Mining students' actions for understanding of complex systems: Students' explorations of gas models in the Connected Chemistry curriculum

ABSTRACT

We investigate students' inquiry actions in computer-based models of complex systems, studying whether and how they adapt to different mathematical regularities in the system; examining how these explorations may relate to prior knowledge and learning. Students' data-collection choices while exploring models were data-mined and analyzed showing: In most cases, students conduct mathematically-astute explorations, consistently adapting their strategies to the model's mathematical structure; Mathematically-fit explorations are associated with deep conceptual knowledge, specifically understanding of the system as complex; Fit explorations are somewhat associated with learning along complementary dimensions: quantitative problem-solving and bridging micro-to-macro-levels in the system. These results are discussed with respect to learning about complexity through exploring models and the importance of such conceptual understanding even for quantitative problem-solving.

INTRODUCTION

Can we learn from students' inquiry actions in a computer-based learning environment about their understanding of the domain under inquiry? As the methods for data-mining of students' actions in digital environments advance, new expressions of knowing become accessible. Detailed observation of each student's interactions with learning materials is beyond a single teacher's capabilities. However, with the assistance of logging and data mining, such observation and subsequent support for learning are made possible.

This investigation focuses on how students explore computer models of complex systems and whether such explorations are indicative of understanding and learning. We examine data that students collect through experimentation with the models when their goal is to further use this data to construct equations. The spacing of this data is investigated as an indicator of their understanding of the model's underlying mathematical behavior. Usually, work in school science structures a linear addition to the *independent* variable (100-150-20-250…). Yet, when the inspected system behaves in a nonlinear way, varying the *dependent* variable at constant intervals would better capture the system's full range of change. In the study, we look into students' data collection strategies, describe the patterns to which they conform, their fit to the model's mathematical behavior, and relate them to the students' prior knowledge and subsequent learning.

The activities involve students' construction of the gas law equations in high-school chemistry class, as they engage with the Connected Chemistry curriculum (***, 2009a), part of the Modeling Across the Curriculum project (***, 2003). Connected Chemistry (CC1) is a computer-based environment for learning the topics of (macro) gas laws and (micro) kinetic molecular theory (KMT) in chemistry through a complexity perspective. CC1 employs NetLogo (***, 1999a) agent-based models that compute a system's behavior from its components' actions and interactions. These models are embedded in Pedagogica scripts (Horwitz & Christie, 1999) that provide several forms of guidance, assistance, and assessment, while logging students' actions and responses to questions.

Students' understanding is described in terms of conceptual/qualitative and "algorithmic" or well-practiced quantitative problem-solving types of knowledge. Several researchers demonstrate how students may be capable of solving problems that involve using equations to predict the properties of gases under a variety of conditions; nevertheless their conceptual understanding lags far behind (Niaz & Robinson, 1992; Nakhleh, 1993; Russell et al., 1997; however, see Chiu, 2001; Costu, 2007). Moreover, such understandings may be limited when the problems do not fall into familiar and practiced problems (Lin & Cheng, 2000).

Students' learning is examined through a complexity perspective: understanding the micro and macro levels, and relating the two; shifting between equation-based and macroscopic descriptions. Complex systems are made up of many elements, which interact among themselves and with their environment, resulting in the system's coherent self-organized behavior (Holland, 1995; Kauffman, 1995). NetLogo is a programming language that supports creating agent-based models, such as those used in CC1. Exploring such models of chemical systems, that integrate multiple representations (visual representations of both the micro and macro levels and symbolic representation of its properties) have been shown to be effective in helping students gain a deeper understanding (Ardac & Akaygun, 2004; Kozma, 2000; Russell, et al., 1997; Snir et al., 2003; van der Meij & de Jong, 2006).

RESEARCH QUESTIONS

1. What patterns describe how students explore computer models when they are engaged in experimentation for the purpose of determining a system regularity--constructing an equation relating the system's variables?
2. How does prior knowledge impact students' model exploration patterns?
3. What associations can be found between students' exploration patterns and their learning gains regarding the related content?

METHOD

Procedure

The students engaged with seven activities of CC1 as part of their high school chemistry course during the 2005-6 year, replacing the topic of gas laws and KMT. Before and after the activities, spaced 2-3 weeks apart, the students completed content knowledge questionnaires. The students' interactions with the computerized environment, answers to open and closed questions and manipulations of the models were logged through the Pedagogica environment.

We focus on sections from three of the later activities that involve mathematical modeling. In these sections, the students constructed the gas laws based on scatter-plots of data they collected from models in which they could manipulate one variable: the number of particles (N), temperature (T) or volume (V) and observe the resultant pressure (P) (Figure 1). They explored the models prior to the activity, and following it, tested their equations with the models.

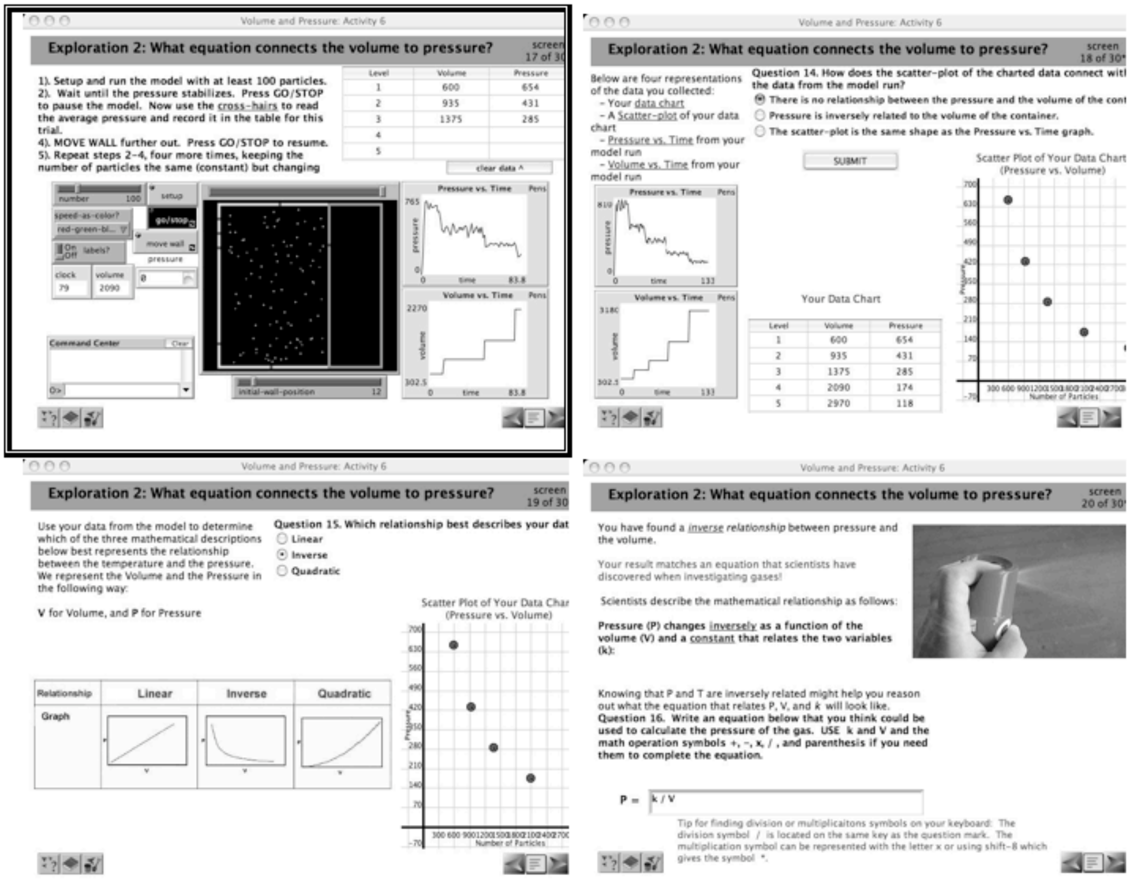


Figure 1: A sequence of four screenshots from the "Volume and Pressure" activity, portraying construction of Boyle's Law through experimenting with the agent-based model. The focus screen in the study is highlighted.

<u>Participants</u>

The sample included 81-368 students (depending on the portion of the data analyzed) who participated in both pre-test and post-test, engaged with the activities, and their logs were successfully captured. Of these, 49% were male students and 51% female students; 13% in 9th grade, 22% in 10th grade, 61% in 11th grade; 4% in 12th grade. 41.3% were in a regular class, 30.4% in an honors class, 17.2% in a Pre-AP class, 7.6% in a college-level class and 3.5% were unspecified. These students come from 12 diverse high schools across the US.

<u>Instruments</u>

In this study, two main instruments have been used: a pre-test/post-test questionnaire and logs of the students' actions in exploring the models. The pre- and post-test content knowledge questionnaire assessed students' understanding of the gas laws and KMT. It is fully described elsewhere (***, 2009b).

<u>Data Analysis</u>

The content knowledge pre- and post-test questionnaires' responses were coded as correct or incorrect and a total score was averaged. Two sets of subscales were created: one distinguishing between conceptual/qualitative-algorithmic/quantitative questions; the other, more fine-tuned, focused on complex systems and mathematical modeling (micro--macro--mathematical--micro/macro--model/math).

Students' data collection was analyzed in the following way. The series of five values they entered into the table for the manipulated variable (N for the N-P relationship; T for the T-P relationship; V for the V-P relationship) was converted into patterns (Figure 2): The first and second derivatives of these sequences were calculated. The second derivative was then recoded as its sign: positive (indicating increasing intervals), negative (indicating decreasing intervals) or zero (constant intervals). The 27 combinations of the three signs of the second derivative were sorted into four categories: mainly constant intervals (at least 2/3 constant additions), mainly increasing intervals (at least 2/3 increasing additions), mainly decreasing intervals (at least 2/3 decreasing additions) and mixed (otherwise).

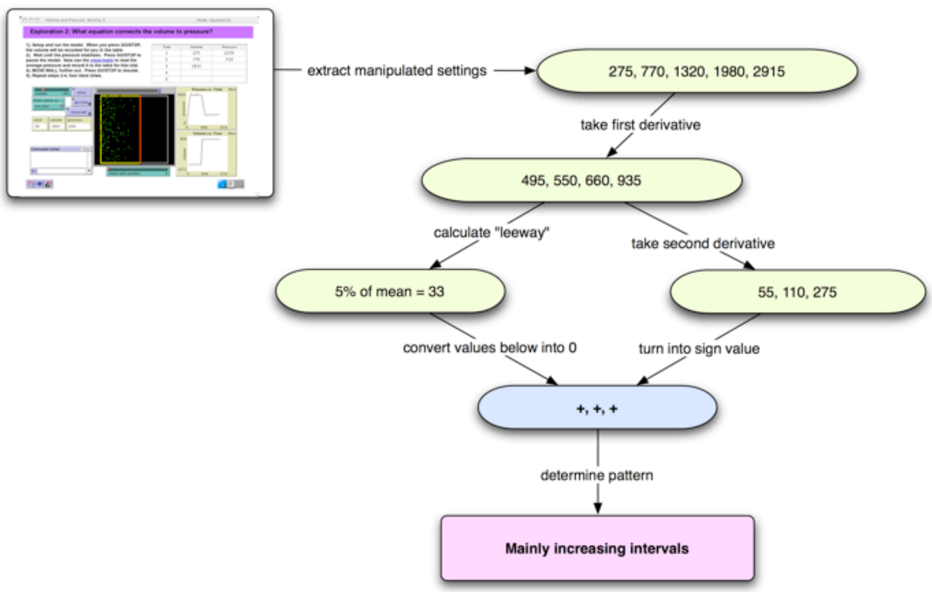
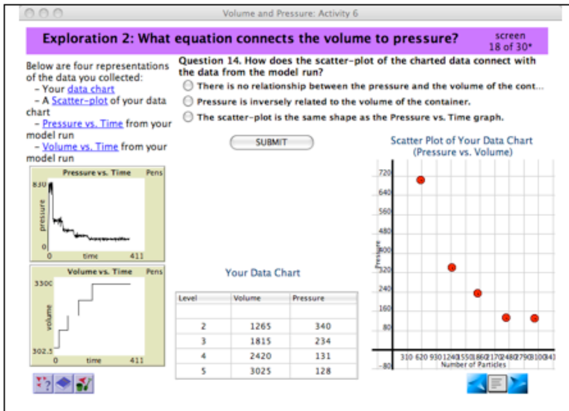
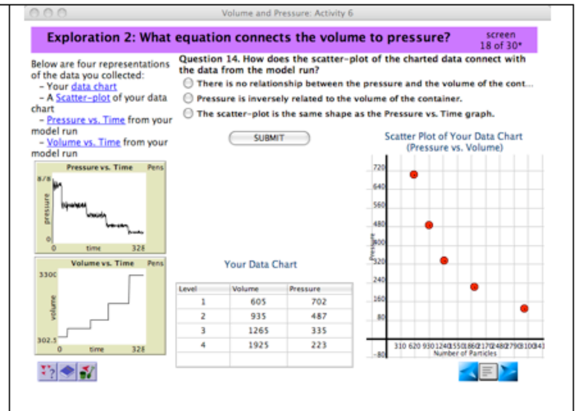


Figure 2: Example describing method of extracting students' exploration patterns from their table data entries.

These patterns were further coded as "fit" or "unfit" with respect to the model's behavior. Regarding the *linear* N-P and T-P relationship, a fit exploration strategy is "constant intervals" as it covers the parameter space systematically; for the *inverse* V-P relationship, it is "increasing intervals" as it captures the faster change in pressure for lower volumes. Figure 3 demonstrates how "increasing intervals" for an inverse relationship captures a larger range of change than "constant intervals". These patterns are described, related to prior knowledge to learning gains. Learning gains are calculated as (posttest-pretest)/pretest.



"Constant intervals" sampling while exploring the inverse pressure-volume relationship.



"Increasing intervals" sampling while exploring the inverse pressure-volume relationship.

Figure 3: Demonstration of the greater coverage of the parameter space afforded by exploring an inverse function with "increasing" rather than "constant" intervals.

FINDINGS

Model exploration patterns

Students' explorations of the gas models are described in Figure 3. Distinct distributions are observed for the different models. For the NP model, a clear mode is seen - the "constant intervals" exploration pattern (66%). For the TP model, a rather flat distribution is observed, the mode at "increasing intervals" (34%). For the VP model, the mode is the "increasing intervals" pattern (47%), however the "decreasing intervals" pattern soon follows. When the patterns are recoded for fitness, we can see that the students are using mainly fit patterns for the NP and VP exploration, however no clear result is seen for TP. The number of fit explorations is $M = 1.41$, $SD = .851$, about half of the explorations.

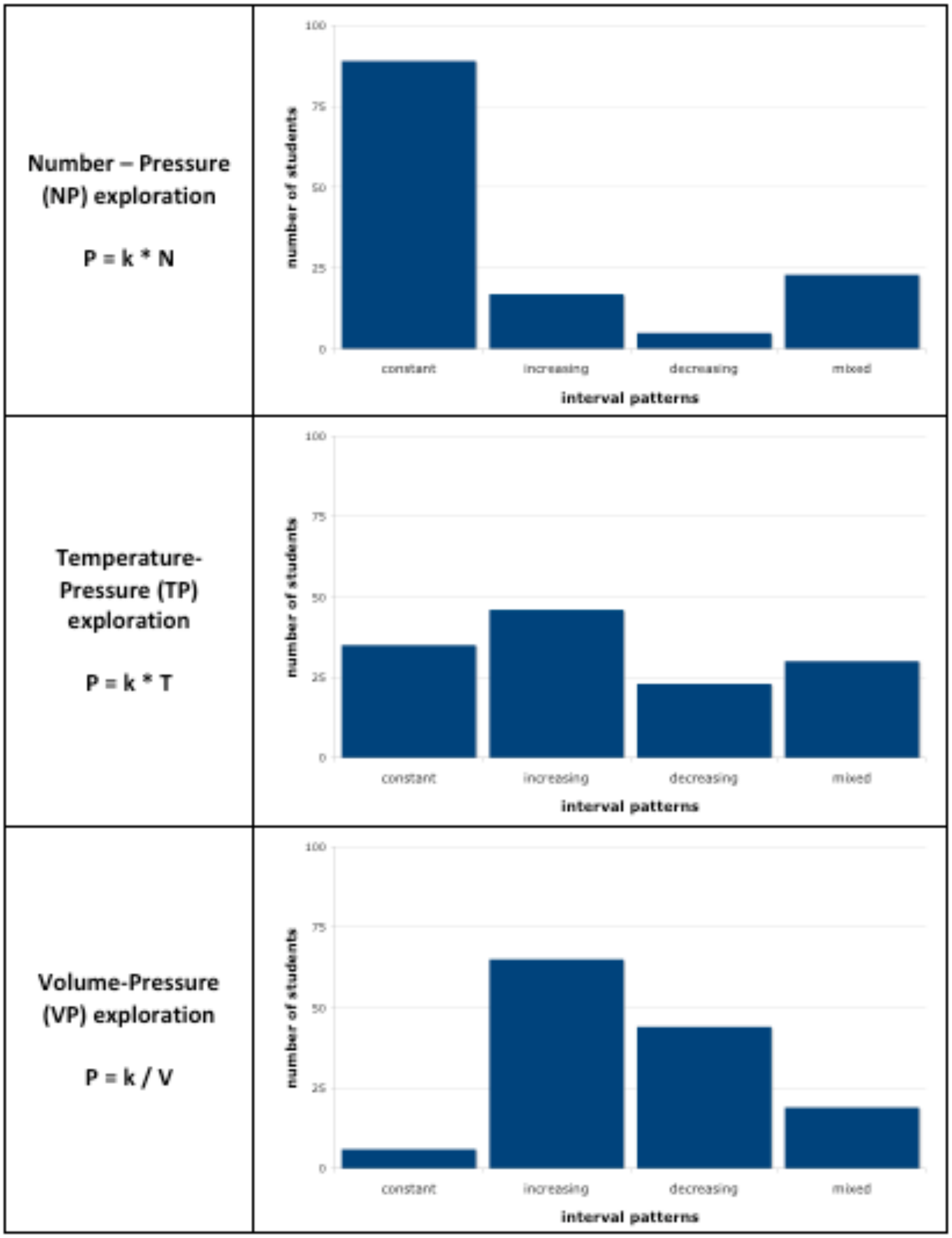


Figure 3: Students' model exploration strategies. Corresponding gas law equations are on the left: k is a constant; P is pressure; N is the number of gas particles in a container; V is its volume; T is the temperature of the gas. n = 134.

Students' consistency in the three explorations was tested by looking at the actual patterns and at their fitness. Regarding the first, the Chi-square Test shows no consistency between the three explorations. As for the second, the Chi-square Test shows no significant relationships between NP and TP ($\chi^2(1)=0.27$, *ns*), or between VP and TP ($\chi^2(1)=0.15$, *ns*). A significant relationship is found between NP and VP ($\chi^2(1)=4.55$, $p=.033$). For the latter, the main part of the students (37%) are using fit strategies in both explorations

<u>Prior Knowledge and Model Exploration Patterns</u>

Students' pre-test and post-test results (Table 1) show greater learning gains for the micro- and micro-to-macro subscales and for conceptual understanding.

Table 1
Descriptive and Comparative Statistics of Students' Content Knowledge in the Connected Chemistry Curriculum (CC1) with Respect to the Conceptual Framework

Conceptual framework component (# of items in questionnaire)	Test Pre *M (SD)*	Post *M (SD)*	Paired t	Effect size Cohen's *d* (95% CI).
All (19)	56 (17)	66 (19)	-17.61**	0.55 (0.46-0.65)
Form of access				
Micro (3)	45 (28)	60 (31)	-13.14**	0.51 (0.41-0.60)
Macro (3)	76 (29)	82 (27)	-6.01**	0.21 (0.12-0.31)
Mathematical (1)	42 (49)	58 (49)	-8.21**	0.33 (0.23-0.42)
Bridge				
Micro/Macro (8)	56 (21)	65 (22)	-11.72**	0.42 (0.33-0.51)
Conceptual/ Mathematical models (4)	56 (28)	62 (29)	-6.83**	0.21 (0.12-0.21)
Quantitative	56 (29)	65 (30)	-7.37**	0.31 (-1.62-2.16)
Qualitative	55 (18)	65 (20)	-17.33**	0.53 (-0.76-1.68)

Note. Scores are mean percentages of correct answers on pre-test and post-test questionnaire.
**$p < .01$

Results of a logistic regression between the students' prior knowledge and the exploration patterns' fitness, and ANOVA for the cumulative fitness are presented in Table 2. Figures 4-5 show significant associations. These results show that only some of the components of knowledge impact the students' model exploration patterns, and in different ways for the different relationships: NP exploration is impacted by conceptual but not algorithmic understanding; and more specifically by prior knowledge of both the micro- and macro-levels as well as bridging the macro-level with its mathematical representations; TP exploration by understanding of the macro-level alone; VP exploration by understanding of the micro-level alone.

Table 2: Logistic regression testing the impact of prior knowledge on the three model exploration patterns, and ANOVA of their cumulative fitness score.

Prior knowledge component	NP exploration pattern χ^2	p	TP exploration pattern χ^2	p	VP exploration pattern χ^2	p	All exploration patterns F(1,83)	p
All	6.36	.01	0.02	ns	0.19	ns	.584	ns
Micro	3.50	.03	0.20	ns	5.43	.02	2.324	ns
Macro	3.51	.03	3.11	.04	0.10	ns	.580	ns
Math	0.36	ns	0.14	ns	0.03	ns	.147	ns
Micro/Macro	0.36	ns	0.40	ns	0.05	ns	.213	ns
Math/Macro	8.53	.003	0.15	ns	0.32	ns	1.316	ns
Conceptual	7.67	.006	0.10	ns	0.11	ns	.329	ns
Algorithmic	1.98	ns	0.01	ns	0.15	ns	.018	ns

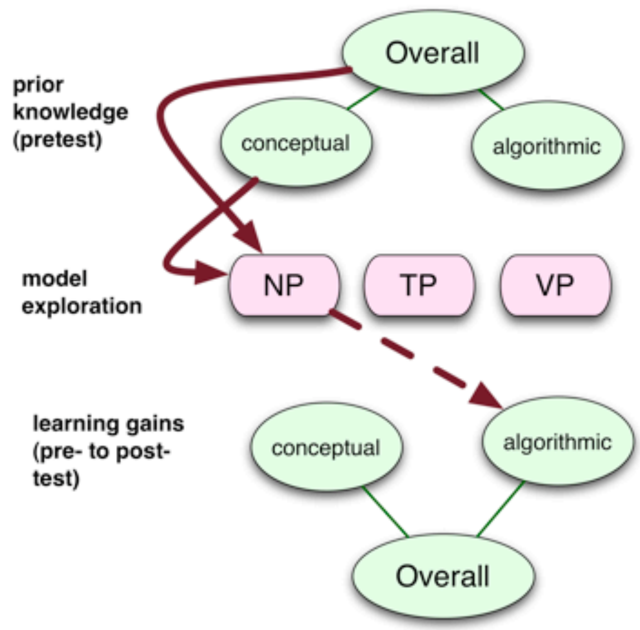


Figure 4: Significant associations between prior knowledge and model exploration patterns, exploration patterns and learning gains, sorted by conceptual and algorithmic. Associations significant only at the 0.1 level are marked in dashed lines.

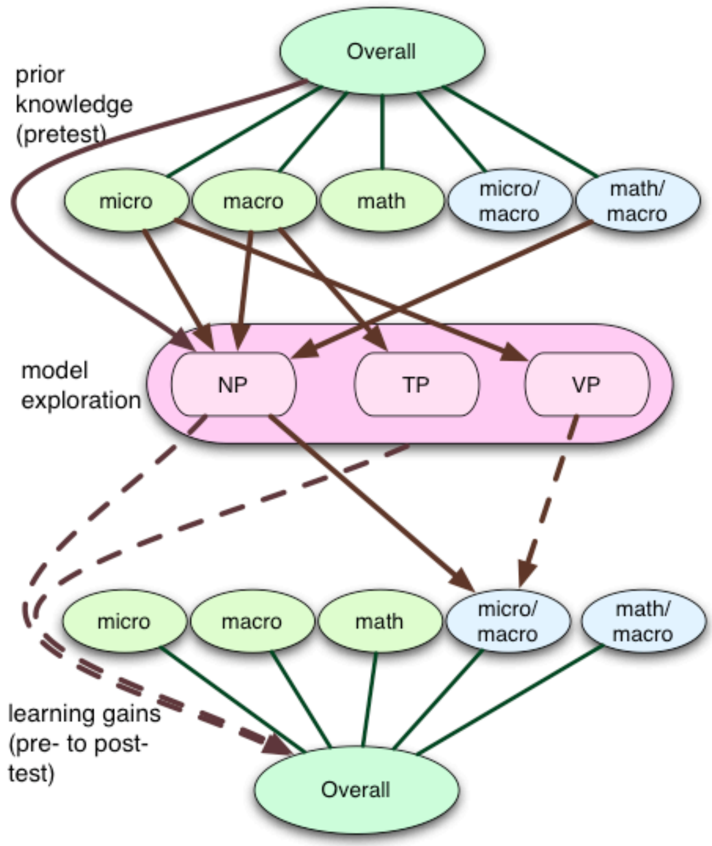


Figure 5: Significant relationships between prior knowledge and model exploration patterns, model explorations and learning gains, sorted by levels (micro, macro, math) and bridges (micro/macro, math/macro) from the pre-test. Associations significant only at the 0.1 level are marked in dashed lines.

<u>Model Exploration Patterns and Learning</u>

Independent t-tests related the explorations' fitness and learning gains (Table 3). Small associations between exploration patterns and learning are found. Fitness in NP exploration (and more weakly for VP exploration) is associated with learning more advanced complexity reasoning: bridging of micro/macro-levels. Less significant, cumulative fitness of the three explorations and fit NP exploration is associated with overall learning gains, and, more specifically along the quantitative dimension.

Table 3: Impact of model exploration patterns on subsequent learning gains.

Knowledge component	NP exploration pattern		TP exploration pattern		VP exploration pattern		All explorations	
	t-test[a]	p	t-test	p	t-test	p	F(3,*n-1*)	p
All	1.76 (244)[b]	.08	1.01 (222)	.31	.55 (264)	.58	2.64 (81)	0.06
Micro	-.911 (241)	.36	-1.063 (178)	.29	.977 (212)	.33	.349 (66)	.79
Macro	.611 (232)	.54	1.429 (212)	.15	.198 (258)	.84	.418 (78)	.74
Micro/Macro	2.08 (258)	.039	.112 (225)	.91	1.82 (280)	.07	1.516 (81)	.22
Math/Macro	-1.497 (227)	.14	.824 (202)	.41	-.022 (254)	.98	.286 (76)	.84
Conceptual	.116 (249)	.91	1.249 (230)	.21	-1.414 (268)	.16	1.987 (85)	.12
Algorithmic	-1.728 (222)	.085	.506 (201)	.61	1.010 (246)	.31	.578 (73)	.63

[a] Independent samples t-test
[b] n's are in parentheses

DISCUSSION

How do students search for information within computer models? We have found that the underlying model behaviors as well as prior knowledge interact in shaping the particular form by which information is searched for. Furthermore, learning is related to how students explore models, though to a lesser extent. These findings point to the potential of using such probes in learning environments, as indicators of understanding and possibly as providing information upon which support for learning can be based.

This study investigated students' model exploration strategies as a possible reflection of their understanding of the domain under inquiry. More specifically, the data they collected in the process of constructing equations was examined. The data-points' spacing was inspected as providing the best information regarding the model's mathematical behavior. We have found that for two of three explorations, the students used mainly fit strategies, even when one of them was the more difficult inverse function. Moreover, they were mainly consistent in their adaptation to the model's mathematical behavior.

Deep conceptual knowledge was found to impact how some of the models are explored even though the students were focused on obtaining quantitative information. Though less significant, one can also see that while conceptual knowledge guides search for information, successful searches are also related to learning along the complementary quantitative/algorithmic dimension of knowledge.

Regarding a more detailed framework that highlights a complexity perspective and the relationship between conceptual and mathematical models, it was found that the explorations were impacted by different components of prior knowledge. The fitness of the NP model exploration, one more indicative of prior knowledge and learning, was found to be related to understanding both micro- and macro-levels of the system, as well as bridging the conceptual and mathematical models. Fit explorations of this model are also related to greater learning gains in the more advanced problems of bridging the micro-and macro-levels. The TP model, one more ambiguous with respect to the students' strategies, was explored with more fit strategies when students had a greater understanding of the macro-level of the gas system. VP model exploration, one more difficult as a result of its inverse function behavior (Nemirovsky, 1994), was more fit when students had a greater understanding of the micro-level.

In the final paper, we will provide a discussion of these results with respect to students' understanding of complexity and the three relationships they explored, the study's limitations and extensions, and potential applications of these findings in supporting learning about complex systems in computer model-based learning environments.

REFERENCES

Ardac, D., and Akaygun, S. (2004). Effectiveness of multimedia-based instruction that emphasizes molecular representations on students' understanding of chemical change. *Journal of Research in Science Teaching* 41: 317–337.

Chiu, M-H. (2001). Algorithmic Problem Solving and Conceptual Understanding of Chemistry by Students at a Local High School in Taiwan. Proceedings of the National Science Council, ROC(D), 11(1), 20-38.

Costu, B. (2007). Comparison of Students' Performance on Algorithmic, Conceptual and Graphical Chemistry Gas Problems. *Journal of Science Education and Technology*, 16, 379-386.

*** (2003). Modeling across the curriculum: Scaling up modeling using technology. Paper presented at the 25th *Annual Meeting of the Cognitive Science Society, CogSci* 2003, Boston, Massachusetts, USA, July 31-August 2, 2003.

Holland, J.H. (1995). *Hidden Order: How Adaptation Builds Complexity*. Cambridge, MA: Helix Books/Addison-Wesley.

Horwitz, P., and Christie, M. (1999). Hypermodels: Embedding curriculum and assessment in computer-based manipulatives. *Journal of Education* 181: 1-23.

Kauffman, S. (1995). *At home in the universe: The search for the laws of self-organization and complexity*. Oxford, NY: Oxford University Press.

Kozma, R.B. (2000). The use of multiple representations and the social construction of understanding in chemistry. In Jacobson, M.J., and Kozma, R.B. (Eds.), *Innovations in science and mathematics education: Advanced designs for technologies of learning*, Lawrence Erlbaum, New Jersey, London, pp. 11-46.

*** (2009a). Crossing levels and representations: The Connected Chemistry (CC1) curriculum. *Journal of Science Education and Technology*, 18(3), 224-242.

*** (2009b). Students' Learning with the Connected Chemistry (CC1) Curriculum:Navigating the Complexities of the Particulate World. *Journal of Science Education and Technology*, 18(3), 243-254,

Lin, H-S, Cheng, H-J (2000). The assessment of students and teachers' understanding of gas laws. *Journal of Chemical Education* 77(2): 235-238.

Nakhleh, M. B. (1993). Are our students conceptual thinkers or algorithmic problem solvers? Identifying conceptual students in general chemistry. *Journal of Chemical Education*, 70(1), 52–55.

Nemirovsky, R. (1994). On Ways of Symbolizing: The Case of Laura and the Velocity Sign. *Journal of Mathematical Behavior*, 13(4), 389-422.

Niaz, M., & Robinson, W.R. (1992). From 'Algorithmic mode' to 'conceptual gestalt' in understanding the behavior of gases: An epistemological perspective. *Research in Science and Technological Education* 10(1): 53-64.

Russell, J.W., Kozma, R.B., Jones, T., Wykoff, J., Marx, N., and Davis, J. (1997). Use of simultaneous-synchronized macroscopic, microscopic, and symbolic representations to enhance the teaching and learning of chemical concepts. *Journal of Chemical Education* 74(3): 330-334.

Snir, J., Smith, C. L., and Raz, G. (2003). Linking phenomena with competing underlying models: A software tool for introducing students to the particulate model. *Science Education* 87: 794–830.

*** (2003). Connected Chemistry – Incorporating interactive simulations into the chemistry classroom. *Journal of Science Education and Technology, 12*(3), 285-302.

van der Meij, J., and de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction* 16: 199-212.

*** (1999a). *NetLogo*. Center for Connected Learning and Computer-Based Modeling, Northwestern University. Evanston, IL. http://ccl.northwestern.edu/netlogo.

*** (1999b). GasLab: An extensible modeling toolkit for exploring micro- and macro- views of gases. In Feurzig, W., and Roberts, N. (Eds.), *Modeling and simulation in science and mathematics education*, Springer, New York, pp. 151-178.

*** (2003). Statistical mechanics for secondary school: The GasLab modeling toolkit. *International Journal of Computers for Mathematical Learning*, 8(1), 1-41 (special issue on agent-based modeling).