High School Students' Parameter Space Navigation and Reasoning during Simulation-Based Experimentation

Hee-Sun Lee, University of California, Santa Cruz, 218 ISB, Physics Department, hlee58@ucsc.edu
Amy Pallant, Robert Tinker and Paul Horwitz, The Concord Consortium,
Email: apallant@concord.org, rtinker@concord.org, phorwitz@concord.org

Abstract: Students formulated their own questions for a virtual spring/mass system and collected and analyzed data in the InquirySpace environment featuring probes, computational models, and data visualization software. We investigated how students navigated and reasoned with the parameter space defined by a set of manipulative variables related to a virtual spring/mass system. We analyzed logging data of 31 high school student groups and a student group's Screencast video and found that (1) students' investigations followed stages: exploration, crude initial investigation, refined investigation, and data analysis, (2) some logging events acted as markers for these stages, (3) students used more extreme parameter values during exploration than collecting data to answer their questions, and (4) students' discourse was mostly centered around their parameter space navigation and analysis.

Introduction

Inquiry-based science learning has been emphasized in recent science education reform documents in the last fifteen years. In the original National Science Education Standards (NRC, 1996), inquiry was stated as "a multifaceted activity that involves making observations; posing questions; examining books and other sources of information to see what is already known; planning investigations; reviewing what is already known in light of experimental evidence; using tools to gather, analyze, and interpret data; proposing answers, explanations, and predictions; and communicating the results" (p.23). Since then, what constitutes inquiry-based learning has gone through several revisions, but what remains consistent is to promote student-initiated inquiry, the type of inquiry that seeks students' active participation, contribution, reflection, and learning.

At the heart of empirical inquiry is experimentation where evidence is generated for scientists to answer their questions, connect to theory, elucidate or hypothesize mechanisms behind phenomena, and develop arguments. Scientific experiments have traditionally involved physical apparatus. However, the advancement in computer technologies has provided scientists with an additional tool for exploring complex scientific topics. In the literature, students' experimentations with both physical apparatus and simulations have been investigated. With physical experimentation, research has focused on whether, how, or to what extent students can design and conduct experiments (Hackling & Garnett, 1992; Kanari & Millar, 2004). Students' experimentation skills include recognizing multi-covariate relationships (Amsel & Brock, 1996), dealing with experimental errors (Allchin, 2012), addressing variability in the data (Masnick, Klahr, & Morris, 2007; Petrisino, Lehrer, & Schauble, 2003), applying statistical reasoning (Lubben et al., 2001), treating anomalous data (Chinn & Brewer, 1993), and revising hypotheses, experiments, and questions after reflecting on evidence (Schauble, 1996). Studies have found that students have difficulties in recognizing, identifying, and controlling variables (Toth et al., 2000). When studying students engaged in simulation-based experiments, McElhaney and Linn (2011) found that students' experimentation patterns can be characterized as intentional, random, and exhaustive based on the number of trials attempted by students with or without experimental coherence and found that these patterns resulted in what parameter values students explored during their simulations.

In this paper, we argue that the type of reasoning that enables students to conduct an empirical inquiry is much broader than student reasoning investigated in any of these studies such as controlling variables or identifying outliers in the data. The purpose of this paper is to characterize student reasoning in a much broader sense to capture students' planning, operationalizing, navigating, and reflecting on multiple experimental runs to generate evidence to answer their questions. Since each experimental run can be summarized based on a set of parameters, we name this type of student reasoning *parameter space reasoning* (PSR). In this paper, we focus on PSR involved in simulation-based experimentations. Research questions are (1) how did students navigate and analyze the parameter space in their experimentation with a simulated spring/mass system?; and (2) what aspects of parameter space reasoning were demonstrated in the discourse of one typical student group through the course of a simulated experimentation?

Theoretical Framework: Parameter Space Reasoning

A parameter is referred to herein as a measurable factor that defines a system or determines the conditions of a system's function. In science, a parameter space is defined as all possible combinations of values related to a set of parameters that define a system. A different set of parameters is used for a different system or a phenomenon under investigation. Parameter space also depends on the conceptualization of an experimenter or a modeler who studies a system. An experimental run can be described as a point in an *n*-dimensional parameter space where the number of parameters related to a system is *n*. Some parameters are salient to determine a given outcome while others are not. The range of a parameter can be plotted on one axis (typically, x) and the outcome of a system can be plotted on another axis (y, for example). In a spring-mass system, an outcome variable of interest such as period can be estimated from a sinusoidal graph between the distance from the center and time, instead of directly measuring it. We call a graph that shows changes in a variable over time as a time-series graph. If all other remaining parameters are kept constant, then a point representing (parameter space value, outcome value) can effectively summarize the result of an experimental run. We call this plot a parameter-outcome graph. It is quite possible that experimenting with different regions of the parameter space yields different results, e.g., a spring too stretched to lose its elasticity. Therefore, it is important to explore the adequate range of the parameter space to test a model or a relationship in experimentation.

An investigation includes multiple experimental runs, each of which is defined uniquely by a set of values chosen for the run (as one run sets a single value for each parameter, two values of the same parameter cannot be investigated simultaneously in a single run). The region of parameter space examined by students can be traced. Before experimental runs, students need to plan for which variables to vary and how to vary them. At the end of each experimental run, students must make decisions on their next step such as redoing the same experiment, varying parameter values, checking their equipment, and eventually concluding their investigation. After all the data are collected, students need to think about the quality of the data and recognize relationships between the parameters they manipulated and system outcomes they measured. PSR captures this array of cognitive processes as they relate to empirical inquiry, and entails, among other things, the ability to compare an experimental run to other runs that differ in the value of at least one parameter. Table 1 lists PSR in three phases of an investigation: parameter setting, data collection, and data analysis and describes the reasoning in each investigation step and how PSR can be observed.

Table 1: Characterization of parameter space reasoning (PSR) during student investigation.

Investigation phases	Investigation steps	PSR occurs as students:
Parameter setting	• Formulate a question with parameters and outcome variable • Identify parameters and outcome variables for an experimental setup • Know how to vary parameters and measure outcome variables	• Set parameters for an investigation based on the question and the setup • Conduct test runs to build a mental model between the phenomenon under investigation and the data to be acquired • Describe which variables will be varied and how
Data collection	• Select a parameter set, carry out a run and measure the outcome variable • Reflect on the quality of the run. • Determine whether to rerun or stop data collection	• Make multiple runs purposefully to answer the question • Determine when to rerun, modify a run, or stop data collection • Select data for analysis
Data analysis	• Calculate a way to characterize a run with a single value, in order to compare runs • Create a parameter-outcome graph • Use a time series graph to obtain an outcome value • Identify patterns in a parameter-outcome graph • Reflect on quality of data • Answer the question using evidence generated from the investigation	• Calculate and incorporate outcome into analysis • Create and explain parameter-outcome graphs • Explain a point in the parameter space in connection to a time series graph of a run • Recognize the shape and important features parameter-outcome graphs (linear, nonlinear, periodic, etc. or break points where the nature of shape changes) • Identify and treat outliers • Identify and treat noise in the data and noise sources • Communicate conclusions using evidence

Methodology

InquirySpace (IS) Learning Environment

The IS environment works with both physical and simulated systems. Figure 1 shows the IS environment for a simulated spring/mass system. Students can conduct multiple experimental simulation runs by varying parameters such as gravity, spring constant, starting position, mass of the ball, and damping. When students finish a simulation run, they can view their data in a table and export the data for analysis after clicking the "Analyze" button. For instance, Figure 1 shows that students conducted four simulation runs by varying the spring constant parameter. A column for the period of oscillation of the system was added in the table in Figure 1, and students inserted period values estimated from the time series graph (bottom left). In Figure 1, students created two graphs: (1) time vs. distance (i.e. time series graph) shown at the bottom left of the screenshot and (2) spring constant vs. period (i.e., parameter-outcome graph) shown at the bottom right.

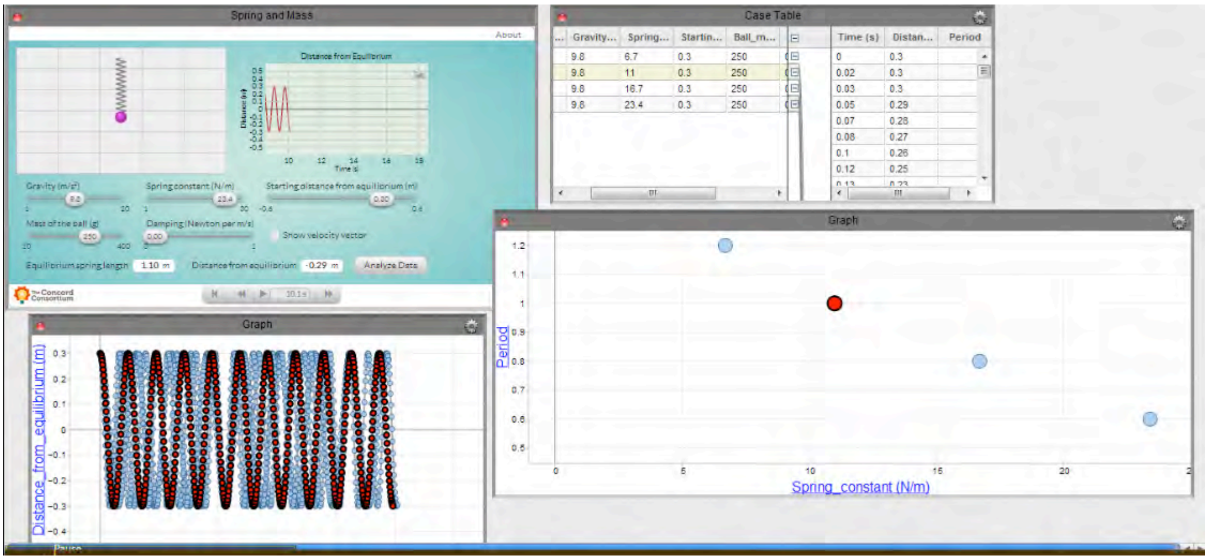

Figure 1. InquirySpace learning environment.

Data Collection and Analysis

The simulation-based spring/mass experimentation took place over two class periods in four ninth grade physics classrooms in an urban high school. Eighty-two students worked in 32 small groups. Each group chose its own question. 92% of the students spoke English as a first language, 52% were female; 20% self-reported to have used computers regularly for school learning. The school is a public charter high school where 98% of the students are from minority populations and 77% receive reduced or free lunch. According to the teacher, the IS curriculum sequence provided the opportunity for students to work in computerized lab experiments for the first time. The teacher was very structured and organized and made class performance expectations clear at the beginning of each class and was implementing the IS curriculum for the first time.

The IS curriculum sequence consisted of three investigations. First, all groups worked on the same hands-on investigation, using probes connected to the data collection and graphing environment, to answer how the period was affected by the mass of the ball. In the next investigation, groups were encouraged to explore how other variables in the physical spring/mass system impact period. The third investigation was conducted in the simulation-based spring/mass system. In this paper, we focus on the third investigation where groups had more choices for independent variables such as spring constant and gravity than were available to them with the physical spring/mass system. Each investigation was guided by accompanying worksheets and was carried out in the order of Explore, Plan, Create a Screencast Video about Plan, Experiment, Analyze, Explain, and Summarize conclusions on a Screencast Video. In this study, we focus on logging events recorded in the data analytics component of the IS environment and Screencast Videos that summarized conclusions of the third investigation. For the exploration stage, the worksheet said "Play with the model until you see what the model does to the spring." Throughout the worksheets, it was clear to students that they were investigating the impact of one variable on another and they were encouraged to go back to collect data if they were confused. The teacher told students to collect data from at least four simulation runs.

We used logged events to track students' parameter space navigation. A total of 5,277 events were logged for 31 student groups during the simulation-based mass/spring experiment. One group's logging data were lost. Fifty-three different syntaxes were used for logging events. Among the logged events, 62.6% were associated with simulations such as starting models and exporting models to the data analysis interface; 16.0% were related to data analysis involving tables, graphs, and data points; 17.9% involved creating and deleting components such as table, graph, and model; 2.8% were related to logging in and out. Each logged event was time-tagged, allowing duration and temporal order analyses. In order to investigate how students explore and analyze the parameter space to answer their investigative questions, we used the logging data to develop an event map for each student group over the period of the group's investigation and plotted key logging events such as beginning of exporting model results, creating period in a table, creating a time-series graph to estimate the period, and creating a parameter-outcome graph that students needed to use to answer their questions. These key logging events signal overall progression of their experimentation from exploration to refined experimentation to data analysis. We also counted the number of simulation runs that were exported vs. not exported and compared students' parameter space navigation patterns between the two. Students' screencasts made at the end of the investigation were used to determine which independent and dependent variables students chose to investigate and how many and what values were chosen for the parameter-outcome graph. In order to determine whether PSR occurred as conceptualized in Table 1, we transcribed a 45 minute video of a student group's investigation and selected all segments where students reasoned for, within, and about the parameter space defined by their experimentation.

Findings

Student groups selected their own independent and dependent variables for their investigations. For the independent variable, 18 groups selected spring constant, 11 groups selected gravity, one group (T5) used mass, and one group (T9) used damping. All but one group selected period as a dependent variable. T9's investigation of the relationship between damping and the amount of decrease in amplitude was unique among all investigations. Students' investigations were carried out in the order of exploration, crude initial data collection, refined data collection, and data analysis. Identification of these stages in the logging data was facilitated by the presence of key logging events that signaled changes in students' experimentation focus. For example, students did not export their simulation results until they were serious about analyzing the data. Thus, *exported model* became an important logging event for students' moving from the data exploration stage to the data collection stage. Another important logging event was *created attribute period* because period was an outcome variable and needed to be estimated from the time-series graph. This event signaled students' moving from the crude initial data collection stage to the refined data collection stage. The syntax, *changed plot horizontal/vertical axis [variable name]* indicated that students were creating either a time series graph or a parameter-outcome graph. If students changed the horizontal plot axis with time and the vertical plot axis with distance from the starting position, they were making a time series graph to make measurements on period, the outcome variable for most student groups. If they assigned the horizontal axis to their independent variable and the vertical axis to their dependent variable, they were creating a parameter-outcome graph that was necessary evidence for their conclusion.

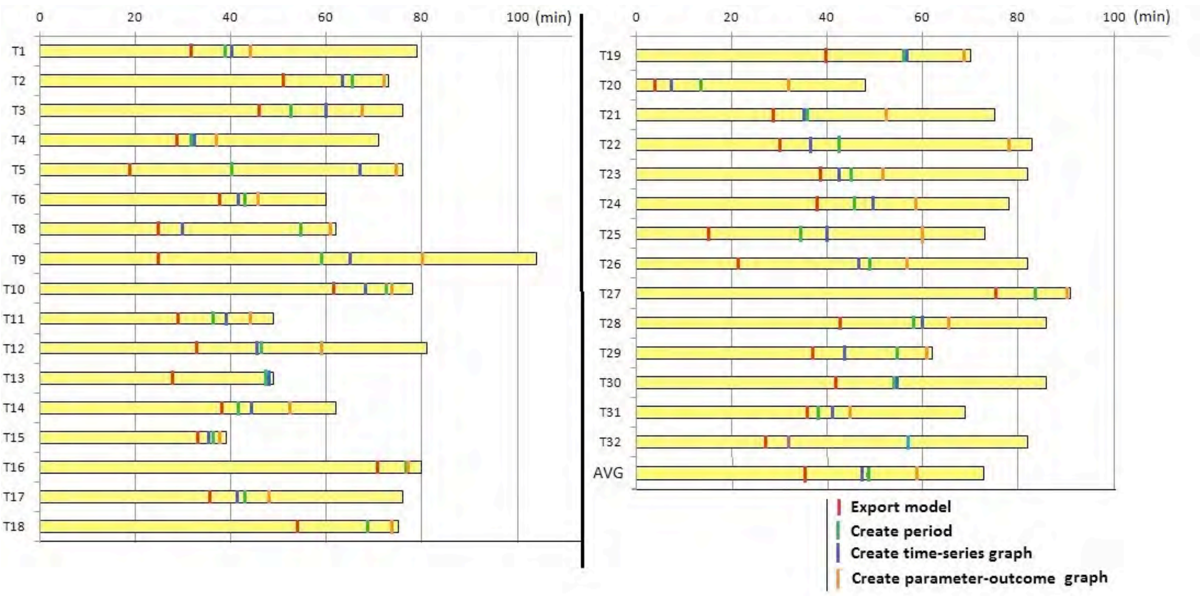

Figure 2. Event map for experimentation

As shown in Figure 2, on average, students' investigation lasted 72.5 minutes, ranging from 39 to 104 minutes. The four main logging events are marked in Figure 2. Creating a time-series graph ensued after several data runs were exported at the average of 47.3 minute mark and immediately followed by creating a column for period at the 48.6 minute mark. Some groups created the period column in the table before they created the time-series graph while other groups did it in the opposite order. All but one group (T17) were able to create a time series graph. On average, groups created their parameter-outcome graphs towards the end of their investigations at 59.8 minutes. Three groups (T13, T30, and T32) failed to create parameter-outcome graphs. Among the 28 groups who created the parameter-outcome graphs, two groups created incorrect graphs. For example, T5 created a graph of gravity vs. period, instead of mass vs. period, showing gravity was the same for all four simulation runs even though the group wanted to investigate the relationship between mass and period. T27 created a graph of gravity vs. elapsed time, rather than gravity vs. period. Nine groups were able to plot parameter-outcome graphs on their first attempt while others needed to try two or more times to put the correct independent variable on the x-axis and the correct dependent variable on the y-axis. Groups that created the parameter-outcome graphs earlier used the remaining time to refine the graphs such as connecting points or making sense of what the graphs represented.

When we examined parameter values students chose to investigate, we discovered that students did not appear to export their data for analysis in the increasing or decreasing order. Rather, students were more concerned about having a set of parameter values that could cover a fairly good range of their chosen independent variable. There were several noticeable differences between exploration and formal data collection stages. According to Table 2, the average number of simulation runs student did not export was 45.5 while that of simulation runs students exported was 7.3. This indicates that students' exploration covered more of the parameter space than they actually analyzed. Moreover, in this exploration stage, students tested all available independent variables before settling into one independent variable of their choice. This indicates students' interest and need for tinkering with the simulation model as recognized in other studies (Berland, Martin, Benton, Smith, & Davis, 2013). All student groups plotted four points in their parameter-outcome graphs as suggested by their teacher. However, many groups exported more than 4 simulation run results for data analysis. Our further inspection of this discrepancy indicates that (1) some groups had to repeat their runs the following day for analysis because they did not save the data, (2) some groups ran simulation runs with the same set of parameters multiple times and exported them, or (3) other groups refined values for intervals between parameter values to be even. Table 2 lists five parameters with ranges that were manipulated by students. We considered the bottom and the top 10% as extreme value ranges. We obtained percent frequencies for these parameters that were set in that extreme ranges for exported vs. not-exported simulation runs. When students were just exploring they used more extreme values than when they were collecting data to answer their questions. One exception to this rule was damping. Students set the damping parameter "0" for most of their simulation runs for data collection because, even though damping does not affect the period, it does decrease the amplitude over time, making estimating the period from the time-series graph more difficult.

Table 2. Navigated parameter space.

Parameters	Value range	Exported parameter sets (n = 235)		Un-exported parameter sets (n = 1,457)	
		Middle (%)	Extreme (%)	Middle (%)	Extreme (%)
Gravity (m/s^2)	0.8 - 19.8	77.5	22.5	40.0	60.0
Spring constant (N/m)	1.0 - 30.0	76.6	23.4	45.1	54.9
Amplitude (m)	-0.6 - 0.6	91.5	8.5	57.9	42.1
Mass (g)	10 - 400	82.1	17.9	48.3	51.7
Damping (N per m/s)	0.0 - 1.0	3.8	96.2	14.3	85.7

Example: A Group's Parameter Space Navigation

In this section, we describe the types of PSR that occurred when a group of students (Group T1) explored the parameter space. Group T1 consisted of one male and two female students. Figure 3 shows a time-lapsed event map for the first 45 minutes of their 79-minute long investigation. The curriculum investigation sequence consisted of exploration, planning, creating a Screencast plan video, data collection, and data analysis. T1 chose a question of how spring constant affected period and was able to complete four main key logging events during this time period. T1 conducted 25 simulation runs for exploration and 10 runs for data collection. They chose the spring constant values of 8, 10, 12, and 16 N/m. From the video transcripts, we identified six occasions where students were actively engaged in PSR as envisioned in Table 1. While not all features of PSR listed in Table 1 occurred, students' discourse shows that PSR played a central role in various stages of their investigation.

Making a Hypothesis

An important aspect of PSR is to setup a parameter and what it means in order to develop a hypothesis. After exploration, students wanted to come up with a hypothesis. In formulating a hypothesis involving a spring constant, students were confused as to what a spring constant meant. The excerpts below indicate students' conceptual clarification of the spring constant variable and how it would affect the period.

S3: We are dong spring constant, right guys?
S2: Less mass it is, the more the spring to be constant?
S1: What?
S4: I don't know.
S1: so, the constant, he means this right?

S3: Is that constant?
[Unsure what spring constant is, students attempting to get the teacher's attention]
S2: We thought the greater the spring constant the faster it would go.
Teacher: So, the higher the spring constant, and what?
S2: the time…
Teacher: What about the time? longer time or shorter time?
S2: I don't get it.
Teacher: Do you know what spring constant is?
S2: No.
Teacher: Do you know something's really stiff.
S2: Right!
Teacher: The spring constant is a measure of the stiffness ...
S2: So, if the spring constant is high, then the faster the period to finish?
Teacher: Maybe...that's the hypothesis you need to figure out.
S1: You got the new hypothesis?
S2: The higher the spring constant, the faster the period.

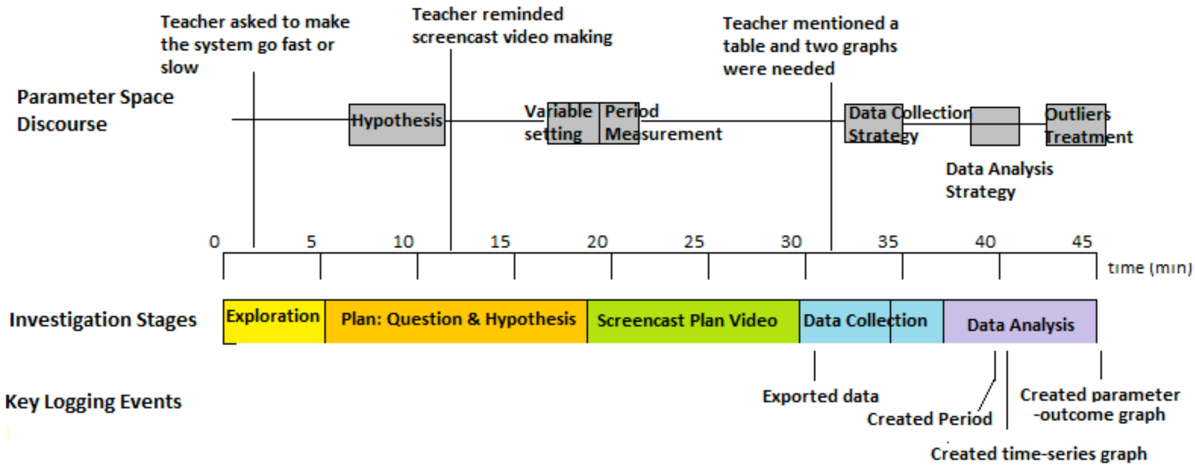

Figure 3. A student group's detailed event map

Setting Variables

About 12 minutes into the class, the teacher reminded students to make a Screencast video communicating their experimental plan. The teacher particularly asked students to focus on choosing a parameter to vary at least four times. This teacher's request had students think about how to manipulate parameter space using the simulation.

Teacher: All you need to do is what variable to vary and what are the four numbers you are going to change variables to, to test it.
S2: How am I going to change numbers?
Teacher: Move the bar before you chose random numbers....like evenly spaced out
S1: Which bar?
Teacher: What are you doing? Gravity or spring constant?
S2: spring constant.
Teacher: spring constant bar....choose four numbers.
S2: don't do odd numbers, do even numbers that are far apart.
[T1 chose 8, 10, 12, and 16 N/m for their experiments.]

Measuring Periods

Now, T1 had a hypothesis of "the higher the spring constant, the lower the period" and chose four spring constant values to test. As they started recording a Screencast video about their plan, they recognized that they did not know how to measure the period:

S2: Our question is how does spring constant affect the period. The two variables that we are using today are spring constant and period.
S1: OK, how do we measure the period?
[At first, confused...asking around...from distance "it's the difference between crests"...].
Teacher: Remember when you drew a graph...
S1: OK, then we need to find out the difference?

S2: OK, we measure the spring constant and period. We do period by subtracting the distance between the two times, but we haven't started yet so we don't have it.

Data Collection Strategies

Immediately after their recording of the Screencast video, T1 started collecting data by changing spring constants from 8, 10, 12, to 16 N/m. Then, the teacher mentioned that a table and two graphs (time series and parameter-outcome graphs) were necessary. Then, T1 realized that they had not been saving data for analysis:

Teacher: You need to create a chart and two graphs.
S2: Oh.
Teacher: I see nothing related to charts. You have to analyze data.
S2: Do we need to start all over?
Teacher: yes. Do you know what we are saying? You need to push that button to send to the chart.
S2: Then, we got to do one by one again, then.

Data Analysis Strategies

T1 had to redo the same simulation runs so that they could save their data for analysis. This time, after each time they varied the spring constant value, they exported the data to the chart. The four simulation runs went fast. T1 then wanted to create a graph and first went for a parameter-space graph. After setting the x-axis as spring constant, they realized that they did not have period in their data. This realization allowed T1 to work on obtaining the dependent variable from time-series graphs.

S1: I need to get period. Where is the other graph [*time-series graph*] I was looking for.
[S1 deleted spring constant x-axis and changed it to time series graph for the spring constant 8 one]
S1: Is time x or y?
S2: Time? x.
[S1 then put the distance from the equilibrium variable on the y-axis to create the time series graph where students could obtain period from subtracting time difference between two crests]

Detecting and Treating Outliers

For the period of the spring constant of 8 N/m, they calculated period as 1.1 seconds. For the period of the spring constant of 10, 12, and 16 N/m, they calculated as 9.9, 9.0, and 7.9 seconds. The first period estimate was a correct one. However, students made mistakes for the other three. A student surmised that the next three periods were too big because she thought, according to their hypothesis, that the periods should be getting smaller. However, her observation went unnoticed. This subtraction mistake was caught when the group plotted their first parameter-outcome graph and immediately recognized something was not right:

S1: Does it supposed to look like that?
[students were examining the first point because it looks an outlier from the time-series graph. ...then went to the second point.]
S2: This is .9
[after examining the corresponding points in the time-series graph]
S1: That's not 9.9. It is .99
[They checked all the other points so that their final periods were 1.1, .99, .90, .79 for spring constants of 8, 10, 12, and 16 N/m. This gave a reasonable spring constant vs. period graph]

Conclusion and Significance and Connection to the Conference Theme

The time progression of students' investigations generally followed directions on the accompanying worksheets. This means that the suggested curriculum sequence was able to accommodate students' needs in simulation-based experimentation. However, we must emphasize that the teacher played an important role in moving and refocusing students' attention on important parts of their investigations, and clarifying questions and confusions students might have had. Students' parameter-value choices were made purposefully. For the exploration, students used extreme values across all available parameters to quickly grasp the general tendency of how each parameter affected the virtual spring/mass system. For the data collection purpose, prompted by their teacher, they chose a parameter to investigate and selected the range of values that could roughly determine the overall shape of the relationship between the parameter and the system outcome variable. Conversations within a student group centered on various aspects of parameter space reasoning: selection, definition, and measurement of independent and dependent variables, data collection and analysis strategies, and outlier treatment. In these efforts, visualizations of the raw or analyzed data appeared to be critical in prompting students' immediate responses, such as forming a hypothesis, modifying period estimates, and recognizing the next actions to take.

As shown in this study, during experimentation, students reasoned exclusively with the parameter space, rather than with mechanics of the spring and mass system. Our currently ongoing research efforts are focused on developing logging data analytics for teachers and researchers to use, defining converging evidence of learning that is unique to student-generated experimentation from multiple sources, and defining and validating PSR.

Inquiry-based investigations in science class have been promoted as an important pedagogical approach in science education reforms. We characterized and illustrated student reasoning associated with the parameter space that defines individual experimentation runs in the context of simulations. However, PSR —the ability to think about the simulation runs as simply data points at a higher–level—can get confounded with other "inquiry skills" such as the ability to ask "interesting" questions, the ability to come up with experimental designs likely to shed light on a question, the ability to control potentially confounding variables, and the ability to reliably distinguish signal from noise. The science education literature refers to "systematicity" (often mindlessly recommending the "one size fits all" strategy of varying only one parameter at a time) or suggests as a normative standard that students exhaustively cover every region in the parameter space no matter what question they are trying to answer. Such supposed universal markers for PSR are clearly too simplistic to capture the richness of the phenomena under study. As more and more learning technologies are integrated to form powerful learning environments, it becomes necessary to reconceptualize student reasoning in a more nuanced and multi-faceted manner. We believe that a learning environment such as InquirySpace can provide opportunities for students to be *learning and thus becoming in practice* while also enabling designers and researchers to study resulting student learning to a greater extent. In this study, we outlined what PSR might mean in a simulation-based learning environment and illustrated that students were indeed engaged in PSR frequently throughout planning, exploration, data collection, and analysis stages.

References

Allchin, D. (2012). Teaching the nature of science through scientific errors. *Science Education, 96*(5), 904-926.

Amsel, E., & Brock, S. (1996). The development of evidence evaluation skills. *Cognitive Development, 11*(523-550).

Berland, M., Martin, T., Benton, T., Smith, C. P., & Davis, D. (2013). Using learning analytics to understand the learning pathways of novice programmers. *The Journal of the Learning Sciences*, 22, 564-599.

Chinn, C. A., & Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research, 63*(1), 1-49.

Hackling, M. W., & Garnett, P. J. (1992). Expert-novice differences in science investigation skills. *Research in Science Education, 22*, 170-177.

Kanari, Z., & Millar, R. (2004). Reasoning from data: How students collect and interpret data in science investigations. *Journal of Research in Science Teaching, 41*(7), 748-769.

Lubben, F., Campbell, B., Buffler, A., & Allie, S. (2001). Point and set reasoning in practical science measurement by entering university freshman. *Science Education, 85*(4), 311-327.

Masnick, A. M., & Klahr, D. (2003). Error matters: An initial exploration of elementary school children's understanding of experimental error. *Journal of Cognition and Development, 4*(1), 67=98.

Masnick, A. M., Klahr, D., & Morris, B. J. (2007). Separating signal from noise: Children's understanding of error and variability in experimental outcomes. In M. C. Lovett & P. Shah (Eds.), *Thinking with data* (pp. 3-26). New York: Lawrence Erlbaum Associates.

McElhaney, K. W., & Linn, M. C. (2011). Investigations of a complex, realistic task: Intentional, unsystematic, and exhaustive experimenters. *Journal of Research in Science Teaching, 48*(7), 745-770.

National Research Council (1996). *National science education standards*. Washington, DC: National Academy Press.

Petrisino, A., Lehrer, R., & Schauble, L. (2003). Structuring error and experimental variation as distribution in the fourth grade. *Mathematical Thinking and Learning, 5*, 131-156.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32*, 102-119.

Toth, E. E., Klahr, D., & Chen, Z. (2000). Bridging research and practice: A cognitively based classroom intervention for teaching experimentation skills to elementary school children. *Cognition and Instruction, 18*(4), 423-459

Acknowledgments

This material is based upon work supported by the National Science Foundation under grant No.1147621. Any opinions, findings, conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The authors acknowledge Edmund Hazzard who developed IS activities, William Finzer and Daniel Damelin who developed the simulation and logging technology used in this study, and students and teachers who participated in this study.