Computers & Education 56 (2011) 556–573

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Mining students' inquiry actions for understanding of complex systems

Sharona T. Levy [a,*], Uri Wilensky [b]

[a] Faculty of Education, Learning, Instruction and Teacher Education, University of Haifa, 31905 Mount Carmel, Haifa, Israel
[b] Departments of Learning Sciences and Computer Science, Center for Connected Learning and Computer-Based Modeling, Northwestern Institute on Complex Systems (NICO), Northwestern University, Evanston, IL, United States

ARTICLE INFO

Article history:
Received 26 May 2010
Received in revised form
18 September 2010
Accepted 22 September 2010

Keywords:
Simulations
Modeling
Agent-based
Interactive learning environments
Secondary education
Improving classroom teaching

ABSTRACT

This study lies at an intersection between advancing educational data mining methods for detecting students' knowledge-in-action and the broader question of how conceptual and mathematical forms of knowing interact in exploring complex chemical systems. More specifically, it investigates students' inquiry actions in three computer-based models of complex chemical systems when their goal is to construct an equation relating physical variables of the system. The study's participants were 368 high-school students who interacted with the Connected Chemistry (CC1[1]) curriculum and completed identical pre- and post-test content knowledge questionnaires. The study explores whether and how students adapt to different mathematical behaviors of the system, examines how these explorations may relate to prior knowledge and learning in terms of conceptual and mathematical models, as well as components relating to understanding systems. Students' data-collection choices were mined and analyzed showing: (1) In about half the cases, mainly for two out of the three models explored, students conduct mathematically-astute (fit) explorations; (2) A third of the students consistently adapt their strategies to the models' mathematical behavior; (3) Fit explorations are associated with prior conceptual knowledge, specifically understanding of the system as complex, however, the three explorations' fitness is predicted by the understanding of distinct sets of systems' components; (4) Fit explorations are only somewhat associated with learning along complementary dimensions. These results are discussed with respect to 1) the importance of a conceptual understanding regarding individual system elements even when engaged in large-scale quantitative problem solving, 2) how distinct results for the different models relate to previous literature on conceptual understanding and particular affordances of the models, 3) the importance of engaging students in creating mathematical representations of scientific phenomena, as well as 4) educational applications of these results in learning environments.

I have always thought the actions of men the best interpreters of their thoughts. ~John Locke [Excerpt from An Essay Concerning Humane Understanding, Volume 1 MDCXC, Based on the 2nd Edition, Books 1 and 2]

1. Introduction

1.1. Motivation

What does exploratory action reveal about understanding? More specifically, what can we learn from students' inquiry actions regarding their understanding of the investigated domain? This study involves students' active exploration of computer models in a chemistry-learning unit and educational data mining of such explorations. It will be shown that their data-collection choices reflect an understanding of the system at hand.

* Corresponding author. Tel.: +972 8 931 8784; fax: +972 4 824 9354.
E-mail addresses: stlevy@construct.haifa.ac.il (S.T. Levy), uri@northwestern.edu (U. Wilensky).

[1] CC1 denotes Connected Chemistry, Chapter 1, a unit devoted to the topic of gas laws and kinetic molecular theory. Under the umbrella of Connected Chemistry, the Center for Connected Learning and Computer-based Modeling (CCL) has created several curricular units.

0360-1315/$ – see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.compedu.2010.09.015

Furthermore, the study addresses a central question in science learning: How do conceptual and mathematical models of knowledge interact with each other, and in particular – with regards to exploratory action? Research has demonstrated that students' ability to solve quantitative chemistry problems does not necessarily reflect a conceptual understanding of the problem at hand (Nakhleh, 1993; Niaz & Robinson, 1992). It will be shown that this picture changes in the context of inquiry activities that involve exploring well-designed computer models and constructing the associated mathematical representations. Many students adapt their search for quantitative information to the model's mathematical structure. Moreover, this adaptation corresponds with a deeper prior understanding of the conceptual model, but not the mathematical model. Thus, in an inquiry context of exploring models with mathematical goals, conceptual knowledge plays a central role.

Commonly, students' understanding of science is assessed through their verbal and textual articulations as well as their performance in solving problems. These expressions are all part of the domain of symbols. However, students' rich expressions through action – exploratory (bringing knowledge from the world) or performatory (enlarging desires onto the world) – (Bruner, 1973; Gibson, 1988) – while informative to expert teachers, are rarely drawn upon in educational settings. As data mining of students' actions in digital learning environments advances, new expressions of knowing become accessible (Baker & Yacef, 2009; Pahl, 2004). 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. In this study, we propose a simple method by which students' mathematical understandings of the behavior of complex systems can be assessed through the actions they perform with computer-based models.

In the current study, the students are involved in forming quantitative descriptions of the explored systems. The activity of creating rather than consuming and manipulating given equations is infrequent in science curricula. Previous studies into activities that engage students in constructing mathematical representations of scientific phenomena are sparse and target more advanced undergraduate students (Bopegedera, 2007; Laugier & Garai, 2007). It is proposed that experiences in which students explore models that provide access to causal explanations and use them to create symbolic representations may promote a deeper understanding of the mathematical representations and how they relate to conceptual explanatory models.

1.2. Educational setting: the Connected Chemistry (CC1) curriculum

The students in this investigation interacted with the Connected Chemistry curriculum (named CC1 to denote Connected Chemistry Chapter 1; Levy, Novak & Wilensky, 2006; you are invited to download the curriculum through the provided website link). Connected Chemistry is a computer-based environment for learning the topics of gas laws and kinetic molecular theory in chemistry. It views chemistry from an "emergent" perspective, how macroscopic phenomena result from the interaction of many submicroscopic particles. In analyzing systems, the terms "macro" and "micro" are used to describe aggregation levels in the system – whether a global or local perspective is approached – with the goal of relating these two levels. In this paper, the term "micro" is supplanted by the term "submicroscopic" to align with a chemistry perspective, where "micro" describes a real size, and is not appropriate for describing gas particles that are much smaller. CC1 is part of the Modeling Across the Curriculum project (Gobert et al., 2003). It employs NetLogo (Wilensky, 1999a) agent-based models, models that compute a system's behavior from its components' individual actions and interactions with other components. 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. Under the umbrella of Connected Chemistry, the Center for Connected Learning and Computer-based Modeling (CCL) has created several curricular units (Stieff & Wilensky, 2003; Wilensky, 1999b, 2003). These Connected Chemistry units focus on topics in chemistry and employ multi-agent NetLogo models to enable students' inquiry: creating, manipulating and observing interactions between objects at the molecular level in order to gain insight into emergent patterns and macroscopic phenomena.

As the system under study involves a multitude of interacting entities, the topic is introduced through a complexity perspective. This perspective combines an understanding of the objects, actions and interactions at the submicroscopic level with that of the macroscopic level, and relating the two. 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 (Wilensky, 1999a) is a modeling language and environment that supports creating, exploring and analyzing agent-based models, such as those used in CC1. Exploring this type of 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, Smith, & Raz, 2003; van der Meij & de Jong, 2006).

A conceptual framework was developed to structure students' experiences and learning through exploring the models (Levy & Wilensky, 2009a). The same framework will be used in the current study to investigate students' understanding of the chemistry ideas they explore with the Connected Chemistry activities (see Fig. 1). The framework describes three spheres of knowledge: conceptual, symbolic and physical; and four forms of access to understanding the system: descriptions at the submicroscopic and macroscopic levels, mathematical representations and physical experiences. Activities were designed to help students build an integrated view of the chemical system, by focusing on understanding along each form of access, and promoting multiple transitions between the spheres of knowledge. The macro-level observable phenomena were used to bridge between the three spheres and support these shifts.

An investigation was conducted into high-school students' learning with Connected Chemistry (Levy & Wilensky, 2009b). Results have shown a strong effect size for embedded assessment and a medium effect size regarding pre-post-test questionnaires. Stronger effects were seen regarding learning about the submicroscopic level and bridging between it and the macroscopic level, two central components in understanding complex chemical systems. Significant shifts were found in students' epistemologies of models: understanding models as representations rather than replicas of reality and as providing multiple perspectives (Gobert et al., 2010).

The focal activities in the current study involve students' construction of the gas law equations. It was found that more than half the students succeeded at this task, including the four-variable Ideal Gas Law: Pressure $\times$ Volume $=$ constant $\times$ Number of particles $\times$ Temperature (Levy & Wilensky, 2009b). Through collecting data about the system's response to various changes, and further

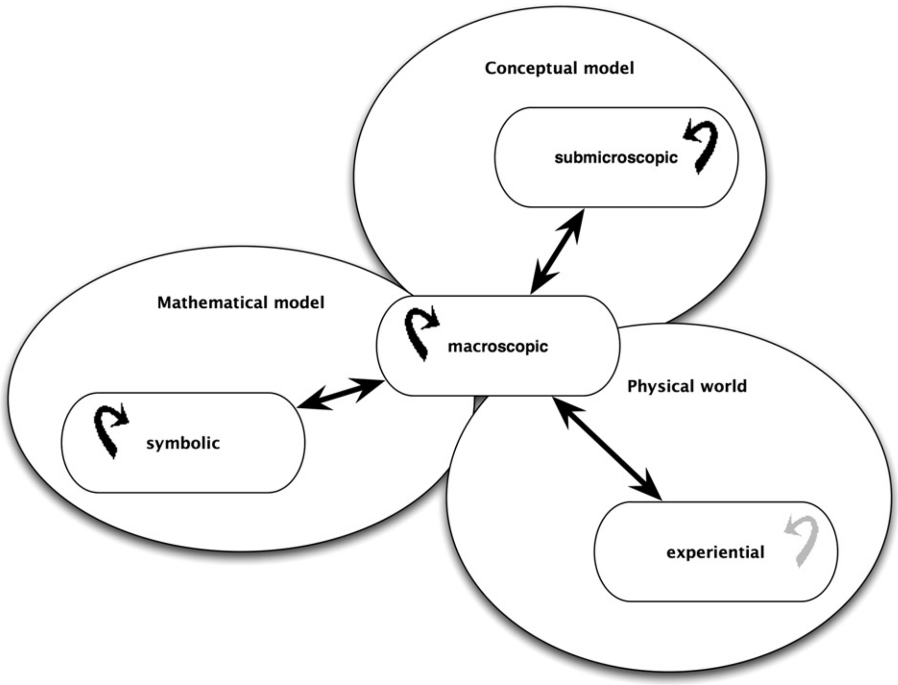


Fig. 1. Conceptual framework for supporting learning through model exploration in the Connected Chemistry curriculum (CC1). With kind permission from Springer Science+Business Media: Journal of Science Education and Technology, Crossing Level and Representations: The Connected Chemistry (CC1) Curriculum, 2009, 229, Sharona T. Levy & Uri Wilensky, Fig 1.

transformations of this data, the students were able to formulate the algebraic form of these relationships, the gas laws. The current study focuses on how they collect the data, from which such constructions were created.

1.3. Related literature

1.3.1. The importance of conceptual understanding in learning chemistry

Relevant to this study is the relationship between students' conceptual and mathematical understandings of how gas particles are impacted by changes to the system, such as heating or expanding the container, which holds them. These issues can be addressed with mathematical representations such as the algebraic gas laws (e.g. Boyle's Law: Pressure $\times$ Volume $=$ constant; when temperature and the number of gas particles are constant). They can also be discussed within a conceptual model of the gas particles' properties, behaviors and interactions. For example, expanding the containing vessel results in greater space through which the gas particles can travel; due to the greater travel distance within the container, each particle hits the container walls less frequently so that the pressure exerted on the walls, which is related to the rate of such collisions, is reduced. In learning such topics in high-school chemistry and physics, the focus is usually upon mathematical problem solving. A typical test question provides a bare setting involving objects, materials, values of several variables, and asks the students to find the value of the missing variables: "At constant pressure and 25 °C a sample of gas occupies 4.5 L. At what temperature will the gas occupy 9.0 L?" To answer this question, the student needs to select the correct equation ($Volume_1/Temperature_1 = Volume_2/Temperature_2$) and insert the correct values for volume and the temperature in Kelvin degrees.

This focus on the mathematical model has elicited critique regarding the "mindless" form of knowledge application as well as research into how students' conceptual and mathematical models relate to each other. In the literature on chemistry education, they are usually named conceptual and "algorithmic" types of knowledge, the latter referring to well-practiced quantitative problem solving. 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; yet their conceptual understanding lags far behind (Gabel, Sherwood, & Enochs, 1984; Mason, Shell, & Crawley, 1997; Nakhleh, 1993; Niaz & Robinson, 1992; Nurrenberg & Pickering, 1987; Russell et al., 1997; Sawrey, 1990; Zoller, Dori, & Lubezky, 2002; however, see Chiu, 2001; Costu, 2007). Moreover, such understandings may be even more limited when the problems do not fall into familiar and practiced problems (Lin & Cheng, 2000).

In this study, we focus in on how the conceptual and mathematical models may interact within the setting of mathematically geared explorations that involve computer-based models that provide access to conceptual understandings.

1.3.2. Educational data mining

Educational data mining (EDM) is an emerging field aimed at developing methods of exploring data from computational educational settings and discovering meaningful patterns (Baker & Yacef, 2009). Pahl (2004) focuses on the interactive nature of learning in such settings and has defined EDM in the following way: "Data mining is a non-intrusive objective analysis technology that is proposed as a central evaluation technology for the analysis of the usage of computer-based educational environments and in particular, of the interaction with educational content". Due to this study's concentration on students' interactions with models, EDM has been the method of choice.

The *goals* of EDM are varied: from constructing and improving student models, designing support in digital settings to scientific discovery about learners and learning (Baker, 2010). Its *areas of application* are predictive (decision support), generative (create new or improved designs for learning) or explanatory (scientific analysis) (Pahl, 2004) and include studies into individual learning from educational software, computer supported collaborative learning, computer adaptive testing and factors relating to student failure or non-retention in courses (Baker & Yacef, 2009). Hershkovitz and Nachmias (2009) offer another classification of applications for EDM using two dimensions; one is whether a group or individual are the subject of research and the second is the time reference regarding learning – whether the end point or process are of interest. While the present study concentrates on scientific discovery about learners and is of an explanatory type, it is also generative in that its goal is also to support improved designs for learning with models, in which inquiry actions can be used as a diagnostic tool for mathematical and conceptual understandings of the explored phenomena. It can also be classified as focused on process and on the individual learner, as differences among individuals in how they interact with computational media is explored.

Methods in educational data mining include usage statistics, classification and prediction, association rules, sequential pattern analysis and time series, clustering, relationship mining, discovery with models, and distillation of data for human judgment (Baker, 2010; Pahl, 2004). The current research falls into the category of *relationship mining*, identifying strong relationships between a large set of variables (various components of prior knowledge and learning gains) and a particular variable of interest (model exploration strategies).

Few studies investigate students' interactions with exploratory learning environments, open learning platforms such as computer models. In one such recent study, Amershi and Conati (2009) have found that undergraduate computer-science students' interactions with exploratory applets distinguish among successful and less successful learners: the former pause for longer times after presentation of a new problem and tend to follow suggested sequences and guidelines. The current research hopes to further understanding of such interactions and related learning.

1.4. Framework of the study

This investigation focuses on how students explore computer models of complex systems and what such explorations reveal about their understanding. We examine data that students collect 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 laboratories structures a linear constant addition of values to the *independent* variable (e.g. 100-150-200-250). 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[2]. An example addressed in this study is the inverse function. If the relationship between the variables is inverse ($y = k/x$; y as the dependent variable, x as the independent variable, k a constant), the system's behavior changes more quickly for smaller values of the independent variable. If one wanted to observe this region of swift change, it would make more sense to space the data collection at higher density for lower values of the independent variable. Fig. 2 presents a comparison between two strategies in data collection for the same inverse function. On the left, Fig. 2a shows sampling at constant intervals for the independent variable on the x-axis: 50, 100, 150…. On the right, Fig. 2b presents sampling at constant intervals for the dependent variable on the y-axis: 10, 20, 30… One can see that the figure on the right (Fig. 1b) captures a more detailed sampling of the zone of greater change than does the left one.

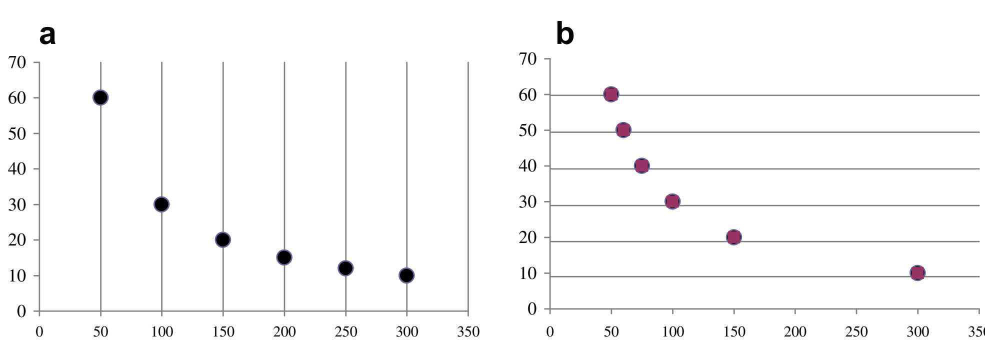


Fig. 2. Graphical representation of sampling six points along the same inverse function. On the left (a), constant intervals are sampled on the x-axis. On the right (b), constant intervals are sampled on the y-axis.

In the study, we consider three possibilities regarding how students would explore the models when their goal is to quantify relationships. One is that the students would employ mainly the commonly taught laboratory strategy: vary the independent variable at even increments. The second possibility is that no dominant pattern would emerge: students were not guided in their collection of data, as this is an open exploratory environment. The third possibility is that they would increment the dependent variable at approximately even

[2] This observation assumes experimental error in collecting data regarding the system.

intervals to capture the full range of change. Varying the independent variable systematically is reasonable when little is known about how the system behaves, as it captures a wide span of values in the parameter space. Knowing how the dependent variable changes requires *prior knowledge* of the domain. Thus, it is proposed that the latter strategy reflects understanding of the system at hand. In the study, we look into students' data collection strategies in terms of spaces between the values of data collected, 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 value of these findings is in detecting knowing-in-action (Schön, 1983) a form of understanding that is more difficult to tap onto. In educational settings, such detection may prove valuable is helping gauge students' progressing understanding in service of appropriate support.

1.5. Research questions

The following research questions guide this study:

Research question 1: What strategies describe how students explore computer models when they are collecting data with which they will construct an equation relating the system's variables?
Research question 2: How consistent are the strategies students use in exploring the computer models in terms of fitness to the models' mathematical behavior?
Research question 3: How does prior knowledge impact students' model exploration strategies?
Research question 4: What associations can be found between students' exploration strategies and their learning gains regarding the related content?

2. Materials and methods

In this study content knowledge questionnaires administered before and after the curricular activities were analyzed, three logged activities were data-mined, consistency between activity patterns was examined, associations between pre-test knowledge and activity patterns and between activity patterns and learning gains were explored, as elaborated in the following section.

2.1. Procedure

The students engaged with seven activities of CC1 as part of their high-school chemistry course during the 2005–2006 year, replacing the topic of gas laws and kinetic molecular theory. Before and after the activities, spaced 2–3 weeks apart, the students completed content knowledge questionnaires. The students' click-type interactions with the models and answers to open and closed questions were logged through the Pedagogica environment. The research was reviewed and approved by an Institutional Review Board.

The focal sections in this analysis are part of three of the later activities and 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 manipulated one variable: the number of particles (*N*[3]), temperature (*T*) or volume (*V*) and observed the resultant pressure (*P*) (Fig. 3). They had explored the models prior to the activity, and following it, tested their constructed equations with the models.

In each of the three activities, manipulating the independent variable is done in a different way and equilibration time until a reading can be made is varied. Perhaps more important, the change in the *independent* variable is immediately observable for two explorations (*NP*, *VP*) but slower to stabilize for the third (*TP*). These differences are described in Table 1 and portrayed in Fig. 4.

2.2. Participants

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 13 diverse high-schools across the United States, whose principal or science department head volunteered to participate in our project; thus, individual students did not themselves volunteer to participate. These schools represent a wide range of socio-economic levels (estimated by the percentage of students in the school who receive free- or reduced-lunch) with schools ranging from 0% to 41% of students receiving free- or reduced-lunch. Across all 13 schools, the average percentage of students receiving free- or reduced-lunch was 16%. School size ranged from 120 students to slightly more than 2000 students, with the average school size 1200 students. The average number of students in the science classes in these schools was 24 students.

2.3. Instruments

In this study, two main instruments have been used: a pre-test/post-test content knowledge questionnaire and logs of the students' actions in exploring the models.

Identical pre- and post-test *content knowledge questionnaires* were administered (Appendix A). They contain 19 multiple choice questions which were designed to assess students' understanding of the gas laws and kinetic molecular theory with particular focus on targeting

[3] The following symbols are used: *N* for the number of particles in the container; *V* for the volume of the container; *T* for temperature; *NP* is the relationship between the number of particles and pressure; *VP* is the relationship between volume and pressure; *TP* is the relationship between temperature and pressure.

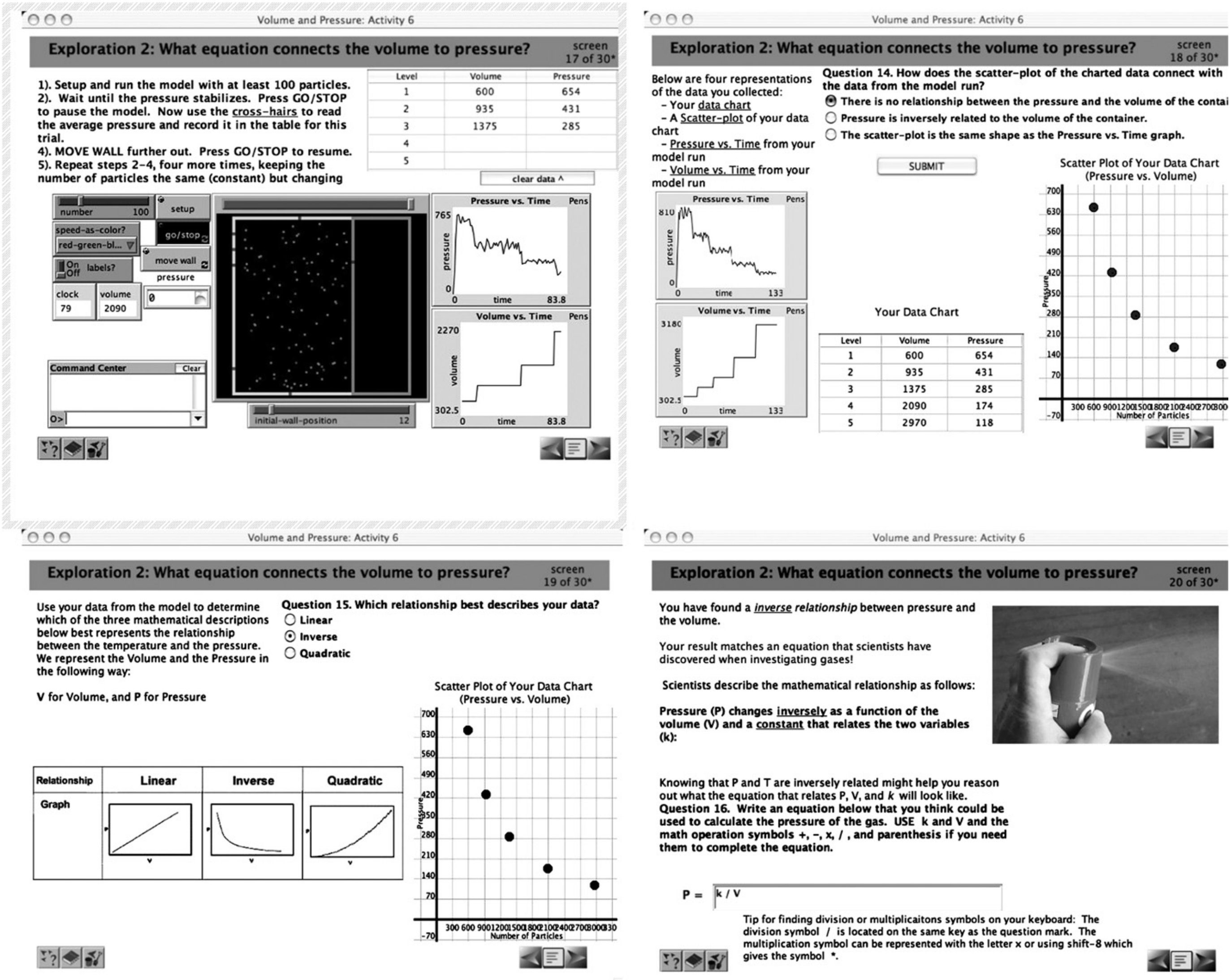


Fig. 3. A sequence of four screenshots from the "Volume and Pressure" activity, portraying construction of Boyle's Law through experimenting with the agent-based model. The focal screen in the study is highlighted.

problematic areas as reported in the science education literature. It is fully described elsewhere in a paper describing research into students' learning with Connected Chemistry (Levy & Wilensky, 2009b). Two sets of subscales were created: one distinguishing between conceptual and mathematical questions; the other, more fine-tuned, focused on complex systems and mathematical modeling (submicro- and macro-description levels, mathematical representations, submicro/macro and model/math bridges). The questionnaire is related to two of the three

Table 1
Comparison between the connected chemistry models in terms of acting on the model and subsequent changes until stabilization.

Model	Manipulation of the independent variable	Change to the independent variable following manipulation	Typical time from manipulation until equilibration of pressure, the dependent variable (seconds)
Number and pressure (*NP*)	A slider is set to the number of added particles, and a neighboring button is pressed to release them into the box.	Increased density of particles in container is observed immediately.	10
Volume and pressure (*VP*)	A button is pressed to enable changing the volume and the container in the model is clicked upon at the location where the wall should move to	Increased volume is observed immediately	4
Temperature and pressure (*TP*)	One of two buttons is pressed – one for heating, one for cooling – changing the temperature of the walls. The number of button presses determines the increment of temperature change.	The container's walls are first heated (or cooled); then, moving particles may hit the wall and become faster (or slower); finally, through collisions with other particles – the energy is redistributed until a stable temperature is reached.	10

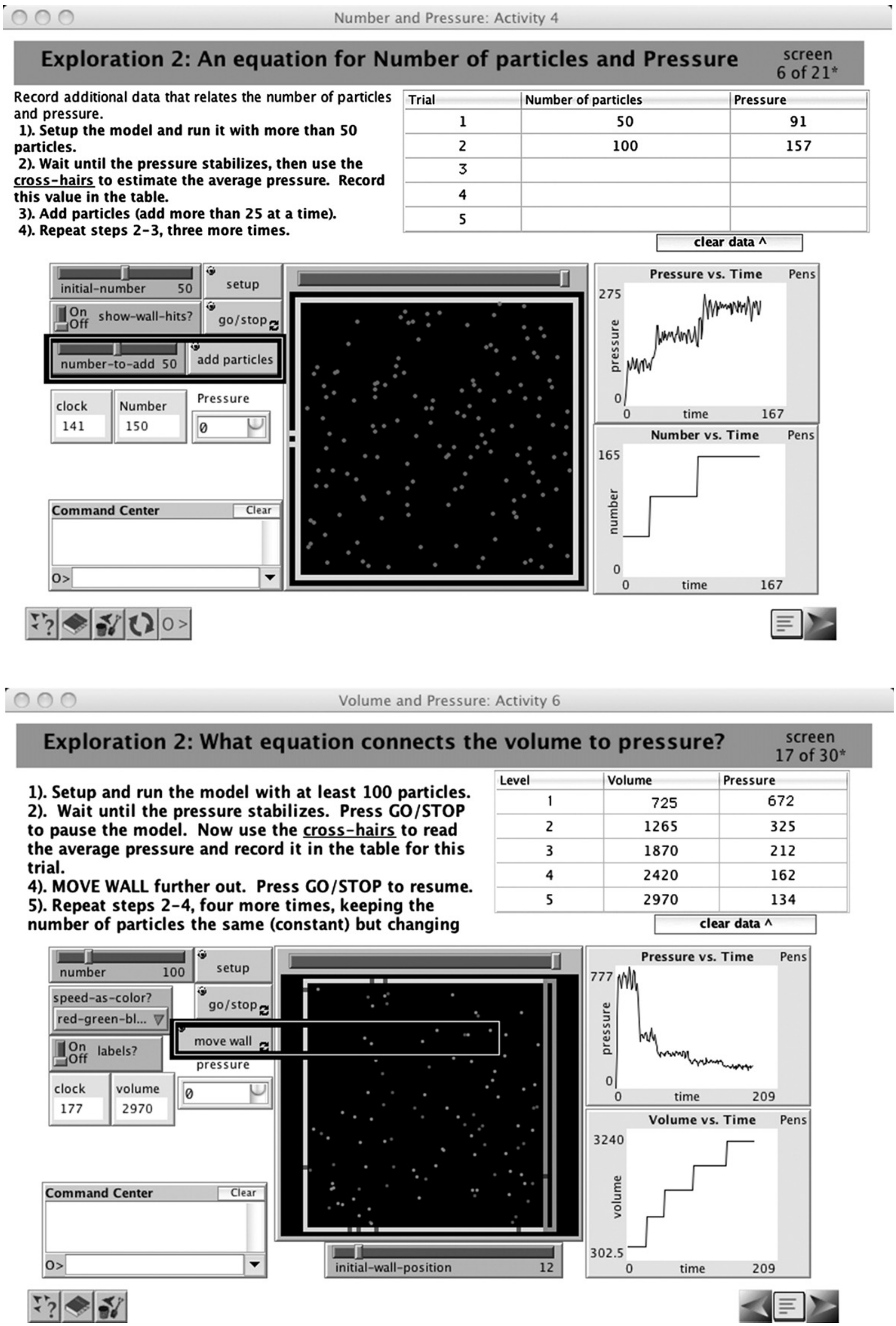


Fig. 4. Focal screens in the three activities (from top to bottom): Number and pressure, Volume and Pressure, Temperature and Pressure. The manipulated model components that change the independent variables are highlighted with double-framed black rectangles.

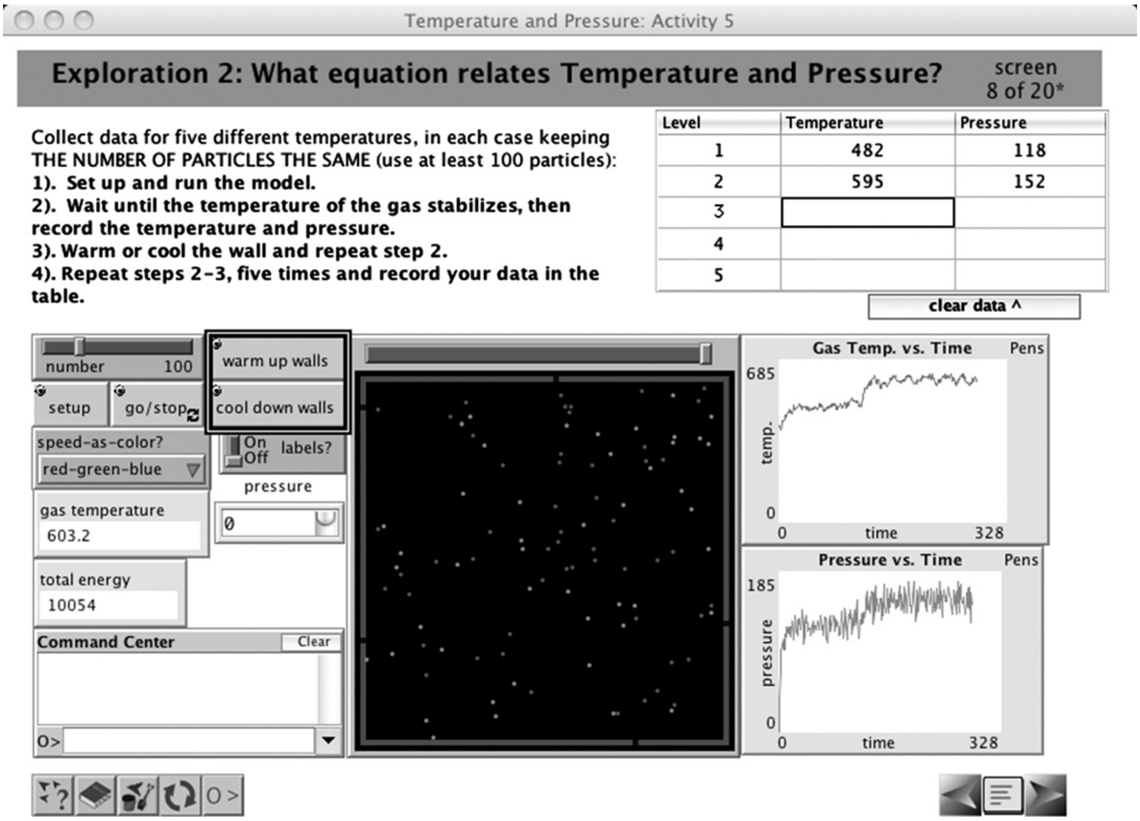
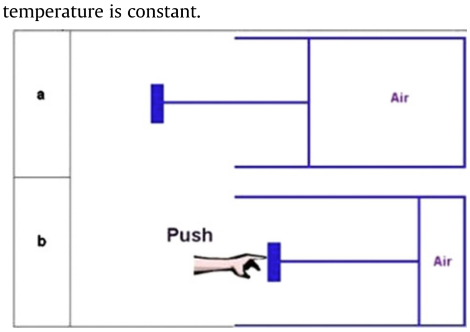


Fig. 4. (*continued*).

spheres of knowledge in the conceptual framework (Fig. 1): the conceptual model and the mathematical model, as well as the links and bridges between the different forms of access. Within the conceptual model, a further separation is made between a focus on submicroscopic and macroscopic levels of descriptions and transitions among them. All questionnaires and activities were scripted and the students' answers were made available for analysis. Typical items are provided in Table 2.

Table 2
Sample questionnaire items and their relation to the conceptual framework of the study.

Conceptual/mathematical sub-scale	Framework components	Sample questionnaire item
Conceptual	Submicroscopic (3 items)	How would including the attraction among particles change the RATE or FREQUENCY at which a single particle reaches the wall? ○ The rate would not change. ○ The rate would decrease. ○ The rate would increase.
	Macroscopic (3 items)	The following diagram shows a piston in a sealed cylinder. In (b) the piston has been pushed in. No air entered or left the cylinder. Let us assume that no energy was added or removed and that the temperature is constant.  The pressure is ○ the same ○ larger in (a) ○ larger in (b)

(*continued on next page*)

Table 2 (*continued*)

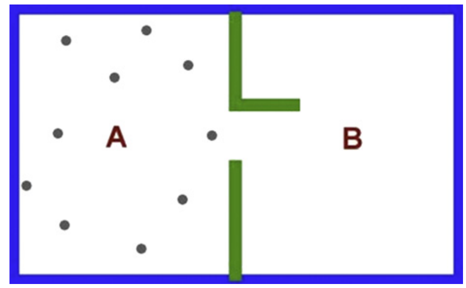
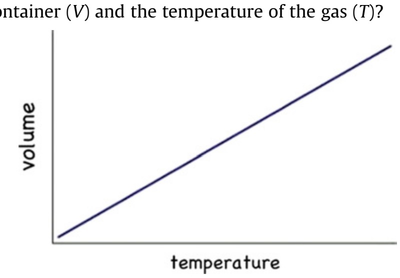
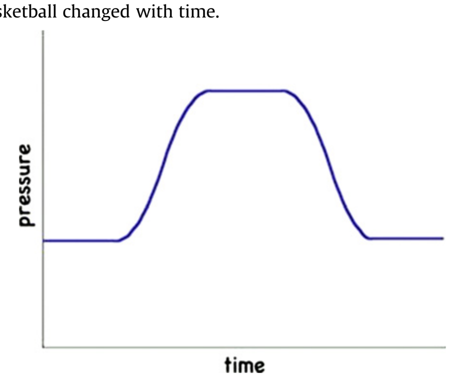
Conceptual/mathematical sub-scale	Framework components	Sample questionnaire item
	Submicroscopic/ macroscopic (8 items)	Imagine a box with a wall inside it as in the following picture. One side of the box [A] contains a gas. A window is then opened in the wall that separates the two parts of the box.  If the gas in the box is heated, will that change what happens to the gas when the window is opened? (C) ○ If the gas is heated, the particles move faster, and hit the wall more often. However, this doesn't make a difference in how they go through the window. Therefore, the gas will go through the window at the same rate. ○ If the gas is heated, the particles become larger, making it more difficult for them to go through the window. Therefore, the gas will go through the window more slowly. ○ If the gas is heated, the particles become larger, making it more difficult for them to go through the window. Therefore, the gas will go through the window more slowly. ○ If the gas is heated, the particles move faster, hit the wall more often, and have a better chance of reaching the window and going through. Therefore, the gas will go through the window more quickly. ○ If they gas is heated, the particles will rise to the top of the box, and will not reach the window. Therefore, the gas will go through the window more slowly.
Mathematical	Symbolic (1 item)	How would you describe the relationship between the volume of the container (V) and the temperature of the gas (T)?  V describes the volume of the container.T described the temperature of the gas.k is a constant number. ○ $V \times T = k$ ○ $V \times T2 = k$ ○ $V = k \times T2$ ○ $V = k \times T$
	Conceptual/ mathematical model (4 items)	The following graph shows how the air pressure inside the basketball changed with time.  Which of the following could have caused the pressure to change in this way? (D)oThe basketball was pumped up with air.

Table 2 (*continued*)

Conceptual/mathematical sub-scale	Framework components	Sample questionnaire item
		○ The basketball had a leak (a hole which lets air come out), the leak was then fixed. ○ The basketball was placed in a very cold container and then taken out until it warmed up. ○ Someone stepped on the basketball, squishing it and making it smaller, then got off, so that it returned to its original shape.

The students' actions in exploring the models, click-type interactions with the models (pressing buttons such as "setup" and "go", clicking on the interface to change a container's volume) and answers to open and closed questions were logged through the Pedagogica environment (Horwitz & Christie, 1999), saved on a server and made available to the researchers. Specific to this study, students' entries into a table of data they have collected from a model in a single screen in three activities are analyzed, as described hence.

2.4. Data analysis

The content knowledge pre- and post-test questionnaires' responses were coded as correct or incorrect and a total score was averaged. The pre-test and post-test results, overall and using the two subscales, were analyzed with descriptive statistics (mean and standard deviation), pre- and post-test scores were compared using a paired samples *t*-test and Cohen's effect size (Cohen, 1988) was calculated.

Students' data collection strategies were analyzed in the following way. The above-described sequences of values the students entered while exploring the model were captured. The first complete series of five values they entered into the table for the manipulated variable (*N* for the *NP* relationship; *T* for the *TP* relationship; *V* for the *VP* relationship). Expert knowledge was used to label one of four patterns, which were mapped as describing a wide variety of monotonic, canonical and familiar mathematical functions: the intervals between values of the manipulated variables were converted into one of four patterns: constant, increasing, decreasing and mixed intervals (Fig. 5). The first and second derivatives of these sequences of values for the manipulated variable 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). The four interval patterns are described in short form as "constant", "increasing", "decreasing" and "mixed". Descriptive statistics – distribution and mode – represent these results.

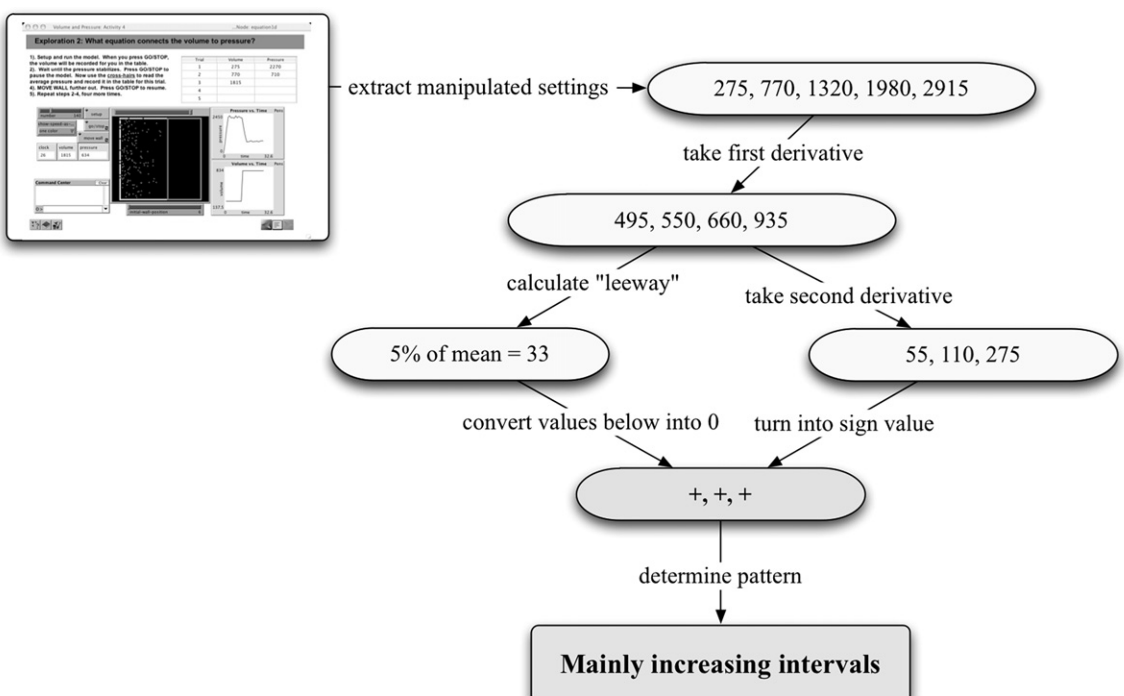


Fig. 5. Method of extracting students' exploration strategies from their table data entries.

These patterns were further coded as "fit" or "unfit" with respect to the models' behavior. Regarding the *linear NP* and *TP* relationship, a fit exploration strategy is "constant intervals" as it covers the parameter space systematically; for the *inverse VP* relationship, it is "increasing intervals" as it captures the faster change in pressure for lower volumes (Fig. 1). A cumulative fitness score was computed as the number of fit explorations (ranging 0–3 for the three explored models).

Consistency between the exploration strategies was computed using the Chi-square test, both for the actual patterns (constant, increasing, decreasing, mixed intervals) and with respect to their fitness to the models' behavior (fit, unfit).

These strategies were related to prior knowledge (overall and subscales) using a logistic regression for the individual explorations' fitness and ANOVA for the cumulative fitness score. It is noted that the subscales used as independent variables may not be completely orthogonal. For example, understanding more about the macro-level behaviors of the system may be related to knowing more about molecular interactions.

Learning gains were calculated as the proportion of gained knowledge with respect to prior knowledge in the following way: ((post-test score) – (pre-test score))/(pre-test score). The exploration strategies' fitness (individually and cumulative) was related to the learning gains using independent samples t-tests.

3. Results

3.1. Model exploration strategies

Students' explorations of the gas models are described in terms of the intervals between the values of the independent variable (constant, increasing, decreasing, mixed) and are depicted in Fig. 6. Distinct distributions are observed for the different models. For the NP

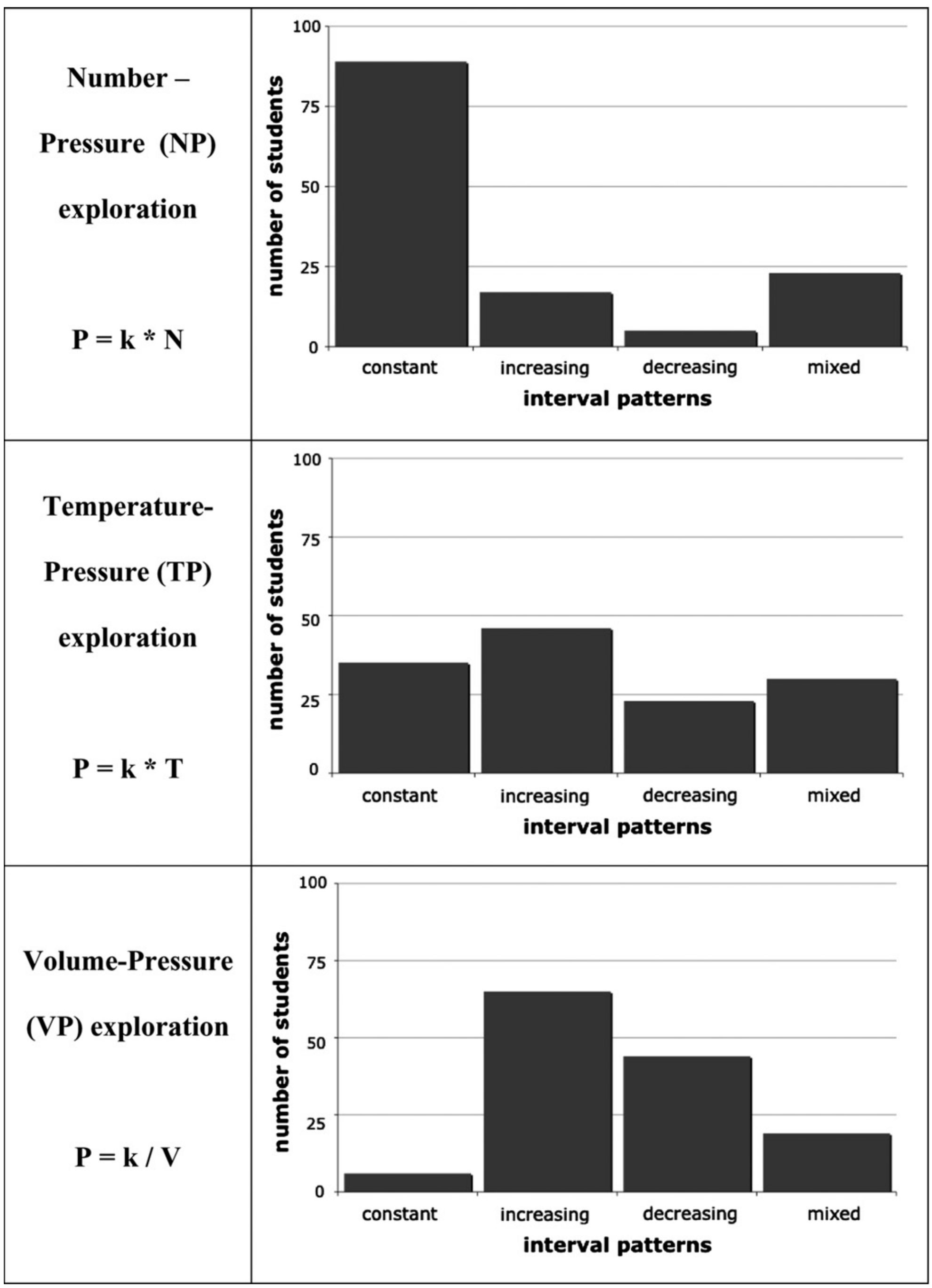


Fig. 6. Students' model exploration strategies. Corresponding gas law equations are on the left: k is a constant; P is pressure; N is the number of gas particles in a container; V is its volume; T is the temperature of the gas. $n = 134$.

model, a clear mode is seen – the "constant intervals" exploration strategy (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" strategy (47%), however the "decreasing intervals" strategy soon follows.

When the strategies are recoded for fitness with respect to the models' mathematical behavior, we can see that the students are using mainly fit strategies for the *NP* and *VP* exploration, however no clear result is seen for *TP*. The cumulative number of fit explorations is $M = 1.41$, $SD = 0.851$, about half of the explorations.

Students' consistency in exploring the different models was tested by looking at (1) the actual strategies; and, at (2) their fitness with respect to the model's mathematical behavior. 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* ($\chi2(1) = 0.27$, *ns*), or between *VP* and *TP* ($\chi^2(1) = 0.15$, *ns*). A significant relationship is found between *NP* and *VP* ($\chi^2(1) = 4.55$, $p = .033$). In exploring both of these two relationships, 37% of the students used fit strategies.

3.2. Prior knowledge and model exploration strategies

Students' pre-test and post-test results (Table 2) show greater learning gains for the submicro- and submicro-to-macro subscales and for conceptual understanding (Table 3).

Table 3
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		Paired *t*	Effect sizeCohen's *d*(95% CI).
	Pre	Post		
	M (*SD*)	*M* (*SD*)		
All (19)	56 (17)	66 (19)	−17.61**	0.55 (0.46–0.65)
Form of access (7)				
Submicro (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)
Symbolic (1)	42 (49)	58 (49)	−8.21**	0.33 (0.23–0.42)
Bridge (12)				
Submicro/Macro (8)	56 (21)	65 (22)	−11.72**	0.42 (0.33–0.51)
Conceptual/Symbolic models (4)	56 (28)	62 (29)	−6.83**	0.21 (0.12–0.21)
Mathematical (5)	56 (29)	65 (30)	−7.37**	0.31 (−1.62–2.16)
Conceptual (14)	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 strategies' fitness, and ANOVA for the cumulative fitness are presented in Table 4. These results show that only some of the components of knowledge impact the students' model exploration strategies, and in different ways for the different relationships: NP exploration is impacted by conceptual but not mathematical understanding; and more specifically by prior knowledge of both the submicro- 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 submicro-level alone.

Table 4
Logistic regression testing association between prior knowledge and the model exploration strategies, and ANOVA of their cumulative fitness score.

Prior knowledge component	NP exploration strategy (χ^2)	TP exploration strategy (χ^2)	VP exploration strategy (χ^2)	All exploration strategies *F*(1,83)
All	6.36**	0.02	0.19	0.584
Submicro	3.50*	0.20	5.43*	2.324
Macro	3.51*	3.11*	0.10	0.580
Symbolic	0.36	0.14	0.03	0.147
Submicro/Macro	0.36	0.40	0.05	0.213
Symbolic/Conceptual model	8.53**	0.15	0.32	1.316
Conceptual	7.67**	0.10	0.11	0.329
Mathematical	1.98	0.01	0.15	0.018

* $p < .05$.
** $p < .01$.

3.3. Model exploration strategies and learning

Independent samples *t*-tests relate the explorations' fitness and learning gains (Table 5). Small associations between exploration strategies and learning are found. Fitness in NP exploration (and more weakly for VP exploration) is associated with learning more advanced complexity reasoning: bridging of submicro/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 5
Association between model exploration strategies and learning gains.

Knowledge component	*NP* exploration strategy (*t*-test[a])	*TP* exploration strategy (*t*-test)	*VP* exploration strategy (*t*-test)	All explorations $F(3,n-1)$
All	1.76 (244)[b]	1.01 (222)	0.55 (264)	2.64 (81)
Submicro	−0.911 (241)	−1.063 (178)	0.977 (212)	0.349 (66)
Macro	0.611 (232)	1.429 (212)	0.198 (258)	0.418 (78)
Submicro/macro	2.08* (258)	0.112 (225)	1.82 (280)	1.516 (81)
Conceptual/symbolic model	−1.497 (227)	0.824 (202)	−0.022 (254)	0.286 (76)
Conceptual	0.116 (249)	1.249 (230)	−1.414 (268)	1.987 (85)
Mathematical	−1.728 (222)	0.506 (201)	1.010 (246)	0.578 (73)

*$p < .05$.
[a] Independent samples *t*-test.
[b] *n*'s are in parentheses.

4. Discussion

The present study addresses two issues related to learners' interactions with computer models of complex systems. One involves advancing the field of educational data mining by developing a method that associates inquiry actions with understanding. The second concerns the relationship between conceptual and mathematical knowledge in a setting where students construct mathematical representations of complex chemical systems. Students interacted with three models and extracted data from them regarding two system-wide variables. Their goal was to construct an algebraic equation that connects the manipulated and observed variables. Among the three equations, two involved linear relationships and one involved an inverse relationship. Their choices in extracting such data were investigated and related to prior knowledge and learning. In the following discussion, we address the research questions, interpret the study's results, extend the literature on learning about complex systems and offer implications for educational data mining and the design of learning environments.

4.1. Research question 1: what strategies describe how students explore computer models when collecting data with which they will construct an equation relating the system's variables?

This first question focuses on describing patterns in students' data collection strategies within unguided exploratory environments. We had considered three possibilities. One is that the students would employ mainly a commonly taught strategy in science laboratories: vary the independent variable at even increments. A second option is that no dominant strategy would evolve, as these interactions are unguided and may be randomly distributed. A third possibility is that the students would space the data they collect so that it would capture the functional change in the dependent variable and support extracting the mathematical equation. Overall, about half of the students' explorations are well-adapted to the model's mathematical behavior. More fine-tuned, it was found that the three models elicit distinct strategy distributions. For two of the three models – one involving a linear relationship (*NP*), and one an inverse relationship (*VP*) – the mode strategy is one most fit to capture the full range of change to the dependent variable. For the first, an "even increments" strategy was most commonly used (two-thirds of the students). As for the second model, many students (about half) displayed an "increasing increments" strategy. However, for the third model that involved a linear relationship (*TP*), no dominant pattern is observed.

Thus the first important finding is that *students' unguided exploration of models with the goal of constructing an equation is well-adapted to portraying change for two out of the three models*. Rather than randomly manipulating the model or using the commonly taught laboratory strategy, in half the cases, students space the data they collected to support obtaining a scatter-plot that would provide the span and density to help them conclude regarding the associated mathematical representation.

Why are *the explorations of the first model more fit than the second*? In the first model varying the number of particles in the container, the relationship between the two variables is linear. An inverse function describes the relationship between the variables in the second, volume-varying model. Inverse relationships among physical entities prove more challenging in a variety of domains (e.g. Bar, 1987; Nemirovsky, 1994; Stavy & Tirosh, 1999). Given many students' greater difficulties with inverse relationships in comparison to linear relationships, fewer students adapted their explorations to the second model's mathematical behavior with respect to the first.

Why are *the results different for the third model, where no dominant exploration strategy is observed*? All three models are used to explore the effect of one variable (number of particles, volume and temperature) on the pressure exerted by gas particles inside a container. We offer two interpretations for the "no pattern" result regarding inquiry actions in the temperature-varying model. One relates to students' greater difficulties in understanding temperature with respect to the other two variables. The other focuses on the model itself and greater challenges it may present in concluding about how the variables relate to each other.

People's understanding of temperature has been explored in the domain of science education. Many high-school students tend to confuse temperature and heat energy (Erickson, 2006; Tiberghien, 1983; Wiser & Carey, 1983), conceive of heat as a substance (Reiner, Slotta, Chi, & Resnick, 2000), view heat and cold as separate entities rather than a continuum (Engel Clough & Driver, 1985), believe that temperature is an inherent property of materials and differs among different materials and find the idea of equilibration difficult to understand (Driver, Squires, Rushworth, & Wood-Robinson, 1994). Volume and the number of particles, the independent variables in the first two models, are quite straightforward and can be immediately perceived – volume as a bigger and smaller container, number of particles as more or less dots on the screen – and no reports in the literature have shown students' difficulties in understanding them. Different from these two variables, the widespread difficulties in understanding temperature may provide the basis for confusions regarding the mathematical relationship based on understanding this very concept.

A second interpretation focuses on the model itself. As previously described (Table 1), the changes to the model following manipulation are distinct. For the first two models – adding particles and increasing the volume of a container – once the student makes the change, its effect is almost immediate. The student needs to focus only on the resultant changes to the system. However, in the third model, manipulating temperature requires attending to a more gradual process of equilibration. The container's walls are first heated (or cooled); then, moving particles may hit the wall and increase (or decrease) their speed; finally, through collisions with other particles – the energy is redistributed until a stable temperature is reached. The challenge in observing change to the manipulated variable is exacerbated by the fact that these are not three distinct steps, but co-occurring interactions that produce the change. Only following these three concurrent and mutually dependent processes, does the pressure in the container equilibrate and the resultant changes can be observed. Thus, with respect to the other two models, this model presents a more difficult task in relating causes and effects in the system.

Given initial difficulties in understanding temperature and the greater challenge of following its change in the computer model it would be reasonable that students' strategies in making sense (both conceptual and mathematical) would be more varied and less focused than for the relatively simpler relationships in the other models.

4.2. Research question 2: how consistent are the strategies students use in exploring the computer models in terms of fitness to the models' mathematical behavior?

Students' consistency in adapting their data collection strategies to the model's mathematical behavior was examined. Consistency was found, with a third of the students using well-adapted strategies in both the particle number (*N*)- and volume (*V*)-varying models.

This points once again to the mindful way in which many students explored the models. *Students that explored one model adaptively explored the other one adaptively as well*. Not only are many students adapting their explorations to the model's mathematical behavior, but they are also doing this consistently. Thus, examining the quality of students' adaptation to the model's' behavior forms two groups: those that conduct mathematically-astute explorations and those that do not. It would seem that the method developed in this study to portray students' inquiry actions is useful as a data mining metric. *It taps onto a significant difference among students in the strategies they employ to interact with models when the goal is to quantify relationships*. In the next section, this argument is continued as this distinction is shown to reflect an understanding of the phenomena explored.

Regarding the lack of consistency in exploration of the temperature-varying model, the same difficulties described above (Section 4.1) would counter consistency in students' adaptation to the models.

4.3. Research question 3: how does prior knowledge impact students' model exploration strategies?

This question turns to the possible impact of prior knowledge upon how the models are explored, asking whether fit explorations may be related to greater prior knowledge. Previous research has shown that learners with greater domain knowledge select more sophisticated strategies in solving problems (Alexander & Judy, 1988), particularly in exploring models and simulations (Gredler, 1996; Hmelo, Nagarajan, & Roger, 2000; Lazonder, Wilhelm, & Hagemans, 2008; Njoo & de Jong, 1993; Schauble, Glaser, Raghavan, & Reiner, 1991). In exploring submicroscopic and macroscopic graphics of cell transport, it was found that students with high prior knowledge tended to focus on the submicroscopic representations, those more critical to understanding the phenomena depicted; while students with low prior knowledge observed the macro-level phenomena and transitioned among them; moved frequently between these and the submicroscopic representations and expressed more difficulty in coordinating these representations (Cook, Wiebe, & Carter, 2008). Liu, Andre, and Greenbowe (2008) have found that students with higher level prior knowledge tend to use chemistry computer simulations mainly in a confirmatory mode after working out the problems with equations and formulas; while students with lower level prior knowledge tend to use the simulations as a main resource to accomplish their tasks. Thus one would expect that students expressing greater prior knowledge would perform a more efficient search for information in the simulations.

What has been found in this study supports the above-described research and extends it: *prior knowledge impacts interactions with the model, however the particular qualities of such knowledge are distinct among the relationships explored*. Thus, we offer a more discriminate look into the kind of knowledge that guides explorations. It was found that only for one of the models, a general prior knowledge impacts the strategies with which the models are explored – that varying the number of particles in the container (*NP*). It is interesting to note that conceptual knowledge rather than mathematical knowledge are associated with the more fit explorations. *In some cases, while the goal of the interaction was to extract a quantitative representation of the model's behavior, conceptual and not mathematical understanding of the domain guides the interaction itself*. This finding is consistent with previous research that highlights the greater importance of conceptual knowledge in understanding science and particularly chemistry, the domain of this study (Nakhleh, 1993; Niaz & Robinson, 1992). However, different from these studies, it shows that when "quantitative" problems are posed in a way that requires active exploration in relatively open environments, many students demonstrate a well-adapted connection between their conceptual understanding and their quantitative problem solving. This points to the value of such challenging activities that engage students with creating mathematical representations and not just applying externally provided equations. Previous studies into activities that engage students in constructing mathematical representations of scientific phenomena are sparse and target more advanced undergraduate students (Bopegedera, 2007; Laugier & Garai, 2007). In this study, we show that not only do students succeed at such an advanced activity of creating algebraic representations; they also harness their conceptual understandings in the process and connect the two forms of understanding.

When shifting to a more discriminate examination of the kinds of conceptual knowledge that support more informative explorations of the models, a complex systems perspective is introduced. Items in the questionnaires were designed to address an understanding of gas particles at the submicroscopic level, phenomena at the macroscopic level, mathematical representations and bridges between these forms of describing the system. It was found that *among the three models, students' exploration strategies were impacted by distinct components of prior conceptual understanding of systems*. In the model varying the number of particles (*NP*), three components were associated with fit explorations: prior understanding of the submicroscopic level, the macroscopic level and bridging the latter with its mathematical representation. More adaptive explorations of the temperature-varying model (*TP*) were associated with a greater prior understanding of the macro-level phenomena. Fit explorations of the volume-varying model (*VP*) were associated with a greater understanding of the submicro-level particles' behaviors.

Interpretation of these results is tentative. There is not enough evidence in the study to corroborate and explain these findings. Further studies such as a more detailed observation of knowledge, learning and interaction with such models could inform such an interpretation in a better way. The following explanations are based on comparing the three models with respect to two main features: the conceptual and the mathematical challenges they present.

It was found that explorations of the model varying the number of particles were more tightly related to prior knowledge. With respect to the other two models, this relationship is much easier, in that the scientific concepts are non-problematic and the mathematical relationship is linear. More students explored this model efficiently and the kinds of conceptual understandings related to this efficiency are more varied and relate to critical components of understanding systems: understanding of the two description levels and mapping the conceptual model onto the mathematical one. Given the non-problematic aspect of this model, its exploration provides a clearer result regarding what it means to understand the system. In fact, the one system-related component that does not impact the model's exploration – relating the submicro- and macro- levels – is associated with learning by those that conducted efficient model explorations.

As described above, the concept of temperature is more challenging than the other concepts engaged with in the models. In terms of the mathematical relationship – a linear one – it is easier. Thus, it seems that rather than understanding the mechanism at the particles' level, a more global understanding supports well-adapted explorations. It is possible that the conceptual challenge is too great for most students.

Efficient explorations of the volume-varying model were supported by an understanding of the submicroscopic particles level. This model presents a greater mathematical challenge, an inverse relationship; however, it is less problematic in terms of conceptual difficulty. In this case, students were more apt to resolve how they relate to quantitative symbolic expressions by harnessing their understanding of the causal submicroscopic level.

To summarize this section, prior knowledge does indeed impact model exploration when the goal is to quantify relationships; however different components of this knowledge are important for the different models. Relationships between conceptual understanding and model exploration are more tightly coupled when the model is less problematic both in terms of conceptual and mathematical understandings. When the conceptual model is less accessible, macro-level associations elicit more fit explorations. When the mathematical structure is more challenging, the deeper conceptual level is of impact on fit explorations. These findings suggest a more discriminate view of how prior knowledge and strategies interact, when these are compounded with additional challenges – conceptual or mathematical. As a data mining metric, not only does the form of interaction with the models signify the efficiency of collecting data, but it also reflects prior knowledge.

4.4. Research question 4: what associations can be found between students' exploration strategies and their learning gains regarding the related content?

Different from prior knowledge, *learning gains are only weakly associated with how the models are explored*. The one clear significant association is between more adaptive explorations of the model varying the number of particles, and learning to bridge the submicro- and macro- levels, a critical component in reasoning about complex systems. This shows that when the model is relatively unproblematic in terms of conceptual and mathematical understandings, more efficient searches for information evolve into deeper understandings of the system as complex.

It is important to note the weak associations with learning. One could consider an educational application such as teaching the students how to explore models more efficiently. However, the weak associations between these strategies and learning preclude such an approach. In the study, students were free to explore the model in any form they choose. Their ways of exploration were impacted by prior knowledge but didn't relate to learning gains. Therefore, these strategies, while indicative of understanding, should not necessarily be externally structured. Contrary to this, it is proposed that supporting the development of conceptual knowledge would be of more value in advancing students' learning, an issue taken up in more detail in the following sections of the discussion. This study has found the association between conceptual knowledge and fit exploration strategies to be stronger, even when the mathematical relationship is challenging. Upon detection of a students' less informative model investigation, introducing interventions geared at gaining a greater understanding of the central concepts would be more beneficial towards advancing learning of the mathematical representations of the system.

4.5. General discussion

The discussion turns to the broader issues that have guided this study: the importance of understanding systems as complex, the conceptual basis of mathematical understandings of systems, educational data mining in exploratory environments and applications to educational settings.

4.5.1. Conceptual basis of a mathematical understanding of systems

One of problems raised in the introduction is the disconnect between conceptual and mathematical understandings of chemical systems in typical problem solving situations. Previous research has found that many students can solve quantitative equation-based problems and yet fail to comprehend the related conceptual model. In this study, we have seen that when quantitative problem solving is set in a more constructive setting – having the students create the mathematical representations rather than only use them – a stronger linkage is found between students' conceptual model of the domain and their approach towards constructing a mathematical model. In fact, it is mainly the conceptual and not the mathematical knowledge that guides such activity. Adding to our previous findings that many students succeeded in constructing such equations (Levy & Wilensky, 2009b), it would seem that this activity is certainly beneficial and possibly applicable to additional settings of learning, where conceptual and mathematical understandings meet and interact.

4.5.2. Importance of understanding systems as complex

One central conclusion touches upon the importance of understanding systems as complex even when the goal is to quantify the relationship between global properties. For two out of the three models, understanding the submicroscopic particles' behaviors and interactions was essential to making the search for information in the model pertinent to its macro-level mathematical behavior.

The proclaimed goal of exploring the model was to relate between two global variables, such as volume and pressure. In terms of efficiency, it would seem that understanding the local behaviors of particles and their interactions is unrelated to obtaining this goal – one could reason about the system in terms of its general properties, associate them and attain the relationship among them. Yet results of this study show that understanding local behaviors is critical in the search for global information. Given the above-described centrality of conceptual understanding to strategic explorations of models and simulations (Section 4.3), it would seem that the *kind* of conceptual knowledge that is critical to understanding is that connecting the submicroscopic level particle behaviors and interactions to the emergent system-wide macro-level. Such reasoning makes up the heart of understanding complex systems (Wilensky & Resnick, 1999). In fact, in creating the Connected Chemistry curriculum, our contention was that in the articulation of the submicroscopic behaviors, noticing and expressing in specific and causal terms interactions at the molecular level, our view gradually expands to include local interactions and emergent behaviors (Levy & Wilensky, 2009a). We have seen this kind of reasoning arise spontaneously among sixth-grade students as they explained a social emergent phenomenon – the scattering of classmates in gym class as they prepare for calisthenics (Levy & Wilensky, 2008). Some of the students well described how local interactions and negotiations (e.g. making sure you do not hit each other, move away if there is someone there) emerged into patterns in small groups (e.g. clustering). These very same students also had a deeper understanding of the system as complex, expressing ideas regarding the stochastic nature of the individuals' behaviors and equilibration processes in the system. Blikstein and Wilensky (2009) have restructured an undergraduate level materials science course, in which a multitude of equations is replaced with having the students construct agent-based models based on a small set of simple rules. Their research has shown that students who are engaged with such a curriculum develop a deeper understanding of the central concepts in the domain, and are able to identify unifying principles and behaviors across phenomena in the domain. The results of this study lend additional support to the idea that a complex systems perspective provides accessible and deeper understanding of systems.

4.5.3. Educational data mining in exploratory environments

The developing field of educational data mining seeks ways to extract information about people's knowledge and learning in computational environments. This study provides a contribution in the less researched domain of *exploratory environments*. The students were free to choose their style of exploration. It was found that this style was indicative of an understanding of the system explored. Moreover, students were consistent in whether or not they adapted their exploration to the model's mathematical behavior. These findings suggest that this method of looking at the increments students use in extracting quantitative information from a model provides a metric for knowing-in-action, tapping onto knowledge as expressed in problem solving strategies in an inquiry environment. It can be applied to any learning environments in which models are explored to obtain mathematical relationships. Its application is also limited to models, where students would have fewer difficulties with the scientific concepts behind the manipulated variables. In such inquiry environments, mining students choices of data can suggest interventions to help improve learning. Based on the results of this study, it is suggested that content-based interventions, geared at a deeper conceptual understanding would be of greater value than one that is procedure-based, on how to conduct strategic explorations. For example, if a student is using a non-adaptive strategy, it may be useful to suggest some additional conceptual support in obtaining a deep understanding of the basic scientific properties and concepts, the submicro- and macro- levels of the system and how they relate.

4.6. Limitations of the study

While this study proves useful in developing our understandings of students' mathematical and conceptual knowledge of systems and how this relates to their exploration, it is also limited in a number of ways. One limitation stems from the way data mining was undertaken. Using expert knowledge to presume what may be indicative of knowledge limits the possibilities to those that have been thought of. Possibly a better approach would be to use a much larger set of variables related to the students activity and cluster the data to form the labels on students' behavior. Fewer assumptions may suggest a wider array of information and discovery.

4.7. Future research

Future research should address the differences among the models in the kind of prior knowledge that impacts inquiry. We have seen that both conceptual and mathematical challenges may be related to differences in students' strategies in exploring the models. This could be tested more systematically in more settings, varying such features systematically. In addition, replication studies using the method with additional models and systems would help understand how general the findings are and whether they can be applied across different models and content domains. Finally, based on the method developed in this study for examining students' mathematical understandings of complex systems, the efficacy of adaptive supports in promoting conceptual understanding when model explorations are found less fit may be tested within computational model-based environments.

5. Conclusions

This study has investigated students' exploratory actions with computer-based multi-agent models when their goal is to construct an equation. More specifically – the data they collected and its spacing was used to determine their expectations regarding the model's mathematical behavior. It was found that conceptual knowledge of the fundamental causes underlying the system's behavior impacts these strategies. This suggests that engaging students in constructing symbolic representations may provide a bridge between frequently disconnected conceptual and mathematical forms of knowledge.

This study contributes towards the development of methods in educational data mining that may support detection of knowledge-in-action and suggests ways that may be more beneficial to support deeper learning. In designing for learning about complex systems through exploring models, the data mining method presented in this study can be used to capture students' exploration strategies. In this way, their conceptual understanding can be tapped into via action-based forms of expression, complementing detection of verbally expressed understanding. Using such fine-tuned means of detecting understanding may be used to support appropriate interventions for further learning. Moreover, appropriate supports suggested by this study are conceptual, rather than strategic, particularly based upon a complexity perspective – understanding simple micro rule-based behaviors as generating global patterns.

Acknowledgements

Foremost, we thank Reuven Lerner who has provided important assistance in creating and programming the data mining tools. We thank Shira Yuval for working with us on the statistical analysis. We deeply thank Michael Novak, the lead curriculum developer who collaborated with us in designing and forming the curriculum. Modeling Across the Curriculum is funded by the Interagency Education Research Initiative (IERI), a jointly supported project of the National Science Foundation, the US Department of Education and the National Institute of Child Health and Human Development, under NSF Grant No. REC-0115699. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies. This paper continues and expands the authors' AERA 2010 paper with the same title.

Appendix. Supplementary data

Supplementary data associated with article can be found in online version at doi:10.1016/j.compedu.2010.09.015.

References

Alexander, P. A., & Judy, J. E. (1988). The interaction of domain-specific and strategic knowledge in academic performance. *Review of Educational Research, 58*, 375–404.
Amershi, S., & Conati, C. (2009). Combining unsupervised and supervised classification to build user models for exploratory learning environments. *Journal of Educational Data Mining, 1*(1), 18–71.
Ardac, D., & 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.
Baker, R. S. J. D. (2010). Data mining for education. In B. McGaw., E. Baker & P. Peterson, (Eds.), International encyclopedia of education (3rd ed.). Oxford, UK: Elsevier.
Baker, R. S. J. D., & Yacef, K. (2009). The state of educational data mining in 2009: a review and future visions. *Journal of Educational Data Mining, 1*(1), 3–17.
Bar, V. (1987). Comparison of the development of ratio concepts in two domains. *Science Education, 71*(4), 599–613.
Blikstein, P., & Wilensky, U. (2009). An atom is known by the company it keeps: a constructionist learning environment for materials science using multi-agent simulation. *International Journal of Computers for Mathematical Learning, 14*(1), 81–119.
Bopegedera, A. M. R. P. (2007). An inquiry-based chemistry laboratory promoting student discovery of gas laws. *Journal of Chemical Education, 84*(3), 465–468.
Bruner, J. S. (1973). Organization of early skilled action. *Child Development, 44*, 1–11.
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.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New Jersey: Lawrence Erlbaum.
Cook, M., Wiebe, E. N., & Carter, G. (2008). The influence of prior knowledge on viewing and interpreting graphics with macroscopic and molecular representations. *Science Education, 92*, 848–867.
Costu, B. (2007). Comparison of students' performance on algorithmic, conceptual and graphical chemistry gas problems. *Journal of Science Education and Technology, 16*, 379–386.
Driver, R. D., Squires, A., Rushworth, P., & Wood-Robinson, V. (1994). *Making sense of secondary science: Research into children's ideas*. London/New York: RoutledgeFalmer.
Engel Clough, E., & Driver, R. (1985). Secondary students' conceptions of the conduction of heat: bringing together scientific and personal views. *Physics Education, 20*, 176–182.
Erickson, G. (2006). Children's conceptions of heat and temperature. *Science Education, 63*(2), 221–230.
Gabel, D. L., Sherwood, R. D., & Enochs, L. (1984). Problem-solving skills of chemistry education students. *Journal of Research in Science Teaching, 21*(2), 221–233.
Gibson, E. J. (1988). Exploratory behavior in the development of perceiving, acting and the acquiring of knowledge. *Annual Review of Psychology, 39*, 1–41.
Gobert, J., Horwitz, P., Tinker, B., Buckley, B., Wilensky, U., & Levy, S.T. (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*.
Gobert, J. D., O'Dwyer, L., Horwitz, P., Buckley, B. C., Levy, S. T. & Wilensky, U. (2010). Examining the Relationship Between Students' Understanding of the Nature of Models and Conceptual Learning in Biology, Physics, and Chemistry. *International Journal of Science Education*, doi:10.1080/09500691003720671.
Gredler, M. E. (1996). Educational games and simulations: a technology in search of a (research) paradigm. In D. H. Jonassen (Ed.), *Handbook of research for educational communications and technology* (pp. 521–540). New York: MacMillan.
Hershkovitz, A., & Nachmias, R. (2009). Learning about online learning processes and students' motivation through web usage mining. *Interdisciplinary Journal of E-Learning and Learning Objects, 5*, 197–215, Special series of Chais Conference 2009 best papers.
Hmelo, C. E., Nagarajan, A., & Roger, S. (2000). Effects of high and low prior knowledge on construction of a joint problem space. *Journal of Experimental Education, 69*, 36–56.
Holland, J. H. (1995). *Hidden order: How adaptation builds complexity*. Cambridge, MA: Helix Books/Addison-Wesley.
Horwitz, P., & 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 M. J. Jacobson, & R. B. Kozma (Eds.), *Innovations in science and mathematics education: Advanced designs for technologies of learning* (pp. 11–46). New Jersey, London: Lawrence Erlbaum.
Laugier, A., & Garai, J. (2007). Derivation of the ideal gas law. *Journal of Chemical Education, 84*(11), 1832.
Lazonder, A. W., Wilhelm, P., & Hagemans, M. G. (2008). The influence of domain knowledge on strategy use during simulations-based inquiry learning. *Learning and Instruction, 18*, 580–592.
Levy, S. T., Novak, M., & Wilensky, U. (2006). *Connected Chemistry Curriculum*. Evanston, IL. Center for Connected Learning and Computer Based Modeling, Northwestern University. http://ccl.northwestern.edu/curriculum/chemistry/. Download at http://mac.concord.org/downloads.
Levy, S. T., & Wilensky, U. (2008). Inventing a "Mid-level" to make ends meet: reasoning between the levels of complexity. *Cognition and Instruction, 26*(1), 1–47.
Levy, S. T., & Wilensky, U. (2009a). Crossing levels and representations: the Connected Chemistry (CC1) curriculum. *Journal of Science Education and Technology, 18*(3), 224–242.
Levy, S. T., & Wilensky, U. (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.
Liu, H.-C., Andre, T., & Greenbowe, T. (2008). The impact of learner's prior knowledge on their use of chemistry computer simulations. *Journal of Science Education and Technology, 17*, 466–482.
Mason, D. S., Shell, D. F., & Crawley, F. E. (1997). Differences in problem solving by nonscience majors in introductory chemistry on paired algorithmic-conceptual problems. *Journal of Research in Science Teaching, 34*(9), 905–923.
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.
Njoo, M., & de Jong, T. (1993). Exploratory learning with a computer simulation for control theory: learning processes and instructional support. *Journal of Research in Science Teaching, 30*(8), 821–844.
Nurrenberg, S., & Pickering, M. (1987). Concept learning versus problem solving: is there a difference? *Journal of Chemical Education, 64*(6), 508–510.
Pahl, C. (2004). Data mining technology for the evaluation of learning content interaction. *International Journal on E-Learning, October–December*, 47–55.
Reiner, M., Slotta, J. D., Chi, M. T. H., & Resnick, L. B. (2000). Naïve physics reasoning: a commitment to substance-based conceptions. *Cognition and Instruction, 18*(1), 1–34.
Russell, J. W., Kozma, R. B., Jones, T., Wykoff, J., Marx, N., & 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.
Sawrey, B. A. (1990). Concept learning versus problem solving revisited. *Journal of Chemistry Education, 67*(3), 253–254.
Schauble, L., Glaser, R., Raghavan, K., & Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *The Journal of the Learning Sciences, 1*, 201–238.

Schön, D. (1983). *The reflective practitioner: How professionals think in action*. New York: Basic Books.
Snir, J., Smith, C. L., & Raz, G. (2003). Linking phenomena with competing underlying models: a software tool for introducing students to the particulate model. *Science Education, 87*, 794–830.
Stavy, R., & Tirosh, D. (1999). Intuitive rules: a way to explain and predict students' reasoning. *Educational Studies in Mathematics, 38*, 51–66.
Stieff, M., & Wilensky, U. (2003). Connected chemistry – incorporating interactive simulations into the chemistry classroom. *Journal of Science Education and Technology, 12*(3), 285–302.
Tiberghien, A. (1983). Critical review on the research aimed at elucidating the sense that the notions of temperature and heat have for students aged 10 to 16 years. In *Research on physics education. Proceedings of the first international workshop, 26 June–13 July, La Londe les Maures, France, Vol. 1984* (pp. 75–90). Paris: Editions du Centre National de la Recherche Scientifique.
van der Meij, J., & de Jong, T. (2006). Supporting students' learning with multiple representations in a dynamic simulation-based learning environment. *Learning and Instruction, 16*, 199–212.
Wilensky, U. (1999a). *NetLogo*. Evanston, IL: Center for Connected Learning and Computer-Based Modeling, Northwestern University. http://ccl.northwestern.edu/netlogo.
Wilensky, U. (1999b). GasLab: an extensible modeling toolkit for exploring micro- and macro- views of gases. In W. Feurzig, & N. Roberts (Eds.), *Modeling and simulation in science and mathematics education* (pp. 151–178). New York: Springer.
Wilensky, U. (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).
Wilensky, U., & Resnick, M. (1999). Thinking in levels: a dynamic systems perspective to making sense of the world. *Journal of Science Education and Technology, 8*(1), 3–19.
Wiser, M., & Carey, S. (1983). When heat and temperature were one. In D. Gentner, & A. L. Stevens (Eds.), *Mental models* (pp. 267–297). New-Jersey, London: Lawrence Erlbaum Associated Publishers.
Zoller, U., Dori, J., & Lubezky, A. (2002). Algorithmic, LOCS and HOCS (chemistry) exam questions: performance and attitudes of college students. *International Journal of Science Education, 24*(2), 185–203.