Design Matters: The Impact of Technology Design on Students' Inquiry Behaviors

Engin Bumbacher, Zahid Hossain, Ingmar Riedel-Kruse, and Paulo Blikstein
buben@stanford.edu, zhossain@stanford.edu, ingmar@stanford.edu, paulob@stanford.edu
Stanford University

Abstract: Recent curricular frameworks consider science inquiry as an intertwined set of practices revolving around data, models and theory. This poses major challenges on the design of tools to support science inquiry. We developed a novel hybrid technology for biology classrooms that combines remote laboratories with modeling tools. How to design such systems is of fundamental importance because the design influences students' learning processes (deJong, Linn & Zacharia, 2013). We examined the impact of the design of the modeling interface on learning, using two designs that differ in the type of visual feedback and the degrees of freedom for exploration. We found that neither of the designs was categorically better; rather, they were conducive to different forms of engagement in the inquiry activity, each offering distinct affordances for learning. This suggests that designers of technology for science inquiry need to be explicit about desired learning goals and forms of engagement.

Keywords: science inquiry, remote laboratories, inquiry strategy, modeling, interactive biology

Introduction

A central goal for science education is to help students become "critical observers" (Hodson, 1986), i.e. participants of scientific conversations that use evidence from real-world data to critically review and evaluate scientific claims. Essential to being a critical observer is the capability to *coordinate scientific ideas with real data* into evidence-based explanations and arguments (Duschl & Grandy, 2008).

Main attempts to facilitate science learning engage students in practices of science inquiry (Duschl & Grandy, 2008). However, there are many challenges to successfully integrating scientific practices into K-12 classrooms (Abd-El-Khalick et al., 2004; Chinn & Malhotra, 2002). Research has predominantly dealt with these challenges by focusing on scientific practices in isolation (Berland et al., 2016). While these reductive approaches helped students improve in the respective practices (Zimmerman, 2000), they have yet to prove successful in fostering critical observation (Berland et al., 2016; Chinn & Malhotra, 2002). This is not surprising, as coordinating scientific ideas with real data interweaves multiple practices at once. The bifocal modeling framework provides an approach to integrate practices instead of isolating them, by bringing scientific models and data into the same representational space for real-time comparison (Blikstein, 2014).

Inquiry-based learning activities hinge in large parts on the available technologies (Sandoval & Reiser, 2004). We argue that there is a shortage of technologies that facilitate more integrative approaches: Many technologies such as physical laboratories or interactive computer simulations generally focus on experimentation and interaction with the scientific phenomenon (de Jong, Linn, & Zacharia, 2013); they are not yet designed to facilitate more model-based inquiry practices (Blikstein, 2014). Technologies for scientific modeling on the other hand generally lack affordances for experimentation or interaction with real data (see VanLehn (2013)). Thus, if a science teacher wants to engage students in the core practices of coordinating scientific ideas with real data, she has to pick different technologies for each practice, and bring them together through the design of the learning activity. This is challenging given the range of logistical requirements, data formats, designs, etc.

A technology can be more conducive to integrative approaches if it provides affordances for various scientific practices within one system, while being easy to use and robust against the constraints of a classroom. In this paper, we present a new technology prototype for science inquiry in biology that integrates scientific models with real-world data into one system, drawing on the bifocal modeling framework. Biology is particularly interesting for such hybrid systems for the following reasons: First, recent technological developments gave rise to interactive biology, i.e. interaction with real living cells through various stimuli both remotely and in real-time (eg. Kim et al., 2016; Lee et al., 2015); second, biological phenomena are inherently noisy and complex processes that no model can fully account for, which necessitates explicit coordination of data and models.

There are many possible ways to implement such a technology design, using different visualizations, affordances and scaffolds, each of which is conducive to different science learning. Tools for the same scientific phenomena that differ in their affordances emphasize different aspects of the phenomena, are conducive to different ways of reasoning about them, and hence influence how students learn with them (Bumbacher et al., 2017; Wilkerson et al., 2017). It is not well understood though what dimensions of a technology design impact learning processes, and how (Bumbacher et al., 2017). Furthermore, there are multiple ways by which the effectiveness of a technology design can be assessed. A common approach is to look at learning outcomes alone. Other approaches incorporate measures of the inquiry process, for example of students' experimentation or parameter exploration strategies (e.g. Bumbacher et al., 2017). A third possibility is to assess the *cognitively alignment* of *actual* and *intended* technology use, i.e. the similarity of actual and intended cognitive and discursive processes of students as they work with the technology (Sandoval & Reiser, 2004). The choice of measurement can affect the evaluation of effectiveness of the technology (e.g. Bumbacher et al., 2017): For example, based on learning outcomes alone, one might conclude that affordance of quick variable manipulations is beneficial for learning (learners get exposed to more examples in the same amount of time; Zacharia & de Jong (2014)); however, examination of inquiry processes might suggest the opposite (quick manipulation can encourage learners to carry out play-like, undeliberated interactions; Renken & Nunez (2013)).

We employed our technology in middle school biology classrooms, using an inquiry unit that engages students in experimentation with the remote lab, and in modeling. In this paper, we will only talk about the modeling part, in order to address the question of how the design of technology impacts learning. We designed two different versions of the modeling interface and analyzed their influence on learning outcomes, exploration strategies and cognitive alignment.

Description of technology

We developed an interactive hybrid system that integrates a modeling interface with a remote laboratory, where students interact remotely with real living cells. The phenomenon under study is the phototactic behavior of *Euglena gracilis*, i.e. their movement in response to light stimuli. The remote lab is detailed in (Hossain et al., 2016) and was also incorporated in a MOOC (Hossain et al. 2017), but in short: Students can *remotely control in real-time* four different LEDs placed around the edge of the microscope plate holding the Euglena. Euglena sense light via one single photoreceptor, that can sample the entire space as the microorganism spins about its own body axis. The net result is that that the creature swims away from the LED light (negative phototaxis). This behavior is noticeable within a few seconds already, which makes it particularly well-suited for inquiry activities in class.

We implemented a model of the microorganism with only three parameters: a) *Speed* of the forward movement; b) *Coupling* – the direction and strength of the reaction to light; positive coupling leads to movement towards the light and negative coupling to movement away from the light; the magnitude determines the strength of coupling; c) *Roll* – the rotational speed about the body axis. In the *model exploration interface* (Figure 1), each of the three parameters is controlled by a slider. The parameters can take on only a discrete set of values. The model will never perfectly match the behavior of the real microorganism; there is no unique solution, but a subset of six optimal parameter values. The coupling parameter ranges from positive to negative values, which allows students to create both positive and negative phototaxis. Once the parameters have been configured, the system visualizes one three-dimensional model of the microorganism and simulates its behavior in reaction to the light sequence it is exposed to. Each simulation lasts about 30 seconds. Students can run as many simulations as they want.

Experimental conditions and research question

We created two designs of the model exploration interface that differed in the types of interactive affordances and types of feedback of the model exploration interface (Figure 1): The *Simultaneous (SIM)* condition is very much in line with the traditional bifocal modeling framework; students can see both model and real organism move at the same time, being exposed to the same, pre-programmed light sequence. The real data consisted of a recording of a real experiment with the given sequence of light directions. The *Light (LIGHT)* condition is more aligned with the remote lab interface in terms of how the light sequences were generated: Students could change in real-time the direction of light (by means of the joystick) during the simulation. However, they could only see the model organism and not the real one. Thus, students in the LIGHT condition could not directly compare the model and the real organisms, but in turn do real-time changes to the light intensity and direction for the model.

In sum, the SIM condition has a smaller degree of freedom of manipulation (parameters-only) than the LIGHT condition (light+parameter), and a richer type of visual feedback (model+real vs model-only). In this paper, we examine how the two conditions compare in terms of students' (i) learning outcomes, (ii) parameter exploration strategies and (iii) cognitive engagement with the behavior of model and real Euglena.

The rationale for these two designs was to create designs that were likely to elicit differences in inquiry processes, to get a better sense of the variation in inquiry processes and their interplay with the technology design. It was not to explore the impact of specific design dimensions – degrees of freedom or visual feedback – on student' inquiry processes and learning. Such a targeted study would be premature for this novel technology that has not been implemented in a classroom before. We selected two design dimensions that play an important role in how technology facilitates inquiry-based learning (Ainsworth & VanLabeke, 2004; Renken & Nunez, 2013) and that we could manipulate in our technology.

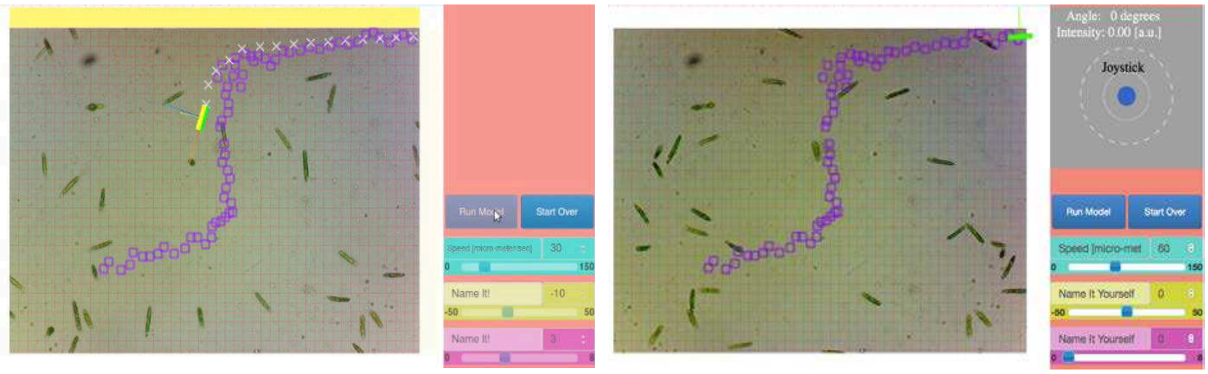

Figure 1. Schema of model exploration interfaces. Left: The SIM condition. Right: The LIGHT condition. The purple squares show the traced path of the real Euglena.

By keeping the light sequence constant, and providing a direct juxtaposition of real and model, we expected the SIM condition to foster more reflection on the interplay between model parameters and model behavior; we hypothesized that this would get manifested in two ways: 1. a more systematic exploration of parameters; 2. more comparisons the behaviors of model and real organisms. In contrast, by focusing on the model only, but introducing the degree of freedom of light, we expected students in the LIGHT condition to engage more in reflection on the dynamics of the model behavior, and on the interaction of light and model structure in the model behavior.

Methods and materials

Student and school sample

The study took place in 7th and 8th grade classes of a private K-12 school in the San Francisco Bay Area. Each class has 50 minutes of lab per week; classes are split in half for the lab session. Over the course of multiple weeks, the researcher team taught 6 sessions of about 8-12 students each, with a total of 59 students. The first two sessions were used to test and adjust the technology and lesson plan. The final study consists of the last four sessions, with 41 students (21 girls, 20 boys). Students worked in groups of 2 to 3, with an overall of 20 groups.

Model exploration activity

In both conditions, the goal of the activity was to discover the mechanism of how Euglena react to light; students were prompted to "understand the three parameters to find out what makes the organism see light from all directions. Find the values for the three different parameters that make the model follow the path of the real organism as closely as possible". In order to stress the discovery aspect, we labeled only the speed parameter explicitly, and left the other parameters unlabeled so students could come up with their own names for them. We added a traced path of a real organism in reaction to the pre-programmed light sequence of the real data used in the SIM condition (Figure 1). In both conditions, the model organism's initial position was at the beginning of that path. Students in the SIM condition could directly compare the model to the real organism that followed this path.

Study design and procedure

The four lab sessions were split equally between the two experimental conditions. The teacher guided the class through the lesson, but minimally engaged with the students during the activities. In the first part (10 minutes), the whole class explored the phenomenon with the remote laboratory; the microscope view of the system was projected onto the wall in front of the class, and one student controlled the light while others told this student what to do. In the second part (5 minutes), student groups examined experiment videos to evaluate and eventually confirm the hypothesis that the organisms move away from strong light. This was followed by a teacher-led classroom discussion (10 min) about the possible mechanisms of the organism behavior. Students came up with ideas about i) how the organism sensed the light (e.g., heat, electricity, vision, etc.) and ii) the potential mechanism. The teacher eventually resolved the first question by showing a microscope view of the single eye-like organelle of the organism. In the third part (~12 minutes), students engaged in the *Model Exploration Phase*: After the activity, students individually completed the test of learning outcomes (10 min).

Data collection

Assessment: We assessed learning outcomes by means of a 5-question post-test. We did not give a pre-test because students had not learned about Euglena phototaxis. The *light question* asked students to infer from a given organism path what light sequence it must have been exposed to. The sequence consists of five direction changes, and every direction was worth 1.0 point. The *parameter questions* were three questions that provided different scenarios of how a model path differed from a real path and asked students to identify what parameter needed to be changed, and how, in order to align the paths of the model and real organism. Each question was given 0.5 points for a partially correct answer, and 1.0 point for a complete answer. **Interaction logs:** We assessed model exploration strategies based on students' interaction logs. Every time students ran the simulation, we logged the current parameter configuration. We also recorded any string students entered in the textboxes for the two parameters. Additionally, for the LIGHT condition, we collected the light sequence of a simulation every time it reached the maximal duration of 30 seconds. **Video data:** In order to examine the cognitive engagement in the inquiry activity, we used video and audio recordings of each student group during the Model Exploration Phase, with the camera pointed towards the computer. This gives us a total of 20 videos of about 12 minutes.

Analysis

Learning outcomes were analyzed on the individual level, and strategy use and student discussions during the experimentation were analyzed on the group level. We found no significant intra-class correlations (less than 5%) in the analysis of student performance by means of two-level mixed models with students nested in groups. Thus, we employed fixed effects MANOVA and ANOVA models, as well as t-tests.

Model exploration strategy

We characterize model exploration strategies by the types of manipulations of any of the three parameters, and the time between manipulations: 1. Manipulations of only one parameter at a time (CTRL); 2. Manipulations of more than one parameter at a time (MIX); 3. Repetitions of preceding parameter configurations (REP); 4. Short experiments (BURST). This characterization builds on previous work on productive exploration strategies (Bumbacher et al., 2017), where we found that short durations between runs (BURST) and confounded parameter manipulations (MIX) were less productive for learning. For each student group, we calculated a 4-dimensional strategy vector with the manipulation types, coding each as percentage of total simulation runs per group. We calculated the proportion of bursts based on the between-manipulation times, as specified in Bumbacher et al. (2017).

Cognitive engagement with real and model Euglena

In order to evaluate cognitive engagement with the phenomenon, we extracted from the audio data simple frequency measures of three codes: 1. Reflections on the functionality of the parameter, and on model characteristics; 2. Comparisons between real and model behavior; 3. Comments about the purpose of the task, i.e. matching the model to the given path. We chose those dimensions because they are reflective of the coordination of scientific ideas with data within a simple inquiry task. We chunked each conversation into short segments of 1 – 3 conversation turns, by topic of conversation, and coded each segment as either "Reflection", "Comparison", "Path-Matching", or "Other" (see Table 1 for examples).

Table 1: Example conversation chunks for each code

"Reflection"	Student1: he needs to take longer to see it, smart huh? Cause it is going to take him longer for him to notice that there is light over there. Right?
"Comparison"	Student1: So, I see the regular one (real Euglena), I think it has to be a bit slower
"Path-Matching"	Student2: Really close, we have the right speed, maybe we should turn the speed up one. Student1: … look, it's just this last turn, like right around here it starts going off course.

Results

Overall effectiveness of the model exploration activity

Students across both conditions executed the simulation in average 23.1 times (SD=4.3), and manipulated each parameter in at least about 20% of the experiments. Student groups in general converged on parameter configurations that enabled them to discover the functionality of all parameters: 12 out of 20 groups found one of the six optimal solutions, while each of the remaining 8 groups had configurations with a negative coupling and a non-zero rotation. There were no differences by condition, p>.3. Students' understanding of parameters is further reflected in how the groups named them: They named the coupling parameter "light sensitivity" (8), "reaction" (3) or "attraction" (3) to light; two names were unclear (number of groups in parentheses). They named the roll parameter "rotation / turning of the eye" (6), "rotation speed" (7), "eye sight" (1); two names were unclear. Four groups did not name the parameters.

Learning outcomes by condition

A MANOVA of the inference question on condition was significant, Wilk's λ=0.80, $F_{(2,28)}$=3.4, p=.05 (see Table 2). The LIGHT condition was marginally better on the light sequence question than the SIM condition, t(39)=-1.8, p=.07, d=0.6. Students in the SIM condition omitted certain light directions, but did not mention wrong directions. The SIM condition performed better on the parameter questions, t(29)=2.2, p=.04, d=0.8. The ten students who did not answer this question were evenly split between the conditions.

Table 2: Descriptive statistics for the inference questions, normalized by the maximal possible scores

		SIM			LIGHT		
Inference Questions	Max	M	SD	N	M	SD	N
Light Sequence	5	0.68	0.26	22	0.82	0.21	19
Parameter Adjustment	3	0.63	0.27	17	0.39	0.33	14

Model exploration strategies by condition

A MANOVA on the percentages of manipulations of the three parameters reveal a trending difference between conditions, Wilk's λ=0.7, $F_{(3,16)}$=2.2, p=.13. We clustered the strategy vectors using hierarchical cluster analysis using Ward's method on the cosine distance between the vectors. We found two well-defined clusters (avg silhouette value = 0.6). These clusters can be characterized as *systematic* and *non-systematic* in terms of model exploration strategies (Figure 2): The systematic cluster had in average a significantly higher CTRL, t(18)=5.1, p<.001, d=2.3, a significantly lower REP, t(18)=-5.5, p<.001, d=-2.5, and a significantly lower BURST, t(18)=-3.1, p<.01, d=-1.4. There was no difference in MIX, p>.3.

Table 3 shows that the majority of SIM groups belong to the systematic cluster, while the majority of LIGHT were in the non-systematic cluster, Fisher's p<.01. The clusters differed also in terms of performance on the inference questions, Wilk's λ=0.8, $F_{(2,29)}$=2.9, p=.07. While there was no difference on the light question, p>.5, the systematic cluster was significantly better on the parameter questions, t(29)=2.4, p=.02, d=0.9.

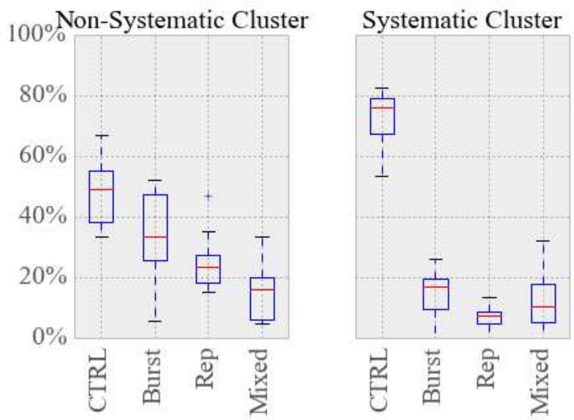

Figure 2. Measures of exploration by cluster.

Table 3: Experimental conditions by cluster of exploration strategy

Condition	Systematic (n = 10)	Non-Systematic (n = 10)
SIM (n = 11)	9	2
LIGHT (n = 9)	1	8

Impact of light as an additional degree of freedom

Given the goal of matching the model to the real path, we expected LIGHT groups that used light sequences closer to the reference light sequence (SIM group) to perform better. Students generated a total of 32 time sequences that *completed* a simulation run, with at least 2 runs per group. We found two clusters, using hierarchical clustering with the centroid linkage method on the correlation distances between the light sequences across all LIGHT groups (avg silhouette score = 0.3). Figure 3 shows the average light sequence of each cluster (dashed lines) with the standard deviation bands. An angle of 0 degrees corresponds to light from the right, an angle of 45 degrees to light from the top-right, and an angle of -45 degrees to light from the bottom-right, etc. The black line shows the reference light sequence of the SIM condition. The *aligned cluster* (18 sequences) contained light sequences that were in average positively correlated with the reference light (r=.3, SD=.4); the *non-aligned cluster* (14 sequences) contained light sequences that were in average negatively correlated with the reference light (r=-.2, SD=.4); the difference in correlations was significant, t(30)=3.8, p<.001, d=1.4.

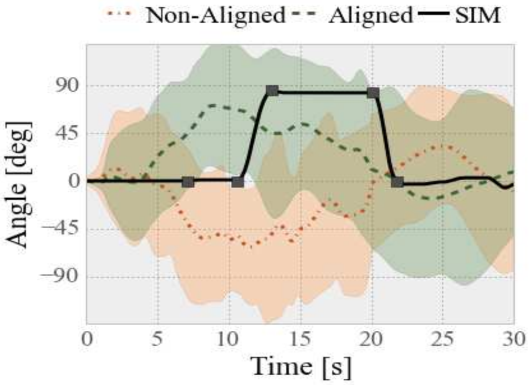

Figure 3. Light sequences by cluster and condition.

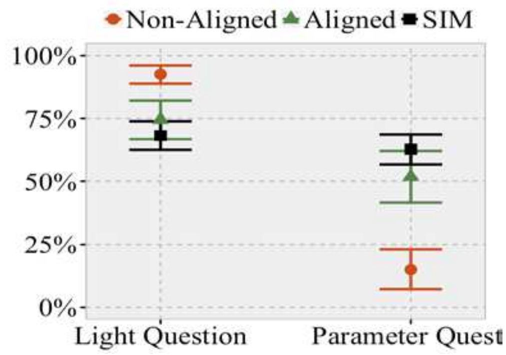

Figure 4. Test scores by cluster and condition.

For each LIGHT group, at least 75% of the sequences belonged to the same cluster. Thus, we assigned the groups themselves to the clusters: 4 out of the 9 groups belonged to the non-aligned cluster. Splitting all student groups into aligned cluster, non-aligned cluster, or SIM condition revealed significant differences on the inference questions, Wilk's λ=0.7, $F_{(2,28)}$=2.9, p=.01 (Figure 4). Post-hoc comparisons between conditions on each question type revealed that the non-aligned cluster performed better on the light question over both the aligned cluster, t(17)=1.9, p=.08, d=0.9, and the SIM condition, t(28)=2.5, p=0.02, d=1.0. They performed worse on the parameter questions, compared to both the aligned cluster, t(12)=-2.2, p=.05, d=-1.2, and the SIM condition, t(20)=-3.5, p=.002, d=-1.8.

Cognitive engagement by condition, and impact of visual feedback

While analysis of the quantitative data indicated that the SIM condition was more systematic in the parameter exploration, and hence performed better on the post-test, analysis of the student conversations paints a different picture of students' engagement in the model exploration: here was no difference between conditions in the frequency of explicit comparisons of model and real Euglena, t(17)=.5, p=.6 (M_{SIM}=1.9, SD=1.5; M_{LIGHT}=2.3, SD=2.1). In both conditions, students hardly compared model and real Euglena. Contrary to what we expected, the LIGHT groups reflected significantly more on the model or parameters (M_{LIGHT}=12.9, SD=5.3) than the SIM groups (M_{SIM}=7.6, SD=2.3), t(17)=2.9, p=.01. In contrast, students in the SIM condition referred significantly more often to the purple path of the real Euglena, or were engaged in conversations about the match of the model with the real path (M_{SIM}= 8.3, SD=7.4) than in the LIGHT condition (M_{LIGHT}=3.0, SD=3.4), W=20.5, p=.048 (Wilcoxon rank sum test due to non-normality of the data).

Discussion and conclusions

We have presented a technology prototype for science inquiry in biology designed to support inquiry-based activities that interweave multiple scientific practices, in line with the bifocal modeling framework. The technology combines a remote biology lab with a model exploration interface that goes beyond physical-only or virtual-only technology approaches. Apart from the feasibility demonstration, a second goal of this paper was to examine potential interplays between the design of technology for science inquiry and how students go about the inquiry activity. We developed two interfaces for model exploration that differed in the degrees of freedom for manipulation and the type of visual feedback provided; Table 4 shows that students engaged in different processes.

Table 4: Summary of results, broken down by measures of learning and inquiry processes

Measures		Results (SIM=model+real; no light control; LIGHT=model-only; light control)
Learning Outcomes	-	SIM better on parameter questions; LIGHT better on light question.
Exploration Strategies	-	SIM more systematic and deliberate in parameter exploration.
Cognitive Engagement	-	SIM more focused on matching the path; LIGHT reflected more on model behavior or parameters.

However, the three-pronged approach for measuring students' learning reveals both strengths and weaknesses in each proposed design when it comes to fostering productive inquiry. Assessment of the relative effectiveness of the interface designs depends on what one considers to be the goals of the activity. We elaborate on this point by means of two pictures that emerge from the analysis:

Picture 1: The SIM condition was more productive for parameter exploration

If one considers the learning goal to be about understanding the model parameters by themselves, the SIM interface design seems to be better suited. Students in the SIM condition explored parameters more systematically; they belonged mostly to the cluster of student groups that manipulated more often only one parameter at a time, did less repetitions and spent more time between manipulations. And we showed that students who were more systematic in the parameter exploration performed better on related questions, which is aligned with literature on inquiry strategies in discovery-based activities (Zimmerman, 2000).

We can only speculate why SIM students explored parameters more systematically, because the study design was confounded at the level of design dimensions. However, we think that difference in visual feedback between the conditions had little to no impact on students' inquiry process, as students in both conditions hardly engaged in explicit comparison of model and real organism. Rather, it seems that the additional light control in the LIGHT condition simply increased the difficulty of the task; LIGHT students had to use the limited amount of time to understand both the light and the model parameters, while the SIM students could focus on only model parameters. Furthermore, the interpretation of parameters hinged on the light sequences students generated; students who generated "good" light sequences performed similarly to the SIM students on the post-test. On the other hand, by keeping the light sequence constant, the SIM interface might have freed up cognitive capacity required to engage in systematic exploration of the parameters.

Picture 2: The SIM condition played the "matching game", and not the "inquiry game"

While the simplified interface of the SIM condition (in terms of reduced degrees of freedom) enabled students to focus more on parameter exploration, they appeared to engage in cognitive processes different from the intended processes of coordinating model with real Euglena. In other words, they played a different *epistemic game* (see Sandoval & Reiser, 2004). Epistemic games are activities that engage people in cognitive and discursive practices involved in making and evaluating knowledge. Students in the SIM condition seemed to play a "matching game", in which they focused mainly on manipulating parameters to get the model Euglena to match the path of the real Euglena, without engaging in discussions *about* the parameters or the model. We knew that the model was too simple to ever perfectly match the real path; we hoped that students might have recognized these limitations and discussed about why that might be the case. However, SIM students continuously tried to optimize the match by doing iterative changes to the parameter values, as exemplified in the excerpt in Table 5. We think that the systematic parameter manipulation was a consequence of students playing the "matching game", rather than a deliberately chosen strategy of inquiry. Thus, what seemed like a productive inquiry behavior based on the interaction and outcome measures alone was less productive in terms of cognitive engagement during the activity.

Table 5: Excerpt of conversation in the SIM condition reflecting the "matching game"

1. Student2: Watch this. It will be perfect. Turn! 2. Student1: Come on, turn. 3. Student2: Yes (Euglena goes down). Yes, that's good, it is kind of a bit far, but that's ok.	4. Student1: Alright, you're good, you're good buddy. 5. Student2: You're good. Now, turn... Aaah... (Turns too early) 6. Student2: Do -35 (parameter value)

The LIGHT condition however was showing more reflective discussions on the parameters and the models, which emerged in their struggles to control the model Euglena through the complex interaction of light sequences and parameter configurations. If one considers the learning goal of our inquiry activity to be about reflecting on models and the interaction of models with the environment (light), the sorts of discussions that emerged among the LIGHT students might have provided a more fruitful ground for subsequent learning.

The results of this study have to be interpreted within its limitations: The group-level sample size was small, and some of the missing data could have introduced potential biases. Furthermore, the lesson still represents a rather simplified version of the full bifocal modeling framework; nevertheless, we were able to go beyond experimentation and evaluation by integrating the computer simulation not as a different version of the real lab with different affordances (see de Jong et al., 2013), but as a means of finding model-based explanations for observations made in the real lab.

Interactive biology laboratories provide opportunities for more integrative approaches to science inquiry; but careful attention needs to be given to how the systems are designed. Using a mixed-methods approach, we found that neither of the modeling interface designs was categorically better; rather, they were conducive to different forms of engagement in the inquiry activity, each offering distinct affordances for learning. However, neither interface design lend itself to the coordination of model and real Euglena, which is ultimately what we want to support. Future research will have to focus on developing designs that are more likely to achieve this goal.

References

Abd-El-Khalick, F., BouJaoude, S., Duschl, R., Lederman, N. G., Mamlok-Naaman, R., Hofstein, A., … Tuan, H. (2004). Inquiry in science education: International perspectives. *Science Education*, *88*(3), 397–419.

Ainsworth, S., & VanLabeke, N. (2004). Multiple forms of dynamic representation. *Learning and Instruction*, *14*(3), 241–255.

Berland, L. K., Schwarz, C. V., Krist, C., Kenyon, L., Lo, A. S., & Reiser, B. J. (2016). Epistemologies in practice: Making scientific practices meaningful for students. *Journal of Research in Science Teaching*, *53*(7), 1082–1112.

Blikstein, P. (2014). Bifocal Modeling: Promoting Authentic Scientific Inquiry Through Exploring and Comparing Real and Ideal Systems Linked in Real-Time. In *Playful User Interfaces* (pp. 317–352). Springer, Singapore.

Bumbacher, E., Salehi, S., Wieman, C., & Blikstein, P. (2017). Tools for Science Inquiry Learning: Tool Affordances, Experimentation Strategies, and Conceptual Understanding. *Journal of Science Education and Technology*, 1–21.

Chinn, C. A., & Malhotra, B. A. (2002). Epistemologically authentic inquiry in schools: A theoretical framework for evaluating inquiry tasks. *Science Education*, *86*(2), 175–218.

de Jong, T., Linn, M. C., & Zacharia, Z. C. (2013). Physical and Virtual Laboratories in Science and Engineering Education. *Science*, *340*(6130), 305–308.

Duschl, R. A., & Grandy, R. E. (2008). *Teaching Scientific Inquiry: Recommendations for research and implementation*. Sense Publishers.

Hodson, D. (1986). Rethinking the role and status of observation in science education. *Journal of Curriculum Studies*, *18*(4), 381–386.

Hossain, Z., Bumbacher, E. W., Chung, A. M., Kim, H., Litton, C., Walter, A. D., … Riedel-Kruse, I. H. (2016). Interactive and scalable biology cloud experimentation for scientific inquiry and education. *Nature Biotechnology*, *34*(12), 1293–1298.

Hossain, Z., Bumbacher, E., Brauneis, A., …, Blikstein, P., Riedel-Kruse, I. (2017) Design Guidelines and Empirical Case Study for Scaling Authentic Inquiry-based Science Learning via Open Online Courses and Interactive Biology Cloud Labs. *International Journal or Artificial Intelligence in Education,* 1-30.

Kim, H., Gerber, L. C., Chiu, D., Lee, S. A., Cira, N. J., Xia, S. Y., & Riedel-Kruse, I. H. (2016). LudusScope: Accessible Interactive Smartphone Microscopy for Life-Science Education. *PLOS ONE*, *11*(10).

Lee, S. A., Bumbacher, E., Chung, A. M., Cira, N., Walker, B., Park, J. Y., … Riedel-Kruse, I. H. (2015). Trap It!: A Playful Human-Biology Interaction for a Museum Installation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (pp. 2593–2602). ACM.

Renken, M. D., & Nunez, N. (2013). Computer simulations and clear observations do not guarantee conceptual understanding. *Learning and Instruction*, *23*, 10–23.

Sandoval, W. A., & Reiser, B. J. (2004). Explanation-driven inquiry: Integrating conceptual and epistemic scaffolds for scientific inquiry. *Science Education*, *88*(3), 345–372.

VanLehn, K. (2013). Model construction as a learning activity: A design space and review. *Interactive Learning Environments*, *21*(4), 371–413.

Wilkerson, M., Shareff, B., Gravel, B., Shaban, Y., & Laina, V. (2017). Exploring Computational Modeling Environments as Tools to Structure Classroom-Level Knowledge Building.

Zacharia, Z. C., & de Jong, T. (2014). The Effects on Students' Conceptual Understanding of Electric Circuits of Introducing Virtual Manipulatives Within a Physical Manipulatives-Oriented Curriculum. *Cognition and Instruction*, *32*(2), 101–158.

Zimmerman, C. (2000). The Development of Scientific Reasoning Skills. *Developmental Review*, *20*(1), 99–149.

Acknowledgements

We thank the teacher who opened her classroom to us. This work was supported by NSF grant #1324753, and the Stanford Graduate Fellowship in Science & Engineering.