Educational Data Mining: Illuminating Student Learning Pathways in an Online Fraction Game

Ani Aghababyan[1]
Taylor Martin[2]
Utah State University
Instructional Technology
and Learning Sciences
2830 Old Main Hill
Logan, UT 84322-2830
+1-435-754-6285[1]
+1-435-797-0814[2]
anie.aghababyan@gmail.com[1]
taylormartin@usu.edu[2]

Nicole Forsgren Velasquez
Utah State University
Management of Information Systems
3515 Old Main Hill
Logan, UT 84322-3515
+1-435-797-3479
nicolefv@gmail.com

Philip Janisiewicz
University of Texas at Austin
Curriculum and Instruction
1 University Station D5700
Austin, TX 78712
(512) 496-1829
pjanisiewicz@gmail.com

ABSTRACT
This study investigates ways to interpret and utilize the vast amount of log data collected from an educational game called Refraction to understand student fraction learning. Study participants are elementary students enrolled in an online virtual school system who played the game over the course of multiple weeks. Findings suggest that students use a variety of splitting strategies when solving Refraction levels and that these strategies are related to learning gains.

Keywords
Educational data mining; hierarchical clustering; learning analytics; mathematics education; fractions; educational games.

1. INTRODUCTION
Electronic games have become a regular part of childhood and adolescence [1]. In recent years, the interest in games for learning has grown, and educational games have increased in their popularity as means of instruction [3]. These games are unstructured environments where students can learn educational concepts through engaging interfaces and at their own pace.

Educational data mining techniques have the potential to illuminate learning patterns across a large number of students who play these games. By analyzing the data generated through these activities and assignments, data scientists can gain insights into when students have mastered a concept or skill, what excites them, where they are getting stuck, and what works to support learning. The ability to discern this for each student and for all students is a key contributing factor in improving the quality of education in the U.S.

2. REFRACTION
Refraction (http://play.centerforgamescience.org/refraction/site/) is an online game based on fraction learning through splitting. It is an open-access, interactive, and spatially challenging game that allows researchers to discover students' fraction learning pathways. In the game level used for this study, students are required to create laser beams of 1/6 and 1/9 using a combination of 1/2 and 1/3 splitters. Four 1/2 splitters, four 1/3 splitters, and seven benders are provided to achieve this goal. One possible solution is shown in Figure 1(b).

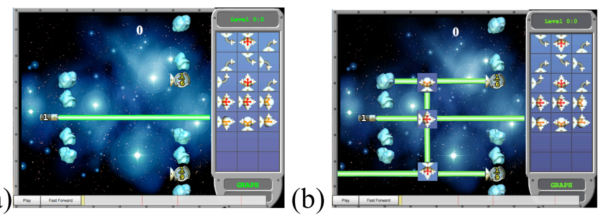
(a) (b)


Figure 1. Refraction pretest and posttest levels (a) at the start of a gameplay and (b) at the completion of the level

3. METHOD
3.1 Procedures
The game begins with a short series of introductory levels, and then students play an in-game pretest level. Following gameplay, which can include several levels, they play an in-game posttest level. The pretest and posttest levels are identical. For this study, we mined only the in-game pretest and in-game posttest levels. We chose this problem because it requires students to move beyond repeated halving (such as creating 1/4 or 1/8), and it requires students to use a combination of 1/2 and 1/3 splitters. Players reach the pretest after introductory levels that teach them the mechanics of the game, in order to avoid any pre-post differences simply being attributable to game's unfamiliarity.

3.2 Variables
Every time a splitter is placed on the laser beam in the Refraction environment, a new board state is logged. We used this data to examine the process of learning by splitting using hierarchical cluster analysis and included the following variables:

a. *Initial 1/2*: this is the percent of board states in a level that have 1/2 splitter as the initial splitter.
b. *Initial 1/3*: this is the percent of board states in a level that have 1/3 splitter as the initial splitter.
c. *Backtrack*: this is a binary variable indicating whether the player returned to using 1/2 as the initial splitter after having used 1/3 as the initial splitter.
d. *Average distance from goal*: Average distance from each board state to the goal state.

3.3 Hierarchical Clustering

Cluster analysis is a common technique for classifying a large amount of information into meaningful groups [2]. Hierarchical cluster analysis was conducted using between-groups linkage, within-groups linkage, centroid clustering, median clustering, and Ward's method [4]. The results for cluster solutions with two to seven clusters were compared in terms of: (a) change in agglomeration coefficients; (b) number of cluster memberships (number of students in each clusters solutions; and (c) results of univariate F-tests (solutions wherein the clusters did not differ in terms of any of the classifying variables were excluded). Based on these analyses, we found the solution with four clusters using between-group linkage method performed the best. Based on F statistics (see Table 1), these four clusters were significantly different from each other in terms of each of the four variables presented above.

Table 1 Cluster definitions based on Duncan post hoc multiple range test.

Variables	*F-values*[a]	*Cluster 1 Halving Strategy*	*Cluster 2 Thirds Strategy*	*Cluster 3 Exploring Thirds Strategy*	*Cluster 4 General Exploring Strategy*
Backtrack	1017.187***	0.58 M[b]	0.00 L	0.00 L	0.63 H
Initial 1/2	17415.955***	3.57 VH	1.00 L	1.04 M	2.99 H
Initial 1/3	4800.857***	1.46 L	3.53 VH	1.82 M	2.88 H
Avg. Distance from Goal	5816.343***	3.58 VH	1.44 L	3.03 H	2.12 M

[a] The significance levels of these F-values are indicated as follows: *** 0.001 level; ** 0.01; and * 0.05 level.

[b] VH, H, M, and L indicates that the mean for the cluster was very high, high, medium, or low, respectively, based on Duncan's Multiple Range Test. If two clusters have the same letter, that indicates that these clusters' means did not have significantly different means.

To interpret the clusters, post hoc comparisons of the means of all four clustering variables were performed. Duncan's Multiple Range Test was used to compare the means of these four variables across four clusters (Hair, Anderson, Tatham, & Grablosky, 1979). In this test, pairwise comparisons are done across clusters and significant differences are identified at the pre-defined significance level; in this case, $p < 0.1$. Furthermore, the test sorts the clusters into groups wherein the means of the clusters within a group are not significantly different from each other, but differ at a statistically significant level from clusters in other groups. For example, for the variable backtrack, the test sorted our four clusters into three distinct groups, as seen by the designation of L, M, and H in Table 1. In this case, the mean of Cluster 1 is significantly higher than the means for Clusters 2 and 3, but significantly lower than the mean for Cluster 4.

4. RESULTS

Clustering results show that there are four distinct ways that students solved the pre-post level. The four clusters can be described as follows:

a. *Halving strategy:* Students using this strategy are primarily exploring the 1/2 space of the game. They display a high percentage of board states that start with a 1/2 splitter. They also have high average distance from the goal and a low percentage of board states that start with a 1/3 splitter. When they do use the 1/3 splitter, they often backtrack to using the 1/2 splitter.

b. *Thirds strategy:* Students using this strategy spend the majority of their time in the 1/3 space of the game. They have a high percentage of 1/3 initial board states. They rarely backtrack. Their average distance from the goal is low.

c. *Exploring Thirds Strategy:* While students using this strategy still experiment with initial 1/2 board states, they have a higher percentage of 1/3 initial board states. They do not backtrack often, but still have a high average distance from the goal.

d. *General Exploring Strategy:* Students using this strategy are exploring the mathematical space of the game more broadly. They have high percentage of board states using both the 1/2 and 1/3 splitters, and they backtrack often. They have medium average distance from the goal.

We conducted one-way ANOVAs on the classifications of students' game play strategy (pre- and postlevel) for both the pre- and posttest. We found that there was a significant main effect for prelevel strategy type on transfer pretest score, $F(3, 2494) = 7.79$, $MSE = 23.10$, $p < .001$. Post hoc tests showed that this effect was primarily accounted for by the Thirds group's significantly greater performance than the Halving and Exploring Thirds groups ($p < .05$). The General Exploring group did not perform significantly differently than any other group.

5. DISCUSSION

Overall, we found that how students used splitting on the prelevel was associated with test performance at that point, but all students developed fraction knowledge by using splitting as they played Refraction, regardless of their splitting strategy.

6. REFERENCES

[1] Greenberg, B. S., Sherry, J., Lachlan, K., Lucas, K., & Holmstrom, A. 2008. Orientations to video games among gender and age groups. Simulation & Gaming, 41(2), 238–259.

[2] Lorr, M. (1983). *Cluster analysis for social scientists*. Jossey-Bass, San Francisco, CA.

[3] Rodrigo, M., Baker, R., D'Mello, S., Gonzalez, M., Lagud, M., Lim, S., … & Viehland, N. (2008). Comparing learners' affect while using an intelligent tutoring system and a simulation problem solving game. Intelligent Tutoring Systems. 40-49.

[4] Ulrich, D. and B. McKelvey (1990), "General Organizational Classification: An Empirical Test Using the United States and Japanese Electronic Industry," Organization Science, 1, 99-118.