he or she spent on each hypothesis test. Students spent an average of 13.12 seconds (SD = 6.50) per hypothesis test.

### Game Choices Predicting Posttest Scores

Students' scores on the posttest served as a measure of in-game learning. On average, students performed at 79% accuracy in selecting the correct teams. Twenty-seven students received a perfect score on the posttest. We found that students who completed the challenge (79 out of 91) during the game performed better on the posttest than those students who did not complete the challenge ($t(89) = -2.57$, $p = 0.01$). Next, we investigated whether children's exploration choices affected their posttest scores. While *mean explore time* did not predict posttest outcomes ($r = -0.09$, $p > 0.05$), *proportion of exploring* positively correlated with posttest scores ($r = 0.21$, $p < 0.05$). Students who chose to use more exploration opportunities learned more.

### Game Choices Predicting Academic Achievement

We obtained students' second trimester $8^{th}$ grade science class grades. Students' average class grade was 81% ($SD = 9\%$). Students' posttest performance was correlated with class grades ($r = 0.29$, $p < 0.01$). We examined relationships between exploration and science class grade. *Proportion of exploring* was not significantly related to class grade ($r = 0.05$, $p > 0.05$). We found that *mean explore time* negatively correlated with class grade ($r = -0.29$, $p < 0.01$); students who complete hypothesis tests faster tend to have higher class grades. Perhaps this implies that students who are more efficient with their in-game exploration do better in school tasks, which fits with the design-thinking model of encouraging rapid iteration. Moreover, *mean explore time* provides predictive power for class grade beyond posttest performance ($r = -0.28$ $p < 0.01$).

## Implications

Most academic assessments place a great deal of emphasis on retrieving verbal memories and executing procedural skills. Outside of school, however, students must continue to learn, as classrooms cannot prepare individuals for everything they must know. Here, we created a computer game where students had to learn to beat the game. We showed that inquiry choices could predict student performance within and outside the game. These choices within the game, not only posttest performance, predicted academic performance.

Going forward, we plan to pursue more fine-grained analyses to better characterize students' exploration patterns. We have created variables to describe the semantics of students' hypothesis tests, such as recording if students isolate an individual character on a side of the tug-of-war. We hope that investigations of detailed choices will complement our results and reveal more about students' inquiry in science domains.

## References

Dunbar, K. (1993). Concept discovery in a scientific domain. *Cognitive Science, 17*, 397-434.

National Research Council (2007). *Taking Science to School: Learning and Teaching Science in Grades K-8.* Washington, DC: National Academy Press.

Schwartz, D.L. & Arena, D. (2013). *Measuring What Matters Most*. Cambridge, MA: MIT Press.

## Acknowledgments

We thank Jacob Haigh and Neil S. Levine for their work programming and designing the Tug-of-War.