Predicting Student Learning using Log Data from Interactive Simulations on Climate Change

Elizabeth McBride
UC Berkeley
4515 Tolman Hall
Berkeley, CA 94709
bethmcbride@berkeley.edu

Mario Martinez
UC Berkeley
Tolman Hall
Berkeley, CA 94709
mario.m.martinez@vanderbilt.edu

Jonathon M. Vitale
UC Berkeley
Tolman Hall
Berkeley, CA 94709
jonvitale@berkeley.edu

Zachary Pardos
UC Berkeley
4641 Tolman Hall
Berkeley, CA 94709
pardos@berkeley.edu

Hannah Gogel
UC Berkeley
Tolman Hall
Berkeley, CA 94709
hgogel@berkeley.edu

Marcia C. Linn
UC Berkeley
4611 Tolman Hall
Berkeley, CA 94709
mclinn@berkeley.edu

http://dx.doi.org/10.1145/2876034.2893410

Abstract
Interactive simulations are commonly used tools in technology enhanced education. Simulations can be a powerful tool for allowing students to engage in inquiry, especially in science disciplines. They can help students develop an understanding of complex science phenomena in which multiple variables are at play. Developing models for complex domains, like climate science, is important for learning. Equally important, though, is understanding how students use these simulations. Finding use patterns that lead to learning will allow us to develop better guidance for students who struggle to extract the useful information from the simulation. In this study, we generate features from action log data collected while students interacted with simulations on climate change. We seek to understand what types of features are important for student learning by using regression models to map features onto learning outcomes.

ACM Classification Keywords
K.3.1 [Computers and Education]: Computer Uses in Education.

Introduction
Interactive simulations are commonly used tools in technology enhanced education. Simulations have the unique affordance of being able to connect small-scale and

large-scale phenomena for students [5]. Connecting the smaller-scale science mechanism to the larger-scale outcome is often confusing for students. Simulations can be used as inquiry activities, especially in science disciplines. In inquiry activities, students are able to develop their own knowledge by asking scientific questions, answering those questions using evidence, developing explanations, and connecting explanations to scientific knowledge [4].

In this study we use a series of four simulations that were part of a week-long curriculum on global climate change. This curriculum was run using the Web-based Inquiry Science Environment (WISE). This particular unit on global climate change had two foci: establishing a causal mechanism between increased concentrations of carbon dioxide and temperature and confronting the commonplace but incorrectly held link between ozone depletion and atmospheric temperature.

We employ feature engineering to understand how students are using the simulations and which activities are most productive for learning, as evidenced by gains in pre- or posttest scores or embedded items. Previous studies have investigated clustering of similar features for use in predictor models [1, 2]. However, since our simulations were embedded within a longer unit, we are also interested in investigating student activity at different visits (first, second, etc.).

NetLogo Models

The four simulations in the unit progress in complexity between the first and the fourth. During the first simulation, students view solar radiation coming from the sun. The solar radiation is either absorbed or reflected by the surface of the earth. The temperature of the atmosphere is shown in a dynamically generated graph next to the simulation of temperature vs. time. In this simulation, students are able to use a slider to change the speed of the simulation or to highlight a single ray and watch it transform into heat, then into infrared radiation. In the second simulation (Figure 1), students are also able to add and remove greenhouse gases. A second graph is added, plotting concentration of carbon dioxide vs. time. In the third and fourth simulations students investigate the role of ozone in global warming, and can add and decrease ozone. We do not analyze the ozone simulation on its own here because there were not clear student gains from this portion of the curriculum. In each of the four simulations, students must press ''Go'' to start or restart the simulation. Students were able to visit simulations as many times as they wanted.

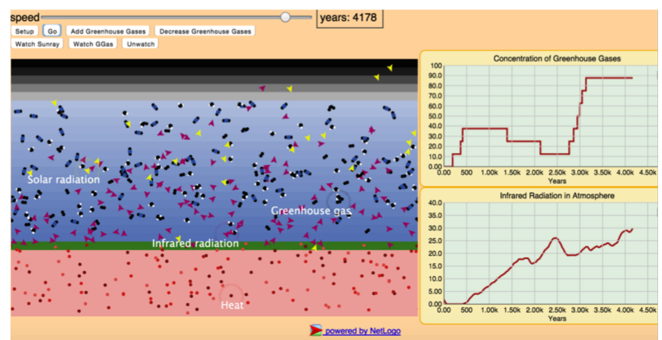

Figure 1: An Example of the NetLogo Simulations students interact with during the unit

Data

Our data on student interactions with simulations includes timestamped actions. Students also completed pre- and posttest items that measured student ability to differentiate between the effects of ozone depletion and

Variable	Mean	St.Dev.	Min	Max
Slider	33.35	126.6	0	1321
Go	21.44	16.92	0	135
Watch Ray	7.10	11.38	0	93
Graph Press	16.77	39.27	0	410
Time Spent (s)	1621.27	801.9	0	4478
Visits	6.49	5.21	1	56
Actions	328.17	288.15	2	1793

Figure 2: Descriptive statistics of generated features

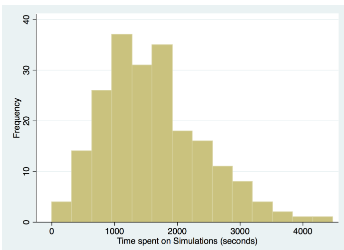

Figure 3: Amount of time (in seconds) spent on all four simulations

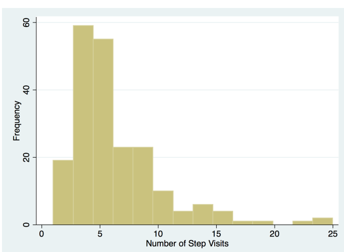

Figure 4: Number of visits students made to all four simulations

Variable	Mean	St.Dev.	Min	Max
Slider	10.06	55.18	0	752
Go	6.02	6.10	0	44
Watch Ray	2.80	5.99	0	46
Graph Press	4.67	13.60	0	131
Time Spent (s)	489.49	311.99	0	1690
Visits	3.40	2.42	1	13
Add GHGas	22.34	52.78	0	400
Decrease GHGas	28.74	47.50	0	367
Actions	117.0	154.06	1	1177

Figure 5: Descriptive statistics of generated features for only Simulation 2

greenhouse gases as well as ability to integrate ideas about energy transformation. Essay items were scored using a knowledge integration rubric [3]. Within the unit students engaged in concept mapping activities, also scored using knowledge integration rubrics [3].

The data comes from 422 students in three classrooms from two different schools. During simulations and the unit, students worked in pairs, but students were assessed individually at the pre- and posttests. We average pre/posttest scores across student pairs during this analysis.

Methodology

We engineered features from the action log data. Using these engineered features and student scores on both embedded and pre/post items, we run regression models. We also examine differences in features based on visit number. We analyze the entire dataset, which includes all four simulations, for productive patterns. Since some feature types differ between simulations, we also analyze a single simulation (simulation 2) to examine features that may be unique by simulation. We chose simulation 2 because it has a wider range of control options than simulation 1. Prior knowledge can be approximated using scores from the first concept mapping activity (after simulation 1).

Feature Generation

We generated features that relate the actions done by students and the time spent on simulations. Using all simulations, these features are: 1) Time spent on simulations, 2) Number of visits to simulations, 3) Number of actions done, 4) Number of times student used the speed slider, 5) Number of times the student pressed ''Go'' to restart simulation, 6) Number of times the student resized graphs, and 7) Number of times the student clicked ''Watch Ray''. Figures 2 - 4 show descriptions of these features.

Using only data from simulation 2, we use the same features as above and add two new features: 8) Number of times a student clicked ''Add Greenhouse Gas'', and 9) Number of times a student clicked ''Decrease Greenhouse Gas''. We also examine student interaction with the model based on the visit number.

Effects of Features on all Simulations

Using a regression model to predict posttest scores that includes the pretest score as well as all features listed above (features 1-7) as covariates, we find that the pretest score [$\beta = 0.81$, $p < 0.001$], features of time spent on simulations [$\beta = 0.0006$, $p < 0.001$], and number of times clicked on a graph [$\beta = -0.006$, $p < 0.05$] are significant. This model also only predicts posttest scores moderately well, with an R-squared value of 0.40. While the time spent on a simulation is an expected predictor, the number of clicks on a graph may be more interesting. Clicking on the graphs may be distracting for students, as shown by a negative regression coefficient, but this needs further investigation to differentiate between how different types of students utilize the graphs in different ways.

Effects of Features on One Simulation

Using data from Simulation 2 across all visits, we ran a regression model to find features that were important to learning. This regression attempts to predict the score of the second concept map activity that follows simulation 2 using the first concept map (done after the first simulation) as a covariate. We use all other features described above (features 1-9).

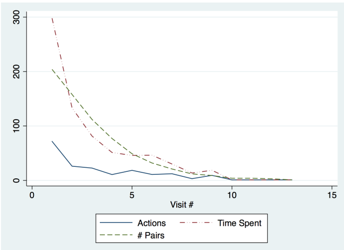

Figure 6: Averages of features by visit numbers

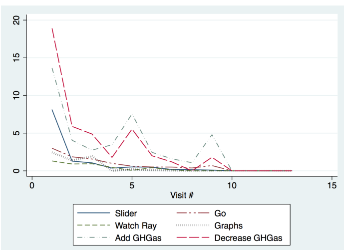

Figure 7: Averages of features by visit numbers

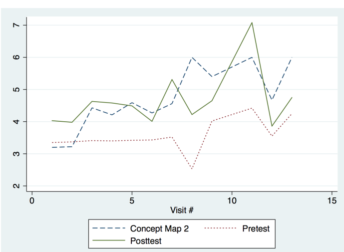

Figure 8: Averages of scores by visit numbers

Again, only the previous concept map score [$\beta = 0.55$, p <0.001] and time spent on the simulation [$\beta = 0.17$, p <0.001] were significant, and the regression model does not fit the data very well, with an R-squared value of 0.32. Descriptive statistics for this simulation can be found in Figure 5. It is interesting that this model explains such a low amount of variance since the concept maps offer a fairly close transfer of knowledge gained from the simulations.

Looking at changes in activity based on visit number, we find that students spend less time on average each time the revisit a simulation (Figure 6). In addition, only about half of students visit the simulation three or more times. Looking at Figure 7, we see that students do not restart the simulation, use the slider button, or ''Watch'' rays with any regularity after visit 2. However, students do continue to examine what happens after adding and decreasing greenhouse gases. The next concept map asks students to map the relationship between greenhouse gases and temperature using a map of how energy transforms. We see from Figure 7 that students may be focused on gaining specific pieces of information to help them construct their concept map during later visits to the simulation; students do not spend as much time or log as many actions during later visits.

Looking at Figure 8, we see that scores on the second concept map generally increase with number of visits. Score on the posttest also increases, but the relationship between number of visits and score is not as clear on the posttest. Interestingly, students who score higher on the pretest seem to revisit the simulations more often, suggesting that there may be a difference between how low and high prior knowledge students interact with the simulation.

Limitations and Future Work

Since students are expected to learn during each activity included in our curriculum, assessing all student interactions over time may be better done using a hidden Markov model. Future research may compare our engineered features to those generated using a deep neural network. This work also begins to examine differences between low and high prior knowledge students, but more work needs to be done. Examining why and when students revisit simulations is also a possibility for future research using a combination of NetLogo action log data and WISE action log data. One specific way to examine this data could also be in the context of spaced practice. Insight into how students revise their understanding of complex topics using the simulations would be important to developing automated guidance for simulations.

References

[1] Conati, C., Fratamico, L., Kardan, S., and Roll, I. Comparing representations for learner models in interactive simulations. In *Artificial Intelligence in Education*, Springer (2015), 74–83.

[2] Kardan, S., Roll, I., and Conati, C. The usefulness of log based clustering in a complex simulation environment. In *Intelligent Tutoring Systems*, Springer (2014), 168–177.

[3] Linn, M., and Eylon, B. *Science learning and instruction: Taking advantage of technology to promote knowledge integration.* Routledge, 2011.

[4] NRC. Inquiry. the national science education standards, 2000.

[5] Wilensky, U., and Reisman, K. Thinking like a wolf, a sheep, or a firefly: Learning biology through constructing and testing computational theoriesan embodied modeling approach. *Cognition and instruction 24*, 2 (2006), 171–209.