“It Must Include Rules”: Middle School Students’ Computational Thinking with Computer Models in Science

10

ELIANE S. WIESE, University of Utah, USA
MARCIA C. LINN, University of California, Berkeley, USA

When middle school students encounter computer models of science phenomenon in science class, how do they think those computer models work? Computer models operationalize real-world behaviors of selected variables, and can simulate interactions between the modeled elements through programmed instructions. This study explores how middle school students think about the high-level semantic meaning of those instructions, which we term *rules*. To investigate this aspect of students’ computational thinking, we developed the Computational Modeling Inventory and administered it to 253 7^{th} grade students. The Inventory included three computer models that students interacted with during the assessment. In our sample, 99% of students identified at least one key rule underlying a model, but only 14% identified all key rules; 65% believed that model rules can contradict; and 98% could not distinguish between emergent patterns and behaviors that directly resulted from model rules. Despite these misconceptions, compared to the “typical” questions about the science content alone, questions about model rules elicited deeper science thinking, with two to ten times more responses including reasoning about scientific mechanisms. These results suggest that incorporating computational thinking instruction into middle school science courses might yield deeper learning and more precise assessments around scientific models.

CCS Concepts: • **Social and professional topics** → **Computational thinking**; **Student assessment**; **K-12 education**;

Additional Key Words and Phrases: middle school science; computer models; computer models in science

ACM Reference Format:
Eliane S. Wiese and Marcia C. Linn. 2021. “It Must Include Rules”: Middle School Students’ Computational Thinking with Computer Models in Science. *ACM Trans. Comput.-Hum. Interact.* 28, 2, Article 10 (April 2021), 43 pages. https://doi.org/10.1145/3415582

1 INTRODUCTION

Designers of educational technology often present complex science ideas through interactive models and simulations (e.g., the University of Colorado’s PhET [1], the Concord Consortium’s Molecular Workbench [5]). Dynamic models of scientific phenomena have gained importance in science classrooms in the United States since the publication of the Next Generation Science Standards [32] and the widespread access to computer technologies in precollege courses [18]. Students in grades 6-12 can and do benefit from instruction featuring dynamic models (e.g.,[4, 29, 52]). State and national assessment items that incorporate dynamic models are being developed [36]. Yet, students

Authors’ addresses: Eliane S. Wiese, University of Utah, 3264 Merrill Engineering Building, 50 Central Campus Dr. Salt Lake City, UT, 84114, USA, eliane.wiese@utah.edu; Marcia C. Linn, University of California, Berkeley, 4611 Tolman Hall, Berkeley, CA, 94720-1670, USA, mclinn@berkeley.edu.

1073-0516/2021/4-ART10 $15.00
https://doi.org/10.1145/3415582

often have limited understanding of the strengths and limitations of these models. Furthermore, assessments of students' learning from models often fail to capture these nuances.

Instruction with dynamic models often asks students to identify the causal relationships in a model. At a high level, this is asking students to identify the key underlying rules that the model is following. We consider a rule to be a logical, natural language statement about the actions of components in the model. A rule is higher-level than programmed code, and is not expressed in a programming language. For example, a simple model of climate might include the rule *When infrared radiation hits a greenhouse gas, the infrared radiation bounces off.* This rule operationalizes the scientific concept that greenhouse gases absorb and re-emit infrared radiation. A model that follows this rule could show an animation with icons representing greenhouse gases, and icons representing infrared radiation bouncing off of them. While exploring the model, students would be expected to notice this kind of pattern of behavior of the objects in the model. To make meaningful, scientific connections between the models and the physical world, students need to reason about these rules. For example, the interaction between greenhouse gases and infrared radiation explains how greenhouse gases contribute to global warming. All of the studies cited above follow this approach in their instructional design.

Interpreting a model as a set of rules requires nuanced computational thinking [54, 60]. In addition to identifying individual rules, the student must consider relationships between the rules, and the difference between a computer model and the physical world. One relationship between rules in a computer model is that they cannot contradict each other. Therefore, the existence of one rule constrains what other rules can also be part of the model. A difference between computer models and the physical world is that the the rules of a computer model may not include all relevant variables (as is the case with all models). Finally, although we can express rules in natural language, they are still particular implementations or operationalizations of scientific concepts. Computer programs cannot directly implement scientific concepts like "absorb and re-emit radiation." Rather, these concepts are operationalized as actions that can be programmed (e.g., showing the movement of shapes across a screen).

Another nuance is the difference between an immediate relationship between two variables and a relationship that relies on a cause-and-effect chain through many variables. For example, a climate model might show an overarching relationship between greenhouse gases and temperature without having a rule that directly links the two. Instead, separate rules would connect temperature to infrared radiation, and then the behavior of the infrared radiation to the presence of greenhouse gases. Typical instruction with computer models in science classes presents a correct model and asks students to interact with it to identify scientifically correct causal mechanisms. Indeed, as evidenced by the success of the systems cited above, students can and do learn overall cause-and-effect relationships this way. However, this approach is less successful in helping students learn the mechanisms underlying those relationships (e.g., [34].)

We report on a needs-finding study with the Computational Modeling Inventory, to examine how students think about the construction of computer models. We investigate whether this focus yields a more nuanced measure of students' science understanding than typical assessments. Unlike typical assessments where the model rules are implicit, our Computational Modeling Inventory addresses model rules directly [57]. The Computational Modeling Inventory examines student understanding of computational models about science, including determining which rules a model follows, and predicting the effect of changing a rule. Focusing on rules rather than behaviors alone can reveal how students think about computer models.

We find that students have difficulty identifying model rules solely from interacting with models. In particular, students struggle to apply the computational thinking skills noted above. Students believe that model rules can contradict. We find that students say that a model includes a rule such as "X always happens" and also "X does not always happen." Students also have difficulty delineating the difference between a computer model and the real-world phenomenon it represents. We find that students say that a model includes a rule even when the rule in question refers to a variable that is not represented in that model. Students also do not see the difference between science concepts and operatonalized rules. We find that students think a computer can implement a science concept directly. Students also do not distinguish between behaviors that follow a single rule and behaviors that emerge due to the interaction of several rules. We find that students assume that salient relationships are enforced by rules that directly link the initial cause and the eventual effect. These struggles matter both because they demonstrate deep misconceptions about the nature of computing, and also because the model rules encapsulate the intermediate scientific mechanisms that students are intended to learn. However, we also found that asking students to imagine the effect of a new, incorrect rule prompted mechanistic reasoning. Asking students to consider a hypothetical (and incorrect) model elicited more mechanistic reasoning than asking students about the science shown in the correct model. Our findings suggest that students may learn more science content with instruction that explicitly supports computational thinking, and that encouraging students to explore incorrect ideas in this context can provide valuable formative assessments.

The design of the Computational Modeling Inventory is based on the Knowledge Integration framework, an instructional framework that elaborates constructivist theory [22]. The Knowledge Integration framework posits that students may have multiple, fragmented, and at times contradictory ideas, and that instruction needs to elicit these ideas so they can be examined, challenged, and made coherent [23]. The Computational Modeling Inventory is designed to elicit students' ideas about computer models, under the assumption that students bring prior conceptions about modeling to their learning environments. Computer models with dynamic visualizations often serve an instructional role of allowing students to test a hypothesis and add new ideas to their repertoire [29]. In these activities with computational models, it is unclear how middle school students are connecting their ideas about modeling and their ideas about the underlying science. The Computational Modeling Inventory illuminates how students are currently integrating their ideas about computer modeling and their ideas about science. The assessment reveals that, consistent with hypotheses in the Knowledge Integration framework, students' ideas about the rules that underlie models are fragmented and contradictory. Student responses also illustrate a drawback of only interacting with models that portray correct science concepts: students may retain incorrect ideas because they cannot directly examine those ideas in a counterfactual model. Results from the Computational Modeling Inventory indicate that computer models in science are worthwhile contexts for instruction on both computational thinking and science.

2 RESEARCH QUESTIONS

(1) Identifying Rules: When students interact with computer models, to what extent are they able to identify the rules a model is following? Specifically, when interacting with our target computer models, do students recognize that:
   - Model rules are not mutually contradictory
   - All relevant variables may not be included
   - Science concepts must be operationalized before they can be implemented in a model
   - Some model behaviors may be emergent patterns that result indirectly from the interaction of several low-level rules (as opposed to the direct result of one rule)

(2) Explaining Mechanisms: How does the depth of students' reasoning differ when responding to our questions that ask about adding hypothetical inaccurate rules to a model versus questions that ask about the science portrayed in the model?

The first research question, Identifying Rules, checks the assumption that students can conceive of model behaviors as actions that follow discrete, logical rules. The second question, Explaining Mechanisms, examines the effectiveness of typical vs. rule-focused questions on revealing students' thinking. The second question also tests the boundaries of students' understanding of both computational thinking and science content, by asking students to consider the effect of one new rule on the behavior of the model overall. Together, these questions could provide a deeper, more comprehensive assessment of students' thinking than either type of question alone.

3 RATIONALE AND PRIOR WORK

The word "model" is used to refer to a variety of constructs in science education, including mental models, physical models, and conceptual models. Unless specified, we use the term "model" to refer to dynamic, animated computer models which show a scientific simulation. *Model-based learning* encourages students to develop conceptual models of science phenomena, generally on paper, and refine their models as they encounter more evidence [26]. Model-based learning with paper models lets students show their ideas, but doesn't allow students to test them, or to see hard-to-anticipate interactions or emergent phenomena. Computer models allow students to test ideas and see how individual mechanisms connect to larger patterns. In a recent review of the model-based learning literature, Louca and Zacharia [26] call for more research on model-based learning with computer models. Like other learning experiences, model-based learning with computer models is shaped by the prior knowledge that students bring [39]. In addition to prior conceptions about the science, students will also bring assumptions and attitudes about technology. These ideas about technology are formed through interactions with systems that are often explicitly designed to hide how the technology works, thus leaving students to make their own conjectures.

3.1 When Technology Design Principles are at Odds with Learning

General human-computer design principles focus on making technology easy to use, which often involves hiding the internal details of how the systems work. Don Norman described this principle over 30 years ago by saying, "In fact, the best computer programs are the ones in which the computer itself 'disappears', in which you work directly on the problem without having to be aware of the computer" [33, p. 80]. Mark Weiser also echoed this idea, saying that "The most profound technologies are those that disappear" [55, p. 94]. This kind of disappearing means that our explicit attention is not on our interactions with the tool, but rather on the goals that the tools help us accomplish. Our attention is not just drawn away from our interactions with the tool, but also from considering the inner workings of the tool. Keeping a user's attention focused on their tasks, in their terms, often results in users forming mental models of a system that are very different from how the system actually works [6]. Today, children are growing up with technology that has both "disappeared" and is also powerful and pervasive.

Technology that does complex things without revealing *how* it does those things provides fertile ground for *animism* - treating a non-living object as if it had life-like qualities, such as desires and goals [27]. Animistic responses such as attributing intentions can be elicited with technology as simple as the Heider & Simmel illusion: an animation of two-dimensional black shapes on a white background [17]. A recent review of extensions of the Heider & Simmel illusion found that these attributed intentions are similar among adults across cultural contexts, and that the animistic response to these shapes seems to emerge by age three or four [43]. Therefore, it is not surprising that children ages 8-11 would ascribe life-like qualities, such as human emotions, to technology such as a robotic dog [56], and even a robotic arm [2]. In turn, designers advocate leveraging humans' animistic responses to develop technology that is more useful, fun, and intuitive to interact with [40].

However, when the goal is for students to learn how technology works, animism is an obstacle. Children are using technology to interact directly with domain objects - characters in a game, tools in a coloring app, atoms in a science simulation. However, they are likely not attending to how those technologies represent those objects - how they are operationalized within the system. For many applications, students do not need to understand the difference between a domain object and its representation in the system. For learning about science and especially about models in science, understanding how objects are represented (that is, modeled), is crucial. Identifying the distinctions between physical objects in the world and the representation of those objects in a computer model is complex. Making these distinctions is likely even harder for students whose prior conceptions about and attitudes toward technology are shaped by assumptions and animism. As a result, students may develop implicit conceptions of how computer models work that inhibit understanding of the rules that actually govern them.

The impact of "dissapearing" technology on how students think about computer science is related to but distinct from the more general differences between technology to support performance versus learning. When the goal is to accomplish a task, it makes sense for the technology to disappear so that the user's focus can stay on the task. However, in educational contexts, the goal is generally to acquire knowledge for accomplishing future tasks, often without the technology. This changes the nature of the interactions. For one, while designers of performance-focused technology try to reduce opportunities for human errors, students learn more from technology that allows them to make more domain-relevant mistakes [25, 28, 30]. Further, even domain-general usability issues that impede performance can sometimes improve learning, such as indirect vs. direct manipulation [15], or a fast vs. slow system response time [12]. In lab studies, when interfaces are harder to use, users respond by planning their steps more carefully, which can help them learn to be better problem-solvers ([12, 15]). Transparent or disappearing interfaces reduce the cognitive load of a task, ideally by reducing *extraneous* load - demands that are unrelated to the task [50]. However, technology cannot reduce *germane* or *intrinsic* loads (demands that are related to learning about the task or inherent in the task itself) [50] without affecting the learning outcomes [3]. The differences in design principles for technology to enhance performance versus domain learning are well-established. Still, in these systems, the technology itself is not an object of learning. In contrast, computer models of scientific phenomena are an object of learning. To deeply learn science *from* these models, students also need to learn *about* the models, so they can analyze their strengths and limitations.

One may argue that if the goal is to learn the science, the workings of the model should disappear into the background. To a large extent, how the model is programmed would be extraneous cognitive load. However, the high-level workings of the model—the cause-and-effect relationships between the variables—are germane. (See Rappin et al.'s [37] DEVICE for an example of an interface that draws students' attention to those relationships and mechanisms, without requiring students to read or write code.) These high-level relationships are often what students are expected to explain as evidence of their learning. And, if computer models are to provide a context for learning about technology and strengthening computational thinking, then the workings of the model are certainly germane. Designing learning environments that reveal useful information about the underlying technology is an important aspect of human-computer interaction. Further, designing these kinds of systems will require an understanding of the prior conceptions about technology that students bring with them. Identifying these prior conceptions will illuminate how children think about technology in a world where it is simultaneously all around them and hidden.

3.2 Computer Models in Science Education

A core component of the practice of science is creating and critiquing conceptual models to explain phenomenon, make predictions, and communicate ideas [9, 26, 36]. Conceptual models can be instantiated with computer models. Students in grades 6-12 can learn from dynamic, interactive models [4, 45, 52]. A key benefit of learning from models is that they visualize the mechanisms behind scientific phenomena [29]. While middle school modeling activities can be designed to enhance students' learning of the target science content and their understanding of modeling as a scientific practice [44, 59], scientific models are often incorporated into instruction to teach science content alone, without addressing how the model was constructed, what assumptions it makes, or its limitations.

However, even when the instructional goals focus on the science content, students' ideas about model construction are still important, as they influence how students proceed to learn from the models. Specifically, students learn more science content from dynamic models when they have a better understanding of the nature of models (ie. epistemology of models), including why and how scientists use and modify models [13]. Further, when interacting with dynamic models, students are much more likely to learn overarching cause-and-effect relationships than the mechanisms that lead to those relationships (e.g., *increasing greenhouse gases will raise Earth's temperature* is a relationship, while *greenhouse gases absorb and re-emit infrared radiation from Earth* is a mechanism [34]). The mechanisms explain why and how scientific phenomena occur, and therefore are crucial to students' understanding of the science content. Since these intermediate mechanisms are governed by specific rules underlying the model, instruction on how models work computationally may help students focus on scientific mechanisms. Further, instruction and assessments with models generally focus on students' interactions with accurate models, and students' explanations of what those models portray [36]. However, allowing students to explore models with potentially innaccurate rules could have benefits for science learning, by allowing students to directly test innaccurate ideas they might hold, and by providing contrasting cases that draw attention to specific model behaviors.

When students create computer models, they are forced to engage directly with model rules and the scientific basis for a model's construction. When students construct computer models, they determine how the model should behave. Thus, an important aspect of model construction, that models are built by humans and can show inaccurate results, is at least present tacitly when students build their own models. Schwarz and White [45] describe an instance where a middle school student used circular reasoning to defend their claim: the model they built generated results that were consistent with the original hypothesis. Other students, however, did not accept this line of reasoning, arguing that the model will show whatever it is programmed to do, which is not necessarily consistent with the real world [45]. When students construct their own models, evaluating their accuracy against real-world data is necessary (though students do not always do this step spontaneously [48]). Creating computer models also helps middle school students learn the target science content [14, 45, 46, 51, 58, 59]. However, learning how to create the models takes time. While researchers are beginning to design modeling tasks that do not add instructional time [51], when most middle school students interact with computer models, they do so with pre-built ones. Since the design and basis for pre-built models is inherently less explicit than it is for student-created models, the question of how to make these issues more salient is important. This paper examines whether students can engage productively with the concept of model rules in the context of pre-built science models, and if doing so can elicit deeper science reasoning than the typical, non-rule-oriented approach.

3.3 Computational Thinking

While most instruction on computational thinking has revolved around computer science alone, integrating it with science modeling has great potential [16]. By *computational thinking*, we mean the attitudes and skills of the discipline of computer science that are useful for other domains and applicable beyond programming, including decomposition, abstraction, and algorithms [60]. Prior research has shown that elementary and middle school students are capable of learning sophisticated aspects of computational thinking, including automata [19], and, through programming, algorithms, abstractions, and debugging [20, 38, 61]. The promise of computational thinking is that it is relevant to many domains. One way to combine computational thinking and other subjects is for students to use programming to display information relevant to another subject, such as by programming a sequence of images or descriptions in an animation [11]. A deeper way to make these connections is to leverage computational thinking to better understand the target domain. Integrating computing and science can strengthen students' understanding of both, and mitigate the risk of disjoint knowledge that results from teaching each topic separately. [47, 53]

The mutually supportive concepts within computer science and science and mathematics are so strong that computational thinking has been specifically defined for these areas. In the context of high school science and math education, computational thinking is a set of practices for data, modeling and simulation, computational problem solving, and systems thinking [54]. Modeling and simulation practices include interpreting models to learn about the phenomena they represent, testing hypotheses with models, and evaluating, designing, and creating models [54]. The Next Generation Science Standards include both computational thinking and developing and using models as science and engineering practices for K-12 students, underscoring the importance of instruction in these areas [32].

Therefore, while models in science are primarily used to teach science content, they can also be used to teach computational thinking. In turn, students' computational thinking skills may impact their ability to learn science from the models. Connections between computational thinking and areas outside of computer science have been understudied. Grover and Pea [16] and Repenning et al. [38] call for more research on leveraging instruction on computational thinking to support learning in other domains, such as science. Computational thinking skills are used implicitly when students learn from models in science. Identifying mechanisms in interactive, dynamic computer models requires, at a high level, reverse engineering the model to uncover the key underlying rules that the objects in the model are following. Therefore, identifying mechanisms involves computational thinking, in particular, decomposition, abstraction, and algorithms [60]. Students use *decomposition* when they examine each object in the simulation individually, and determine its range of behaviors under different circumstances. Students use *abstraction* when mapping objects and behaviors in the simulation to the real-world entities and actions they represent. Finally, students mentally construct a high-level *algorithm* that would generate the display shown by the simulation. Expecting students to achieve these goals assumes that students: (1) conceive of computer models as collections of non-contradictory rules; (2) that students can interact with and observe the model to determine if the model follows or violates a hypothesized rule; (3) that students, realizing that computer models are abstractions and simplifications, can recognize when relevant variables are not included in a model; and (4) that students can differentiate between a correct science concept and the implementation of that concept in a model, with the understanding that computer models require science concepts to be operationalized.

The computational thinking practices described above have specific roles in students' explicit and implicit understanding of computer models. Notably, students may show correct understanding of computation when explicitly asked in the abstract, but still implicitly rely on non-scientific thinking for specific applications [35]. It is important to design assessments that reveal how students are applying computational thinking to specific science topics in educational contexts [7]. This paper examines students' computational thinking in the context of science models, with results that point to instructional designs for how computational thinking and learning from models can support each other.

This paper uses terminology from the field of science education in describing students' ideas. Instead of using the terms "correct" or "wrong", this paper will often use the terms "normative" and "non-normative." This choice of terms is intended to emphasize that scientists' understanding of the natural world is always changing, and a commonly-accepted, normative view may not be fully correct. Further, we want to emphasize that students often draw on valid forms of reasoning and valid sources of data even when they produce "incorrect" responses. Finally, it can be hard to shake the connotation that "wrong" answers hold no educational value and should be avoided. By using the term "non-normative", we hope it will be easier to embrace the educational value of these ideas as contexts for deeper learning.

4 METHOD

The Computational Modeling Inventory examines what ideas students bring to their interactions with computer models. The online assessment includes computer models, each with an animated simulation and graphical data output, for three topics: plant growth (Figure 1), chemical reactions (Figure 2), and global climate (Figure 3). The models included in the assessment had been tested and refined in Web-based Inquiry Science Environment (WISE) units and were known to be interpretable by middle school students (i.e., they were included in units such as [42, 52]). The inventory includes instructions for how to run each model to show the model behaviors that the assessment items ask about. This paper presents results from the subset of assessment items that are aligned with the two research questions above.

4.1 Participants

The Computational Modeling Inventory was given to 253 7th grade students of three teachers at two local public schools (School A and School B). All participating classes had studied the WISE photosynthesis unit earlier in the school year (247 students, 98% of the sample). Of the 202 7th graders who attended School A, 171 of them (85%) had studied a WISE unit on global climate as 6th graders (none of the 51 students at School B had done so; overall, 68% of students in the study had studied the WISE global climate unit). Approximately 40 students saw a version of the global climate unit which included the model on the assessment (with the same animation and both data output graphs); the remaining students saw a version of the model with the same animation but only one of the two data output graphs. None of the students had done the chemical reactions unit.

The two teachers at school A had their 202 students do the assessment individually (95 self-reported as female, 100 as male, and 7 declined to answer), and the teacher at school B had students do the assessment in pairs (46 students worked in pairs, 5 worked individually; we could not collect self-reports of gender from students who worked in pairs).

While the assessment was intended to take one class period, teachers used their discretion in giving students additional time. We report the time spent on the assessment based on the number of seconds logged on the first day the assessment was given. Additionally, if students took at least 3 minutes on a second or third day, that time is included. If students logged less than 3 minutes on a second or third day, we hand-inspected the logs and only included that time if they changed their answers. If students logged less than 3 minutes on a second or third day and did not change their answers, we assume that they were trying to navigate to another activity and entered the assessment by mistake. The majority of students at School A took less than 55 minutes on the

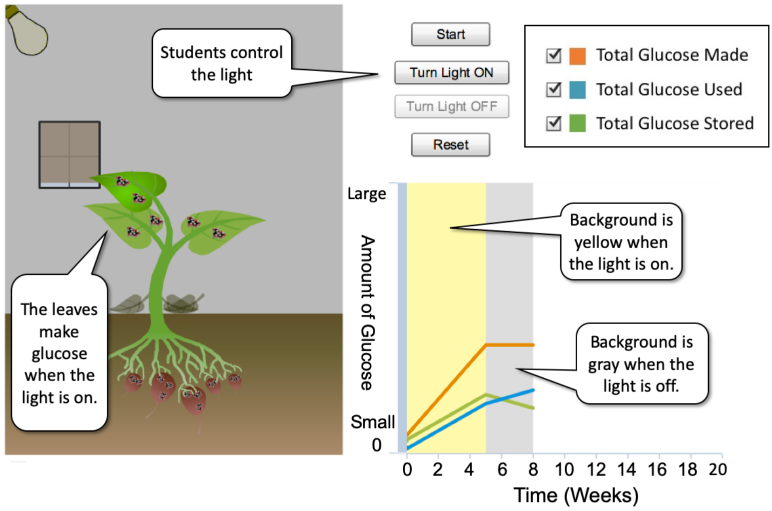

Fig. 1. Screenshot of the plant growth model and graph key (the explanatory callouts are not part of the model). The animation shows glucose forming in the leaves of the plant when the light is on. Glucose that is not used immediately moves down the stem and is stored in the roots. When a molecule of glucose is used, it disappears. The cumulative graph shows the total amount of glucose made and used over time, and the amount in storage. Cumulative glucose made increases when the plant produces glucose in the light, and remains unchanged when the plant does not produce glucose (in the dark). Cumulative glucose used always increases since the plant uses glucose continuously while it is alive.

assessment (mean: 49 minutes; median: 46 minutes; 32% (66) took less than 40 minutes, 38% (77) took 40-55 minutes, 17% (35) took 55-70 minutes, and 12% (24) took more than 70 minutes). 78% of the students at School A (158) took the assessment over one day; 18% (37) took two days; 3% (7) took three days. The majority of students at School B took less than 35 minutes on the assessment (mean: 29 minutes; median: 27 minutes; 10 workgroups took less than 20 minutes; 9 took 20-35 minutes, and 9 took over 35 minutes). At School B, 20 workgroups took one day for the assessment, 7 took two days, and 1 took three days. These durations were for the entire assessment, including seven open-response questions and demographic survey questions which are not reported on here.

Since some students did the assessment in pairs, our analysis is at the level of workgroups, with each workgroup consisting of one or two students. Since the goal of this paper is to document the range of ideas that students have about models, we do not exclude the students who worked in pairs, since their responses represent the ideas of at least one student.

4.2 Instruments and Scoring

230 workgroups participated in the assessment (46 students worked in pairs, and 207 students worked individually). Table 1 shows the question types and the number of workgroups who answered each question for the subset of items discussed in this paper. Rule Sorting questions consisted of a drag-and-drop interface where students sorted text statements into predefined categories. Open Response questions consisted of a prompt and a text box. Multiple Choice and Explain questions consisted of a multiple choice question with a prompt for students to explain their choice in an open-response box. Screenshots of each model are shown in figures 1-3, and the text of each question (along with rubrics for open-response items) is given in their respective results and discussion sections.

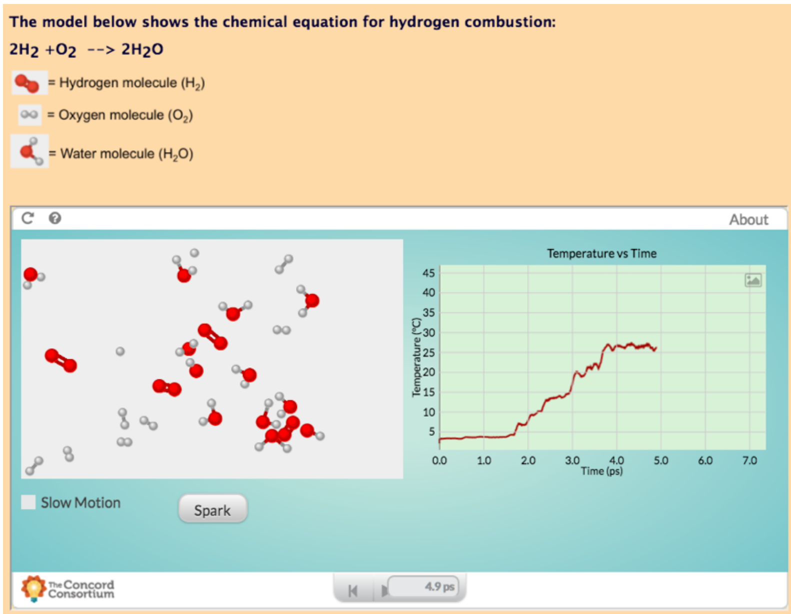

Fig. 2. Screenshot of the chemical reactions model. When the Spark button is pressed, a spark is added to the oxygen and hydrogen gas. This prompts a combustion reaction, which creates water and increases the kinetic energy of the molecules, resulting in a warmer temperature. The red line on the graph shows temperature over time.

To make it easier for students to go back and forth between answering the questions and interacting with the model, the relevant model was included on each question page, and students could interact with the models as much as they wished. The question prompts directed students to run the models in ways that would illustrate all of the behaviors targeted by the assessment items. For the plant growth model, students were told to try running the model for 4 weeks and then turn off the light. For the chemical reactions model, students were told to press the "Spark" button (which starts the reaction). For the global climate model, students were asked to run the model and observe it without greenhouse gasses, and then to press the "Run Factory" button to add greenhouse gasses to the atmosphere. The models did not generate logs of students' interactions, so there is no record of how or how often students ran the models.

Individual rules in the rule sorting questions and responses to the Multiple Choice questions were scored as correct or incorrect. All open response and explanation items were categorized by the first author with emergent, iterative coding schemes. We present response categories and percentages of responses that fall into each category for each question. However, the percentages reported here are not intended to estimate the true prevalence of these ideas in the general population of middle school students. Rather, we include them to show that the selected categories cover most or all substantive responses in the sample, and to give an idea of which ideas are most common.

Rubrics were validated with double-coding by a second coder. The first author developed Knowledge Integration rubrics for the items *Emergent Pattern*, *Greenhouse Gas*, *Temperature*, *Reflect Sunlight*, and *Two Hydrogens Bond* (rubric categories and examples are shown in the Results sections). Knowledge Integration constructs are applicable across science topics, and Knowledge Integration assessments have been shown to offer fine-grained measures of student understanding,

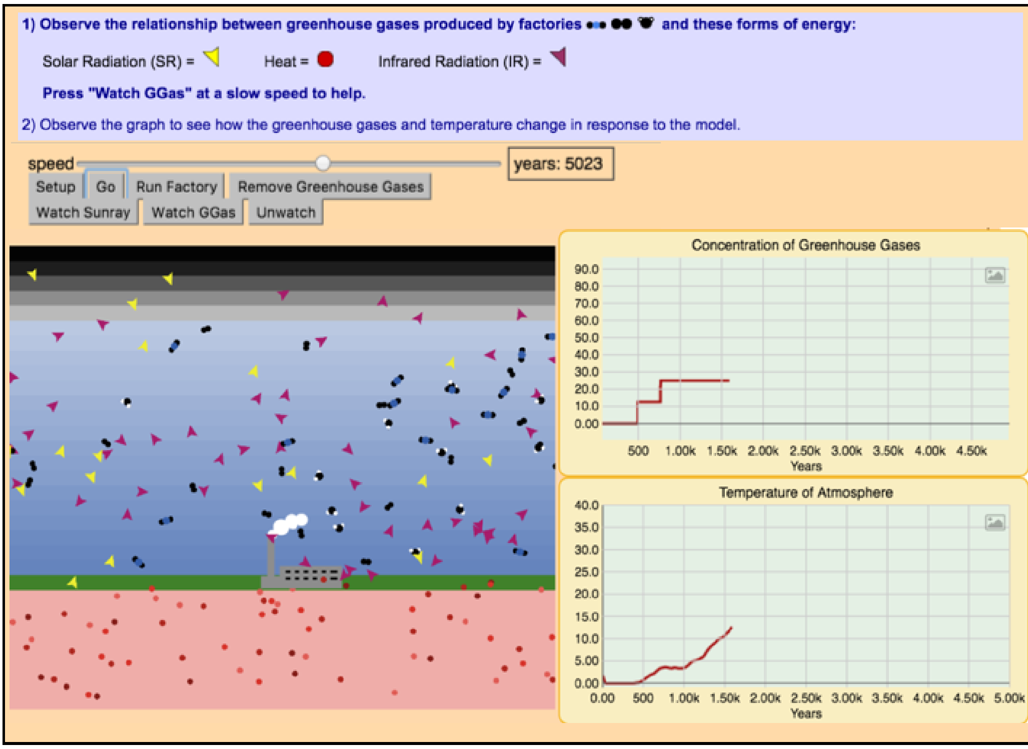

Fig. 3. Screenshot of the climate change model. Students can add and remove greenhouse gases (with the *Run Factory* and *Remove Greenhouse Gases* buttons). Dynamic graphs show the concentration of greenhouse gases in the air (top) and the atmospheric temperature (bottom). Greenhouse gases reflect infrared radiation back toward Earth instead of letting the radiation escape to outer space, which causes an increase in atmospheric temperature.

Table 1. Number of workgroups who answered each question

Question	Number of Workgroups (N = 230)
Rule Sorting, Global Climate	Sorted at least one rule: 169 (73%)
Rule Sorting, Plant Growth	Sorted at least one rule: 207 (90%)
Emergent Pattern in Chemical Reactions (only given to one teacher's classes; 140 workgroups)	Multiple Choice: 133 (95%) Explanation: 132 (94%)
Greenhouse Gas: *Typical* question, Global Climate	Open Response: 186 (81%)
Temperature: *Typical* question, Chemical Reactions	Open Response: 203 (88%)
Reflect Sunlight: *New Rule*, Global Climate	Open Response: 172 (75%)
Two Hydrogens Bond: *New Rule*, Chemical Reactions	Multiple Choice: 200 (87%) Explanation: 198 (86%)

with good psychometric properties [24]. The rubrics differentiate between normative and non-normative ideas, and between mechanistic reasoning and other types of explanations. Each rubric included 20-34 example responses. The first author coded all responses, and a second coder used the rubrics to code one or two sets of random samples of 20 of the remaining non-blank responses for each question.

To measure the reliability of the rubrics, agreements and Cohen's kappas were calculated for each double-coded set. For the Emergent Pattern item, agreement between the two coders was 80% (Cohen's kappa of .75), indicating good reliability of the coding scheme. The rubrics for the other questions included separate categories for different combinations of the correctness of the response and the type of mechanism included, and since some analyses examine only correctness or only the presence of mechanisms, agreement and Cohen's kappa were also calculated for the collapsed categories (with fully correct, partial, and non-normative mechanisms collapsed into *present*, and the remaining categories collapsed into *absent*). On the first double-coded set, agreement on the presence or absence of intermediate mechanisms was 100% for the *Temperature* item, and 45-95% for the other questions, with Cohen's kappa ranging from .08 to .64. For Temperature, agreement on the full rubric was 90%, with Cohen's kappa of .88. For the other three questions, the two coders discussed and resolved disagreements from the first set, and then the second coder coded another set of 20 randomly-selected, non-blank responses. For this second set, agreements on mechanisms (present or absent) were either 85% or 90%, with Cohen's kappa ranging from .57 to .80. While agreements were high for the presence or absence of mechanisms, the kappas were more variable because the distributions were skewed. On all double-coded sets, agreement on the correctness of each response (also referred to as "outcome" in the analyses) was 90% or greater for each question, with Cohen's kappa ranging from .79 to .93. On the second double-coded sets, agreements across all rubric categories were either 80% or 85%, with Cohen's kappa of either .73 or .79. These measures indicate good reliability of the rubrics.

The Two Hydrogens Bond question included a multiple choice component (for the outcome) and an open response for students to explain their choice. The kappas for outcome reported above are based on open-response scoring, not the multiple-choice. Of the 158 non-blank open responses, coding for outcomes matched the multiple-choice selection in 150 (95%) of responses. The remaining 8 responses had the correct multiple-choice selection, but with open responses that were off-task, incomplete, or indicating misunderstanding of the new rule.

5 IDENTIFYING RULES: RESULTS AND DISCUSSION

When students interact with a computer model, to what extent do they analyze the model as a computational tool to identify what rules the model is following? To answer this question, the Computational Modeling Inventory engaged students with the concept of *rules*. Rules sit between science concepts and computer code. They are plain language statements that describe the behavior or quantity of objects in the model. The concept of representing a computational model as a set of rules or instructions is typical of introductory computer science education. For example, students are often asked to write a set of rules for making a peanut butter and jelly sandwich [8]. The rules in the Computational Modeling Inventory were written to be interpretable by middle school students, and to focus on key model behaviors and common alternative ideas.

We examined this research question with the plant growth and global climate models, through *Rule Sorting*. Rule Sorting tasks include a set of proposed rules and ask students to sort them into two categories: Rules used by the computer vs. Rules not used by the computer. These questions examine if students can see computer models as instantiations of programmed rules, as opposed to viewing the models as showing the scientific phenomenon directly, or as a system without underlying logic. If students can identify the rules behind a computer model, then they are closer to understanding that a scientific model is not simply a graph, or an animation, but rather is the underlying rules and assumptions that produce those visualizations. The rules are written at the level of object behaviors, not individual lines of code. The behaviors in the rules follow *computational thinking patterns* for simulations, such as collision, absorption, and generation [38]. Therefore, Rule Sorting questions include proposed rules on specific model behaviors (eg., that total glucose made increases while the light is on) that, combined with other behaviors, results in a model that is consistent with key scientific concepts (e.g., plants make extra glucose during the day, and storing this excess lets them survive at night). While the rules were not written at the level of individual lines of code, they were written to align with the actions that code can perform, such as updating the value of a variable. Therefore, we use the phrasing *When light is on, total glucose made increases*, to align with the idea of a variable `total_glucose_made`, whose value increases when the light is on. Students sorted the rules, but were not asked for their reasoning. For the chemical reactions model, we proposed one rule and asked students if the model was following it or not, and to explain their reasoning.

To address part (1) of this research question, both Rule Sorting tasks include proposed rules which contradicted each other (e.g., Total glucose used always increases vs. When light is off, total glucose used decreases), and we examine how often students thought a model followed both rules in a contradictory pair. The Rule Sorting task for plant growth additionally address parts (2) and (3) by: (2) including proposed rules which refer to variables that are relevant to photosynthesis but are not portrayed in the model; and (3) including "rules" which are phrased as non-operationalized science concepts.

The distinction between a scientific concept and the way it is operationalized in a model is subtle, but important. The underlying code for a computer model must describe concrete system actions, not overarching scientific principles. While the plant growth model is consistent with the scientific idea *Plants use light to make glucose in photosynthesis* it operationalizes the idea in a particular way: when the light is on, the plant makes a set amount of glucose per unit time, which increases the running total amount of glucose made (that is, *When light is on, total glucose made increases*). Another model could operationalize the same idea in a different way, perhaps by including carbon dioxide in the model and specifying that light and carbon dioxide must both be present for glucose to be produced. The alternate model would still be consistent with the scientific idea, but would not follow the same particular rule, since turning on the light would not be the sole criterion for making glucose. Recognizing that an overarching idea can be implemented in different ways is crucial for understanding why different models can have different levels of accuracy, even if they all implement the same general principle. Therefore, we examine if students can make a distinction between a scientific concept (*Plants use light to make glucose in photosynthesis*) and an operational rule that defines the behavior of a model (*When light is on, total glucose made increases*).

Finally, to address part (4), we investigate model behaviors that are not the direct result of a single underlying rule, but are emergent patterns. The distinction between a model rule and an emergent pattern is also subtle, because emergent patterns can be described with operational model rules, and the behavior of the model is consistent with such rules. The subtlety lies in the difference between a rule being *consistent* with the model's behavior and being *necessary* for the model's behavior. For example, in the plant growth model, the amount of glucose in storage goes down when the light is off. However, the rules *When light is off, total glucose stored decreases* and *When light is on, total glucose stored increases* are not necessarily part of the model, since the same behavior would be demonstrated with a rule that defines a change in glucose stored as the difference between glucose made and glucose used. We further examine students' thinking about rules and emergent patterns by asking if a proposed rule must be part of the chemical reactions model, and asking students to explain their reasoning.

Students responses to the Rule Sorting questions illuminate how they think computation happens within the simulation. The concepts of what rules are possible, how rules are connected to behavior, and how rules interact with each other point to students' ideas about *notional machines* [10] for simulations. Sorva's excellent literature review explains "a notional machine is a characterization of the computer's role as executor of programs" [49, p.2]. Just as programmers need a conception of a notional machine to think about the relationship between their code and a running program [49], the students in this study rely on a mental model (albeit an implicit one) of a notional machine to connect a simulation's rules and behaviors.

5.1 Rule Sorting

In each Rule Sorting question, half of the proposed rules are correct model rules, and the rest are not. All proposed rules for global climate refer to the animation, with all incorrect rules directly contradicting the animated behaviors. Some rules for the plant growth model refer to the animation, and some to the graph, with incorrect rules directly contradicting the graph or animation. Additionally, some incorrect rules for plant growth reference variables that are not included in the model, and one rule describes a correct science concept that the model portrays, rather than the way this concept is operationalized in the model. The proposed rules and frequency of selection are shown in Table 2 and Table 3 for plant growth and global climate, respectively (note that percentages may not total 100 due to rounding). Workgroups are included in this analysis if they submitted an answer to a rule sorting question (even if the answer was blank); 213 workgroups are included overall.

Students were directed to run the models in ways that would show the behaviors targeted by the rule sorting questions. On the Rule Sorting item for the plant growth model, students were directed to run the model for 4 weeks (with the light on by default) and then turn the light off. Running the model this way shows all of the behaviors that the rule sorting item targets: when the light is on, new molecules of glucose appear on the plant's leaves, and on the graph, the lines for total glucose made (rule 5), used (rule 3), and stored (rule 7) increase. When the light is off, no new glucose appears on the leaves, and on the graph, the line for total glucose made is flat (rule 1). Since the plant is not making new glucose, it must use the glucose it has already stored. The animation shows glucose molecules fading and then disappearing. On the graph, the line for total glucose used keeps increasing (rule 3), and the line for total glucose stored decreases (rule 6). After several weeks in the dark, the plant's stores of glucose are depleted, and when the plant dies, the graph shows the total glucose stored is at 0 (rule 4) and the line for total glucose made intersects with the line for total glucose used. Both the animation and the graph indicate that the amount of glucose the plant has in storage is the cumulative amount it has made minus whatever has been used already (rule 2), but arriving at this conclusion requires some interpretation.

For the global climate model, students were told to run the model first without greenhouse gases (the default) and then to press the "Run Factory" button. When that button is pressed, greenhouse gases emerge from the factory and start bouncing around the atmosphere. These instructions were part of an item that came immediately before the rule sorting item. The model was included on the same page as the rules to be sorted, but the instructions for using the model were not repeated. All of the behaviors targeted by the rule sorting item can be observed by running the model with greenhouse gases. Even with the smallest amount of greenhouse gases, there are enough molecules that collisions between the gas molecules and solar and infrared radiation can be observed within 5 seconds (likewise for the solar and infrared radiation moving to different parts of the simulation). The model begins with solar radiation coming down from the top, and heat dots bouncing around in the interior of the Earth (the pink stripe). Solar radiation (the yellow arrows) starts from the top of the model and moves downward. When solar radiation hits a greenhouse gas, each continues in its original direction (rule 1). The solar radiation continues until it hits Earth's surface (the green stripe). Then, the solar radiation either gets reflected (bounces off and starts moving back to the top of the screen) or gets absorbed as heat (rule 6, the yellow arrow disappears, a red dot appears in its place, and the red dot continues into the interior of the Earth). When a solar radiation arrow reaches the very top, it disappears (rule 2). When a heat dot hits the surface of the Earth, it either bounces back down or it converts into infrared radiation (rule 5, the dot disappears and a red arrow appears in its place, moving in its original direction, away from Earth's surface). When a red infrared radiation arrow hits a greenhouse gas molecule, the infrared radiation changes direction, and keeps moving in the new direction (rule 3). When a greenhouse gas molecule hits the top of the screen or the surface of the Earth, it bounces off, staying in the atmosphere (rule 4).

Table 2. Rule Sorting for Plant Growth (N=207). Question Prompt: Run the model below. Try turning the light off after 4 weeks and observe the graph. The computer follows a set of rules to create the model for plant growth. The model produces the graph showing the relationship between light and the use of glucose inside a plant. Sara is trying to figure out the rules the computer used to create the graph, but some of her rules are not correct. **She wrote these rules, below, about what happens when the plant is alive. Help Sara by dragging the rules below into the correct buckets.** (Bold in original)

Proposed Rules that the Model is Following While the Plant is Alive	Sorted Correctly	Sorted Incorrectly	Not Sorted
1. When light is off, total glucose made stays the same. *Observable in the graph, but requires the realization that the graph is cumulative.*	53%	36%	10%
2. Total glucose stored = Total glucose made - Total glucose used. *Requires graph interpretation and understanding of the science content in the model. This correct rule is contradicted by rule 13.*	54%	32%	13%
3. Total glucose used always increases. *Observable in the graph. This correct rule is contradicted by rules 12 and 14.*	57%	32%	10%
4. When total glucose stored is 0, the plant dies. *Observable in the graph.*	73%	16%	10%
5. When light is on, total glucose made increases. *Observable in the graph.*	79%	10%	11%
Unnecessary/Redundant Plant Growth Rules	**Sorted In Model**	**Sorted Not In Model**	**Not Sorted**
6. When light is off, total glucose stored decreases. *Observable in the graph.*	77%	14%	8%
7. When light is on, total glucose stored increases. *Observable in the graph.*	77%	13%	10%
Proposed Rules that the Plant Growth Model Is Not Following	**Sorted Correctly**	**Sorted Incorrectly**	**Not Sorted**
8. Plants use light to make glucose in photosynthesis. *A correct science concept that is consistent with the model, but is not phrased as an operationalized rule that specifies a behavior.*	32%	56%	12%

Continued on next page

Table 2: Plant Growth Rules, continued from previous page

Proposed Rules that the Plant Growth Model Is Not Following	Sorted Correctly	Sorted Incorrectly	Not Sorted
9. Plants use carbon dioxide in photosynthesis. *A correct science concept, but not an operationalized rule that specifies a behavior. Further, this model does not include carbon dioxide.*	49%	40%	11%
10. Plants release oxygen in photosynthesis. *A correct science concept, but not an operationalized rule that specifies a behavior. Further, this model does not include oxygen.*	53%	36%	11%
11. When the light is off for 4 weeks, the plant dies. *Observable from the model that this is not true - it depends on how long the light was on previously.*	55%	37%	8%
12. When light is off, total glucose used decreases. *Observable from the graph that this does not happen. This rule directly contradicts rule 3.*	57%	32%	10%
13. Total glucose stored = Total glucose made + Total glucose used. *Requires graph interpretation and understanding of the science content in the model. This rule is directly contradicted by rule 2.*	59%	28%	13%
14. When light is off, plants do not use glucose. *Observable from the graph that this does not happen. This rule is contradicted by rule 3.*	74%	14%	11%

Table 3. Rule Sorting for Global Climate (N=169). Question Prompt: The computer follows a set of rules to create a model. This model produces graphs showing how greenhouse gases and temperature change over time. *1) Observe the relationship between greenhouse gases produced by the factory and these forms of energy: Solar Radiation, Heat, Infrared radiation. Press "Watch GGas" at a slow speed to help. 2) Observe the graph to see how the greenhouse gases and temperature change in response to the model.* Terry is trying to figure out the rules the computer used to create the graph, but some of her rules are not correct. **She wrote these rules, below. Help Terry by dragging the rules below into the correct buckets.** (Bold in original. The portion in italics is embedded in the Global Climate model and includes the key, as shown in Figure 3).

Proposed Rules that the Model is Following *All are directly observable in the model.*	Sorted Correctly	Sorted Incorrectly	Not Sorted
1. When solar radiation hits a greenhouse gas, it will keep moving forwards. *Contradicts rule 9.*	64%	27%	9%
2. When solar radiation reaches the top of screen, it will always disappear. *Contradicts rule 7.*	65%	29%	7%
3. When infrared radiation hits a greenhouse gas, it will change direction. *Contradicts rule 12.*	65%	24%	11%

Continued on next page

Table 3: Global Climate Rules, continued from previous page

4. When greenhouse gas hits surface of the earth, it will always change direction. *Contradicts rule 11.*	70%	23%	7%
5. When heat reaches the top layer of the earth, it will sometimes convert into infrared radiation. *Contradicts rule 8.*	73%	24%	3%
6. When solar radiation reaches ground, it will sometimes convert into heat. *Contradicts rule 10.*	82%	12%	5%
Proposed Rules that the Global Climate Model Is Not Following	Sorted Correctly	Sorted Incorrectly	Not Sorted
7. When solar radiation reaches the top of screen, it will sometimes change direction. *Contradicts rule 2.*	57%	32%	11%
8. When heat reaches the top layer of the earth, it will always change direction. *Contradicts rule 5.*	57%	34%	9%
9. When solar radiation hits a greenhouse gas, it will change direction. *Contradicts rule 1.*	58%	35%	7%
10. When solar radiation reaches ground, it will always convert into heat. *Contradicts rule 6.*	62%	28%	10%
11. When greenhouse gas hits surface of the earth, it will sometimes disappear. *Contradicts rule 4.*	68%	24%	8%
12. When infrared radiation hits a greenhouse gas, it will convert into heat. *Contradicts rule 3.*	72%	18%	10%

5.1.1 Rule Sorting Overall Results: When students interact with computer models, they can identify some of the rules a model is following. Analyses include workgroups that submitted answers to the Rule Sorting questions (including blank answers): 201 workgroups sorted at least one rule for the plant growth question, and 6 workgroups submitted blank answers to that question; 168 workgroups sorted at least one rule for the global climate question, and 1 workgroup submitted a blank answer to that question. More workgroups submitted answers to the plant growth question than the global climate question because the plant growth question was presented earlier on the assessment. In all, 213 workgroups submitted an answer to at least one of those questions. 163 workgroups submitted answers to both questions, and 159 submitted non-blank answers to both questions. No workgroups in the analysis submitted blank answers to both questions, though 3 workgroups submitted blank answers to the plant growth question and did not submit any answer for the global climate question.

The vast majority of workgroups in this analysis sorted all of the rules: 177 (86%) did so for the plant growth model, and 149 (88%) did so for global climate. The pattern of responses indicates that students were not guessing randomly on these items: binomial tests for population proportions on sorted responses, by rule, indicate that most of the proportions are significantly different from 50% (note that these tests excluded workgroups, by rule, that did not sort that rule). 11 plant growth rules and all global climate rules are significant at $p < .01$, and 2 plant growth (rules 1 and 10) are significant at $p < .02$. Only one rule, rule 9 from the plant growth model, had a proportion that did not differ significantly from 50% ($p = .16$). This analysis indicates that, overall, the students sorted the rules thoughtfully, rather than guessing randomly.

In examining the average number of rules correctly sorted for each model, we find that Rule Sorting was not simple for students. For the global climate model, all rules are either directly observable in the animation or are directly contradicted by the animation. These rules are all expressed in an operational manner. On average, workgroups sorted 66% of these rules correctly, with the remaining rules sorted incorrectly or not sorted. 99% of workgroups sorted at least one rule correctly; 80% were correct on six or more; and 21% were correct on all 12. It was easier for students to accept a correct rule than reject an incorrect rule, with 70% of correct rules sorted correctly, and 62% of incorrect rules sorted correctly. A paired-samples t-test on the number of correct and incorrect rules that each workgroup sorted correctly shows that this difference is significant ($t(168) = 5.59, p < .001$). Still, performance on these two types of rules is strongly correlated (Pearson's $r = .80, n = 169, p < .001$), indicating that students who are more able to accept correct rules are also more able to reject incorrect rules. Finally, the climate model rules showed good reliability as a scale (Cronbach's alpha is .78, and that measure would not be improved with the deletion of any single item).

The proposed rules for the plant growth model are more complex than those proposed for the global climate model. One complexity comes from two rules which do not contradict the model, but also are not necessarily part of the model because they are redundant with other correct rules. Since legitimate arguments can be made on their correctness, we analyze them separately. To examine results across types of rules, we score each rule independently: 1 if sorted correctly, 0 otherwise. The average score for the redundant rules is 77%, and the average score for the remaining rules is 58%. 96% of workgroups sorted at least one observable rule correctly; 67% were correct on four or more, and 26% were correct on all seven. The rules that were observable in the model were easier for students to sort correctly compared to the other rules, with average scores of 64% for sorting observable rules and 50% for sorting the other proposed rules (excluding the redundant rules). A paired samples t-test on the proportion of observable and non-observable rules that each workgroup sorted correctly shows that this difference is significant ($t(206) = 7.17, p < .001$), indicating that it is easier for students to tell if a model is following a rule or not if the proposed rule is directly observable or directly contradicted in the model, as opposed to other kinds of rules. As with the global climate rules, it was easier for students to accept correct rules than reject incorrect rules, with 64% of correct rules sorted correctly and 54% of incorrect rules sorted correctly. A paired samples t-test on the proportion of correct and incorrect rules that each workgroup sorted correctly shows that this difference is statistically significant ($t(206) = 6.27, p < .001$). Still, performance on these two types of rules is strongly correlated (Pearson's $r = .76, n = 207, p < .001$), indicating that students who are more able to accept correct rules are also more able to reject incorrect rules. Overall, the plant growth model rules showed good reliability as a scale (Cronbach's alpha is .85, and that measure would not be improved with the deletion of any single item).

While the plant growth question was harder overall, the observable rules posed similar difficulty levels for both plant growth and global climate. Of the 163 workgroups who submitted answers to both questions, average scores were 67% correct on the global climate rules and 64% correct on the observable, non-redundant rules in the plant growth model. A paired t-test shows that this difference in scores for observable rules is not significant ($t(162) = -1.33, p = .18$). While the global climate rules all pertained to the climate animation, the observable plant growth rules pertained to the dynamic cumulative graph. We were surprised that performance was not significantly different across the two contexts (graph and animation), since the graph is more abstract. 23 of these 163 workgroups (14%) sorted all observable, non-redundant rules correctly across both questions.

Analysis of the Rule Sorting tasks indicate that students took the tasks seriously and that the scales are coherent. Each rule was sorted by at least 86% of the workgroups, and binomial tests indicate that for 25 out of 26 rules, workgroups were not randomly guessing between the two options. Both Rule Sorting tasks demonstrate coherence, with Cronbach's alpha of .78 and .85 for the global climate and plant growth tasks, respectively, indicating that each Rule Sorting task has adequate reliability.

5.1.2 Discussion. Students identified model rules to a moderate extent: above chance but far below ceiling for rules that were either directly observable or directly contradicted by the model. Notably, the interactive models were included for each question, and most students in the sample had previous exposure to these models: 98% of students in the sample had studied a unit with the plant growth model, and 68% had studied a unit with a global climate model with the same animation as the assessment (and all of the global climate rules referred to the animation, not the graphs).

Students' sorting of observable rules indicates that it is not easy for them to identify key model behaviors. Correct rules for the global climate and plant growth models (e.g., *When infrared radiation hits a greenhouse gas, it will change direction* and *Total glucose used always increases*) were sorted correctly by 65% and 57% of workgroups, respectively. The behaviors that follow from these rules are directly observable in the model and the graph, and embody crucial ideas about climate change and plant growth. We suspect that difficulty recognizing model rules poses a barrier to students' understanding of the science and of modeling as a practice. We do not believe it is necessary for students to identify every model rule, or to express these rules at the level of computer code. However, if students are to extract important cause-and-effect relationships from these models, logical prerequisite steps include observing the behaviors that portray those relationships and interpreting those behaviors as outcomes of programmed rules. Connecting rules with observable behaviors is the first step in considering the *algorithms* behind the models. The logic of a notional machine may also be helpful [49], by making explicit that all observable behaviors are caused by the underlying rules.

5.1.3 65% of workgroups think model rules can contradict each other. The Rule Sorting tasks also show that students often think a computer model can follow contradictory rules. For the plant growth model, two correct rules were contradicted by incorrect rules (e.g., *Total glucose used always increases* vs. *When light is off, total glucose used decreases*). For the global climate model, each correct rule was contradicted by an incorrect rule (e.g., *When solar radiation reaches the top of screen, it will always disappear* vs. *When solar radiation reaches the top of screen, it will sometimes change direction*). Separately from assessing students' evaluations of the proposed rules, we examined how often students think the computer is following both rules from a contradictory pair. For the plant growth model, 49 of 207 workgroups (24%) indicated that the model followed at least one pair of contradictory rules; for the global climate model, 103 of 169 workgroups (61%) did so. Of the 163 workgroups who submitted answers to both questions, 106 (65%) indicated that at least one model would follow at least one pair of contradictory rules. This was not because students were guessing that the computer followed all proposed rules, which no workgroup did for either model.

5.1.4 43% of workgroups rejected rules that referred to variables that were not in the model. When interacting with computer models, which are necessarily simplifications of real-world phenomena, students may be confused about the many factors which relate to a phenomenon in real life versus the limited number of factors that are portrayed in the model. Two proposed rules for the plant growth model present the correct scientific concepts that plants use carbon dioxide and release oxygen in photosynthesis. However, carbon dioxide and oxygen are not part of the plant growth model. Yet, only 102 workgroups (49%) correctly sorted the rule for carbon dioxide; 109 (53%) correctly sorted the rule for oxygen. 89 workgroups (43%) sorted both rules correctly, and 62 workgroups (30%) incorrectly indicated that the model included both rules. However, 29 workgroups (14%) declined to sort one or both rules. Considering only the 178 workgroups who sorted both rules, 89 workgroups (50%) sorted both correctly, 62 workgroups (35%) thought the model included both rules, and 27 workgroups (15%) thought the model included only one of those rules.

5.1.5 56% of workgroups cannot distinguish between a science concept and the way that concept is operationalized in a model. The most difficult rule to sort was the proposed rule for the plant growth model *Plants use light to make glucose in photosynthesis*. 67 workgroups, or 32%, correctly sorted it as not a rule the model was following, with 116 workgroups, or 56%, saying the model was following that rule. This rule was particularly difficult because light is a variable in the model, and the model overall is consistent with this statement. However, as discussed in the introduction to the Rule Sorting results, this "rule" does not operationalize the concept by describing the behavior or quantities of objects in the model.

61 of the 67 workgroups (91%) who correctly rejected *Plants use light to make glucose in photosynthesis* also correctly accepted *When light is on, total glucose made increases* (Table 4, left). This proportion was not higher among groups who incorrectly accepted *Plants use light to make glucose in photosynthesis* (87%). Examining the 187 workgroups who sorted one or both of those rules, a Fisher Exact test on a Table 4 (left) shows that there is no significant relationship between the two questions ($p = .36$; 95% confidence interval of the difference between the proportions of workgroups who were correct on the concept for light given their correctness on the rule for light is -.1 to .3; considering only the 180 workgroups who sorted both rules, the relationship remains not significant, $p = .45$). There is a relationship between thinking that the model follows

Table 4. Relationships between correctly rejecting the concept *Plants use light to make glucose in photosynthesis* and correctness on other rules: at left, accepting *When light is on, total glucose made increases* (for the workgroups sorting one or both rules); at right, rejecting *Plants use carbon dioxide in photosynthesis* and *Plants release oxygen in photosynthesis* (for workgroups sorting all three rules).

N=187	Accepted the *concept* for light	Rejected the *concept* for light
Rejected the *rule* for light	17	6
Accepted the *rule* for light	103	61 correct on both

N=178	Accepted the *concept* for light	Rejected the *concept* for light
Accepted a rule for O_2, CO_2, or both	74	15
Rejected rules for O_2 and CO_2	39	50 correct on all 3

the rule *Plants use light to make glucose in photosynthesis* and thinking that the model is following rules that refer to carbon dioxide or oxygen (Table 4, right). Examining the 178 workgroups who sorted all three of these rules, a Fisher Exact test on Table 4 (right) shows that there is a significant relationship between rejecting *Plants use light to make glucose in photosynthesis* and rejecting rules with non-represented variables ($p < .0001$). Of the 113 workgroups who incorrectly accepted the concept about light, only 39 (35%) correctly rejected both rules that referred to carbon dioxide or oxygen; of the 65 workgroups who correctly rejected the concept about light, 50 (77%) correctly rejected both rules that referred to carbon dioxide or oxygen. From the other direction, of the 89 workgroups who correctly rejected the two rules that referred to carbon dioxide or oxygen, 50 of them (56%) also rejected *Plants use light to make glucose in photosynthesis*. Of the 89 workgroups who incorrectly accepted either rule, only 15 workgroups (17%) correctly rejected *Plants use light to make glucose in photosynthesis*.

5.1.6 Contradictions, Variables, and Operationalizations Discussion. The majority of workgroups (65%) who answered both questions indicated that a model could follow contradictory rules. Middle school students learning science often hold contradictory ideas simultaneously [23], and therefore they may hold contradictory ideas about the models they interact with. Alternatively, they may not recognize that two proposed rules are mutually exclusive. Students may be considering each rule individually, and may need guidance to reflect on their selected rules as a whole. Selecting contradictory rules illustrates the kinds of ideas students develop when the technology disappears. Instruction that helps students integrate an understanding of algorithms with their learning from models could help them appreciate the problems resulting from contradictory rules. Noticing the impossibility of contradictions (from a computational perspective) may prompt more careful thinking about the science to determine which rule is correct.

Students struggled to sort the two proposed rules for the plant growth model that refer to variables that are not present, with only 43% of workgroups correctly indicating that neither rule was used by the computer. These results suggest that many students think that computer models somehow account for variables which they do not include. Identifying and distinguishing between variables is an important form of decomposition. To learn the science and understand how the simulation relates to the real world, students must take a scientific phenomenon and conceptually break it up into parts. Students must also understand that only some parts will be represented in the model. Correctly interpreting the results from a model requires students to recognize which variables a model includes, and which contribute to a model's assumptions and limitations. A large minority of workgroups, 30%, thought the plant growth model included rules for both carbon dioxide and oxygen. This suggests that students may think models include more variables than they actually do, and therefore may interpret models as being more accurate than is warranted. These findings indicate that when students know that certain variables are relevant to a science phenomenon, they may assume that all models of that phenomenon include those variables. Students who hold that assumption may have difficulty distinguishing between a model and the science phenomena that the model represents.

Rule sorting indicates that most students do not distinguish between a science concept and its implementation in model, with 56% of workgroups indicating that the model followed the "rule" *Plants use light to make glucose in photosynthesis*. Computer models are valuable tools for scientists precisely because they rely on operational rules rather than concepts; comparing different operationalizations of the same idea can illuminate underlying mechanisms. While it is worthwhile for students to learn that plants use light in photosynthesis, it is also important for students to distinguish between a concept and the way the concept is operationalized. This distinction, between what something is and what it represents, is a key part of computational abstractions.

Importantly, the vast majority of workgroups (91%) who correctly sorted this concept were also correct in their sorting of the implementation of this concept in the model: *When light is on, total glucose made increases*. These workgroups recognized that while the model did not directly implement the concept, the model was still consistent with that concept. 29% of all workgroups correctly sorted both the concept and the implementation, suggesting that, while most students cannot make this distinction on their own, it is not unreasonable to expect that they could do it with support.

Further, these results show that there is a significant relationship between thinking that a model can directly follow a concept and thinking that a model incorporates more variables than it does. Of the workgroups who agreed that the model followed *Plants use light to make glucose in photosynthesis*, a majority (65%) accepted rules that referenced oxygen or carbon dioxide. In contrast, of the workgroups who rejected that concept about light, only 23% accepted rules that referenced oxygen or carbon dioxide. This relationship underscores the importance of helping students separate science concepts from their implementations: students who cannot recognize the difference may also believe that models always include all necessary variables, which in turn may lead students to believe that computer models are completely accurate and beyond critique. Together, these Rule Sorting results show students' non-scientific ideas not only about algorithms, decomposition, and abstraction, but also about the notional machine as a whole. A version of Roy Pea's *superbug*, the idea that the computer is intelligent on its own [35], seems to be manifesting here. In this context, instead of knowing about the programmer's goals [35], the notional machine is implicitly credited with knowing about the real world and being able to portray it. If asked explicitly, these students would likely say that the simulation can only do what it is programmed to do, like the novices Pea discusses [35]. Yet, again like Pea's novice programming students [35], these middle school students behave in a way that is consistent with the superbug.

5.1.7 95% of students who can identify necessary rules cannot distinguish between model behaviors that follow directly from rules and emergent patterns. The two redundant rules for plant growth were each accepted by 159 and 160 of the 207 workgroups (77%). 136 workgroups (66%) accepted both rules. The redundant rules state that glucose stored increases when the light is on and decreases when the light is off. This is consistent with the model's behavior, but these rules are not required for this pattern to emerge. Rather, this behavior is the result of four of the five correct rules that, together, specify that (a) total glucose made only increases in the light (rules 1 and 5 in Table 2); (b) total glucose used always increases (rule 3 in Table 2); and (3) total glucose stored = total glucose made - total glucose used (rule 2 in Table 2). Of the 56 workgroups who accepted all four of these required rules, 53 (95%) also accepted both redundant rules. One could make the argument that rules 1 and 5 together could be considered redundant, or rule 3. In such cases the proportion of workgroups accepting the designated "redundant" rules (given acceptance of the other rules) would change to 53 of 65 (82%) and 53 of 62 (85%).

Responses to the redundant rules in the plant growth model suggest that students are likely to agree that a model follows a rule if the model does not contradict that rule, even if that rule is not necessarily part of the model. The distinction between rules that are necessarily part of a model and behaviors that result from those rules in combination is important for understanding emergent patterns. While parsimony is a value in computer programming, it is crucial for scientific modeling, since it shows how complex behaviors of systems can result from simple behaviors of individual actors. We investigate this issue directly in chemical reactions model.

5.2 An Emergent Pattern in the Chemical Reactions Model

Computer models are particularly suited to show emergent phenomena: complex outcomes that result from a group of objects following simple rules. Emergent phenomena, such as a wave in a stadium, are not caused by a single organizer, but rather the distributed actions of individuals who only interact with their immediate neighbors. The increasing temperature in the chemical reactions model (the red line on the graph), is an emergent phenomenon, the result of individual atoms and molecules transforming potential energy to kinetic energy during a combustion reaction (Fig. 2). The model does not directly specify that the temperature must increase, or even that the average speed of the atoms and molecules must increase - it only specifies the behavior of the individual atoms and molecules. However, the same model behavior would result if the underlying model rules *did* specify that temperature must increase. Emergent patterns are a key area where it is important to make a distinction between the *behavior* of a model and the *rules* underlying the model. Using top-down rules instead of rules that support emergence lead to models with less explanatory power (they are less generalizable to other contexts) and also face inherent epistemological challenges: in nature, how are the top-down rules enforced? A foundational step in recognizing an emergent pattern in a model is realizing that an observable behavior is not necessarily the outcome of a single rule that governs the system as a whole, even though such a rule would not be inconsistent with the observed behavior.

To examine if students thought the overall temperature pattern was the result of a single rule, we asked if they agreed with Tom, who states "the model must include a rule that says, 'The temperature must always be higher than 0 degrees Celsius.' " Students indicated if they agreed or disagreed (multiple choice) and then were asked to explain their response. While this rule is not inconsistent with the behavior of the model, its presence would invalidate the model as evidence that the hydrogen combustion reaction is exothermic, because the temperature evidence in the model would come from an arbitrary rule, not one that reflected the physical workings of the reaction. This question aims to direct students' thinking toward the difference between a model's behaviors and its rules.

5.2.1 98% of workgroups cannot distinguish between a model rule and an emergent pattern. The Emergent Pattern question was given to 140 students of one teacher, working individually. (The other classes saw an earlier version in which Tom proposed that the temperature could never decrease. Most students refuted that hypothesis because, while the overall trend was increasing, noise in the data led to occasional decreases, which were observable from the graph). On the multiple-choice item, 84 students (60%) agreed that the model must include such a rule, 49 (35%) disagreed, and 7 (5%) left that section blank (all 7 also left the explanation blank; 1 student agreed on the multiple choice and left the explanation blank).

For students' explanations of their choices, 8 responses (6%) were blank. The coding rubric has seven categories (Table 5). Two categories are for incorrect answers: (1) Not Substantive (off-task or repeating the multiple-choice selection without explaining their reasoning), and (2) Not Scientific (unclear, incomplete, or strongly non-normative reasoning). Two categories are for answers that partially link the rule to the underlying science: (1) Sub-Zero Temperatures are Possible (these responses include assertions that the model included sub-zero temperatures, that sub-zero temperatures existed in the world, or both), and (2) No Disconfirming Evidence (correct observations that nothing in the graph or model contradicted the rule in question). Two categories are for answers that fully link the rule to the underlying science: (1) The Reaction Cannot Occur Below 0 (if the molecules were at or below 0, the combustion reaction would not occur), and (2) The Reaction Can Occur Below 0 (if the molecules were at or below 0, the combustion reaction could still occur). One category is for answers that link the rule to concepts relevant for emergent patterns: Rule Would Be Invalid and/or Unnecessary for the Existing Model (the rule would make the model invalid, and the behavior of the model would be the same with or without the rule). The quotation in title of this paper is in this table (bolded). The complete response is "I agree because it must included rules." Notably, the response indicates that the student believes there must be rules in the simulation, but does not offer any reasoning for determining if a particular rule is part of the simulation.

Table 5. Student responses by category, in order of sophistication, and by multiple-choice response. Students were asked if they agreed with Tom, who states "the model must include a rule that says, 'The temperature must always be higher than 0 degrees Celsius.' " (N = 140). Student responses are presented verbatim.

Categories of Student Explanations, with Examples	Agree	Disagree
Blank: 8 (6%)	1	0
Incorrect: 48 (34%)	**28**	**20**
Not Substantive: off-task or unelaborated multiple-choice selection: 16 • "It doesn't matter the temperature, it matters the color." • "I agree because it is a fact." • "i dissagree because its just not true"	9	7
Not Scientific: unclear, incomplete, or very non-normative reasoning: 32 • "It would be too cold for the molecules and they will die." • **"I agree because it must include rules."** • "That is not absolute zero 0 onthe scale of celvin is absolute zero."	19	13
Partial Link of the rule and the model or the the science: 50 (36%)	**35**	**15**
Sub-Zero temperatures are possible (in the model, real life, or both): 15 • "the temperature can go below zero degrees celsius" • "Because sometimes the temperature is -28 degrees and that's below zero and some places get that cold." • "You are able to go below 0 degrees, it's not a rule."	1	14
No Disconfirming Evidence: the model does not contradict the rule: 35 (27%) • "I agree because on the graph it doesn't show anything that is lower than 0 degrees Celsius. This is so we can see the line on the graph the whole time." • "I agree because on the chart it shows that the temperature is always above 0 degrees Celsius." • "The temperature must always be higher than 0 degrees Celcius, because since the molecules continue to move, it constantly heats up. It's as if it is a routine that the organism is used to."	34	1
Full Link between the rule and the model or the science: 32 (23%)	**20**	**12**

Continued on next page

Table 5: Emergent Pattern Categories, continued from previous page

The Reaction *Cannot* Occur Below 0: the model could not show the reaction at sub-zero temperatures: 19 • "because with no tempture then there is no movement." • "I agree with Tom because the molecules would not move as fast and they most likely wouldn't separate and combine to make water molecules." • "If the molecules aren't above 0 degrees Celsius, the molecules wont move at a speed needed. Causing a set back in the process."	16	3
The Reaction *Can* Occur Below 0: the reaction could occur at (or reach) a lower temperature: 13 • "because if left aloen long enf in a no other energy sores environment the energy will get lower to a sertent point of freesing" [Because if left alone long enough in a no-other-energy-source environment, the energy will get lower to a certain point of freezing] • "It dosen't have to be because they could go slower." • "I agree because if the temperature is cold,the molecules won't move as fast."	4	9
Full Link connecting the rule, emergent patterns, and the model or the science: 2 (1.5%)	**0**	**2**
Rule Would Be Invalid and/or Unnecessary for the Existing Model: 2 (1.5%) • "There are temperatures below 0 Celsius so therefore it wouldn't make sense to only include temperatures above or equal to that in order to get accurate data." • "It doesn't matter if the temperature must be higher than 0 degrees. It will always get hotter."	0	2

5.2.2 Emergent Pattern Discussion. Over 90% of students in the sample could not meaningfully engage with the main issue in the question "Does the given model necessarily include a rule stating that the temperature must be higher than zero degrees Celsius?" This was not because students were off task - only 8% of students left the explanation blank and only 12% of submitted explanations were not substantive. Rather, most students did not distinguish between the idea of a rule being consistent with the model and being necessary for the model. The categories *Sub-Zero Temperatures are Possible* and *No Disconfirming Evidence* categorize responses that look for consistency. In the former category, students assert that sub-zero temperatures are possible, implying that the rule is not consistent with the model (even though sub-zero temperatures are never show in the model). In the latter category, students do not find any evidence of sub-zero temperatures in the model or the graph, implying that the rule is consistent with the model. 50 students (38% of non-blank responses) fall into one of those two categories. Responses in the categories *The Reaction Can/Cannot Occur Below 0* examine a slightly different question: could a similar model show the reaction with temperatures below zero? The 32 students (10% of non-blank responses) in these categories express interesting science ideas, including the difference between absolute zero and zero degrees celsius, but do not directly address emergent patterns. Only two answers touch on ideas that are relevant to emergent patterns and model accuracy: that arbitrary rules not based on the underlying science would make the model inaccurate, and that the pattern of rising temperature would still occur without the rule in question.

These responses suggest that many students may be ready to learn about emergent patterns. 82 students (62% of non-blank responses) discuss evidence that is important for thinking about emergent patterns in this context: that sub-zero temperatures exist in the real world, that sub-zero temperatures may (or may not) affect this reaction, or that the existing model does not contradict the proposed rule. While 7th graders appear ready to learn about emergent patterns, these results indicate that they are unlikely to come up with them on their own. Emergent properties are part of the Next Generation Science Standards [32], within the crosscutting concept of Systems and System Models. However, emergent patterns are not linked to a particular grade level. The teacher confirmed that these students were not taught about emergent patterns in 7th grade, and did not believe they had been taught this topic earlier. It is not surprising that these students could not explain how such a pattern emerged within a computer model. Therefore, we find it promising that many responses indicate readiness to learn about this concept.

The concept of emergent patterns is particularly interesting as an example of the intersection between science, modeling, and computation. Since a computer model can produce similar results with an emergent pattern or a top-down rule, understanding the construction of the model is important for drawing scientific conclusions. A pre-requisite for engaging with these ideas is the computational knowledge of how rules relate to a model's behavior.

5.3 Identifying Rules: Discussion

Model rules are crucial for understanding models as computational artifacts and as representations of science phenomena. When students learn science from models, they are, at a high level, extracting and interpreting the model rules (even if students don't see themselves as doing this). Therefore, tasks in the Computational Modeling Inventory examine the extent to which middle school students can identify model rules when they interact with models. The Rule Sorting and Emergent Pattern results show that students can thoughtfully engage with the concept of model rules. However, these results also show student difficulties with decomposition, algorithms, and abstraction in models. Sorting the rules required students to practice decomposition by examining each object individually rather than considering the model as a whole. Therefore, difficulty sorting individual rules suggests difficulty with decomposition. Thinking that model rules can be self-contradictory, or that every model behavior is the direct result of a single rule, show weaknesses in student understanding of algorithms. Finally, indicating that models include all relevant variables or that they can directly incorporate concepts suggests a misunderstanding of the role and concept of abstraction in modeling. More broadly, this assessment examines computational thinking issues that are both high-level and fundamental: What can computational tools do? How do computational tools work? This assessment explores students' ideas on these questions in a way that is intrinsic to the scientific models. Together, these results suggest that computer models in science are a worthwhile context in which to support computational thinking, and that strengthening computational thinking around these models will enhance students' learning of the science content.

Including incorrect rules in the Rule Sorting task allows for a richer exploration of students' ideas than would have been possible with the correct rules alone. When students build computer models, they can implement rules that directly test a range of ideas, including incorrect ideas. When students interact with pre-built, scientifically correct models, their opportunities to engage with incorrect ideas are more limited. The next section examines using the context of model rules to prompt engagement with incorrect ideas and assess understanding of the science content. Additionally, while the Rule Sorting and Emergent Pattern items examined student ideas about model rules in a formal, explicit way, the items in the next section investigate student ideas about model rules in an informal, more intuitive way: as the low-level behaviors that contribute to overarching relationships.

6 EXPLAINING MECHANISMS: RESULTS AND DISCUSSION

The previous section examined students' thinking about models as computational tools. This section looks at how thinking about computation can affect the quality of students' science reasoning. Additionally, the questions in this section examine rules and algorithm from a different perspective. While the Rule Sorting questions ask if individual rules are present or not, the questions in this section ask what behaviors in the model would change if a given rule was edited.

What kinds of ideas are elicited by asking students to imagine a modified (and scientifically incorrect) computer model as opposed to asking students about the science portrayed by a correct model? Specifically, can students meaningfully engage with questions that ask them to imagine the implementation of a new rule in a model, and can this type of question reveal additional insights on students' understanding of the science content? We ask this research question because when students work with computer models in science, those models typically show correct science content and the students cannot modify the underlying relationships between variables. Allowing students to change what rules a model is following could help students see the connections between model rules and model behaviors, and could improve students' computational thinking, their thinking about the underlying science, and their thinking on the connections between the two. In order for students to deepen their science understanding, students would need to meaningfully engage with a modified model by making predictions for how the new rule would change the model's behavior, and then observing what the actual effect of the new rule is. This section examines the ideas elicited by students' predictions on the effects of a new, scientifically incorrect rule compared to the ideas elicited by students' explanations of correct relationships in a model. In particular, we examine mechanistic reasoning versus other kinds of responses.

6.1 Question Types: Typical vs. New Rule

Students typically learn from computer models in science class by interacting with the models and then explaining relationships between variables they portray. The Computational Modeling Inventory contains these types of typical questions as well as *New Rule* questions. New Rule questions ask what would happen if a programmer changed the underlying code in the model. Since New Rule questions ask about a hypothetical model that students cannot run, students must reason through how the new model would behave. In the Computational Modeling Inventory, the New Rule questions asked what would happen if a scientifically incorrect rule was implemented. We compared students' answers to each type of question to see if the different question types elicited different kinds of ideas.

For the global climate model, students saw two typical questions targeting the same underlying idea but with different phrasing (both were open-response). Students' responses to these two questions were combined and scored together as a single response to *Greenhouse Gas* (Prompts: 1. Press SETUP, then press GO. Watch the model and observe the graphs. Then press RUN FACTORY. Watch the model and observe the graphs. What is the effect of running the factory? 2. Explain how the model shows the relationship between greenhouse gases and the temperature?). When the model is run without greenhouse gases, Earth's temperature is stable and low. When the factory releases greenhouse gases, the temperature rises and stabilizes at a higher level. Students can see evidence for this outcome by observing the temperature graph before and after adding greenhouse gases to the atmosphere. To understand why the temperature rises, students must observe the animation. Greenhouse gases do not increase the temperature directly, but rather through their interactions with infrared radiation. Heat from the Earth is released as infrared radiation. In the model, greenhouse gases reflect this radiation back toward Earth's surface, where it is absorbed as heat again.

The New Rule question for global climate (also open-response) is *Reflect Sunlight* (Prompt: A programmer added a new rule to this model: 'When solar radiation reaches the surface of the earth, the solar radiation will always bounce off.' Given this new rule, what would be the effect upon temperature compared to the original model?). Since the global climate model does not include this rule, students cannot try it out. Instead, they must reason through how the new model would behave: since all solar radiation would be reflected, none would be absorbed as heat. Therefore, the temperature would be much colder than the original model.

The typical question for chemical reactions (open-response) is *Temperature* (Prompt: Press the SPARK button and watch the model and observe the graphs. Explain the relationship between the movement of the molecules and the temperature). The chemical reactions animation starts with hydrogen molecules and oxygen molecules slowly floating around the animation screen. When a spark is added, the molecule that the spark hits starts moving faster, and then any molecules that are hit by that first one also start moving faster, and so on. When a collision occurs at a high enough speed, bonds between two oxygen atoms or two hydrogen atoms can break, and a bond between a single oxygen atom and a single hydrogen atom can form (as long as the oxygen atom is already bonded to at most one hydrogen). When bonds change, the molecules move faster. On the temperature graph, a red line shows the temperature over time. While there is some noise, the general trend shows an increase in temperature as the reaction continues. Students can observe that as the speed of the molecules increases in the animation, the temperature increases in the graph.

The New Rule question for chemical reactions, *Two Hydrogens Bond* was in a Multiple Choice + Explain format (Multiple Choice Prompt: Suppose a programmer adds a new rule that says, 'If a single hydrogen collides with another hydrogen then they bond together.' How will this rule impact the formation of H2O in the model? Multiple Choice Options: It will likely produce more water; It will likely produce less water; The amount of water will remain the same; There is not enough information. Explain prompt: Explain your choice). To infer that a water molecule has formed, students must observe the animation and identify the different molecules based on the key. The model does not show numerically how many molecules of each type are present. However, before the spark is added, all of the molecules are either hydrogen gas or oxygen gas, and when the model is run for about 30 seconds after the spark is added, all or most of the molecules are water.

The typical questions address the key relationships that each model was designed to portray. The key relationship in the Chemical Reactions model is the positive relationship between molecular movement and temperature. The key relationship in the Global Climate model is the positive relationship between greenhouse gases and Earth's temperature. Students could learn these relationships by observing the temperature graphs and animations for each model. In the Chemical Reactions model, the combustion reaction between hydrogen and oxygen produces water. Over the course of the reaction, the molecules move faster and temperature increases. This is an identity relationship since temperature is a measure of the average speed of molecules. In the Global Climate model, temperature rises after greenhouse gases are added to the atmosphere. This is a cause-and-effect relationship, as greenhouse gases effectively trap infrared radiation.

The New Rule questions target some of the mechanisms behind phenomena portrayed in the models. To answer the New Rule questions, students need to interpret aspects of the animations that are not directly portrayed on the graphs. For Chemical Reactions, the new rule question asks how the formation of water would be affected if single hydrogens bonded with each other whenever they collided. To answer correctly, students must notice that water only forms when individual hydrogen atoms bond to oxygen: oxygen cannot bond to two hydrogens that are already bonded to each other. Further, when two hydrogens are bonded together, energy is required to break that bond. Noticing these details is necessary for students to fully understand the combustion reaction. In the reaction, energy is released as weaker bonds in H_2 and O_2 are broken and stronger bonds in H_2O are formed. We did not expect students to understand this reaction in such depth from simply exploring the model on this assessment. However, students' responses to this question did reveal what they noticed. Noticing what happens is necessary for students to then explore and understand why it happens.

For Global Climate, the new rule question asks how temperature would be affected if all solar radiation were reflected. To answer correctly, students must notice that the solar radiation that is reflected from Earth's surface escapes to outer space. Solar radiation that is not reflected is absorbed as heat, which ultimately is re-emitted as infrared radiation. Therefore, if all solar radiation were reflected, the temperature would be much lower. Understanding the role of solar radiation in producing infrared radiation is crucial for understanding the relationship between greenhouse gases and temperature: without infrared radiation, the relationship would not exist.

6.2 Scoring

For all four questions, responses were coded by the first author for outcome (identifying the relationship in the model or the result of the new rule, coded as correct or incorrect) and for reasoning (explaining the intermediate mechanisms behind that relationship or outcome, coded as absent, partial/non-normative, or fully correct). All rubric categories, with examples, are presented in their respective results sections below. Mechanisms for each question were scored as fully correct if they offered a scientific explanation for why the correct outcome would occur; they were scored as partially correct/non-normative if any other kind of explanation for the relationship was given (including for incorrect relationships).

6.3 Global Climate Results

For Global Climate, 168 workgroups answered both the typical question (Greenhouse Gas) and the new rule question (Reflect Sunlight); 40 left both blank; 18 only answered Greenhouse Gas, and 4 only answered Reflect Sunlight. The Global Climate analyses below exclude students who left both questions blank.

6.3.1 Global Climate: Outcomes. Correct outcomes for the Typical question indicate that greenhouse gases and temperature have a positive relationship. Correct outcomes for the New Rule question indicate that temperature would decrease. Determining the outcome was much easier for the Greenhouse Gas question than the Reflect Sunlight question, although neither was trivial. Considering all non-blank responses for each question, 109 of 186 (58%) were correct on Greenhouse Gas, and 53 of 172 (30%) were correct on Reflect Sunlight. Of the 168 workgroups who answered both questions, 95 (57%) were correct for Greenhouse Gas, and 53 (32%) for Reflect Sunlight (Table 7). A McNemar test shows that this difference in performance is significant ($p < .01$).

That the typical question was easier is not surprising, because students could use the graphs to answer it. However, this result illuminates a challenge for students in understanding one of the key causal mechanisms in the model: that all of Earth's thermal energy (heat and infrared radiation) ultimately comes from Earth's absorption of solar radiation, which does not occur when solar radiation is reflected. While 57% of workgroups correctly indicated a positive relationship between greenhouse gases and atmospheric temperature, only 24% of workgroups were correct on both questions, suggesting that many students are detecting an overall cause-and-effect relationship in the model without identifying the intermediate mechanisms that the relationship depends on. That is, students are detecting a relationship between greenhouse gases and temperature, but are not understanding the role of solar radiation: the absorption of solar radiation and the re-emission of that energy as infrared radiation is a necessary precondition for the relationship between greenhouse gases and temperature. Exploring a model that implemented this incorrect rule could help students identify gaps in their own reasoning.

6.3.2 Global Climate: Mechanisms. Workgroups were more likely to include mechanisms when responding to the New Rule question (49/172, 28%), compared to the typical question (26/186, 14%). Rubric categories and sample responses for these items are shown in Table 6. Of the 168 workgroups who answered both questions, 29% (49/168) included mechanisms for the New Rule question, and 14% (24/168) did so for the typical question (Table 8). A McNemar test shows that this difference in including mechanisms is significant ($p < .01$). That the New Rule question was more likely than the typical question to elicit mechanisms is especially notable because the New Rule question did not specifically ask students to explain the mechanism behind the new outcome.

Table 6. Categories of responses to the Greenhouse Gas (typical) and Reflect Sunlight (new rule) questions for the global climate model. Percentages are out of the whole sample, N = 230.

Categories and Verbatim Student Responses for 1) Greenhouse Gas (Typical) and 2) Reflect Sunlight (New Rule)	Typical	New Rule
Total Explanations with Mechanisms	**11% (26)**	**21% (49)**
Fully Correct, Typical: Greenhouse gases trap/reflect infrared radiation, raising temperature	7	11
(1) Th greenhouse gases coming out of the factory trap the infrared radiation inside the atmosphere and redirect them back towards the Earth. This made the temperature rise at a fast pace.		
Fully Correct, New Rule: With the new rule, solar radiation would not get converted to heat (and/or, would not ultimately convert to infrared radiation), so temperature would decrease		
(2) It would cool down because it would not convert light energy into heat so there wouldn't be as much heat.		
Correct outcome with non-normative or partial mechanism:	11	21
(1) The model shows the relationship because the greenhouse gases keep the heat inside the atmosphere.		
(2) The tempurature would become cooler because there would be less heat.		
Missing or non-normative outcome, with correct, partial, or non-normative mechanism:	8	17
(1) The greenhouse gases are a lot more warmer than the actual tempater.		
(2) The solar radiation will never put heat in the earth.		
Total Outcomes without Mechanisms	**43% (98)**	**24% (55)**

Continued on next page

Table 6: Categories and Sample Responses for Global Climate, continued from previous page	Typical	New Rule
Correct outcome:	91	21
(1) The more greenhouse gases, the higher the temperature (2) The temperature would go down		
Incorrect outcome:	7	34
(1) The greenhouse gasses makes the temperature go down (2) The temperature of the atmosphere may increase but the temperature of the Earth itself will most likely decrease.		
Total Other Substantive Responses	**13% (31)**	**7% (15)**
Observations:	31	3
(1) The effect of running the factory is the concentration of green house gases increased. (2) some of the solar energy bounces of the earth and some enters it. we use that energy for plants and stuff like that.		
Misunderstanding the new rule:	N/A	12
(2) solar radiation will never reach earths air		
Vague, nonsensical, or incomplete i dont know	**13% (31)**	**23% (52)**
Blank	**19% (44)**	**26% (59)**

6.3.3 Global Climate: Student Ideas. The mechanisms in students' responses revealed students' thinking much more than the outcomes alone. While the range of possible outcomes was limited (temperature increases, decreases, or is not affected), the range of ideas in students' mechanisms was broad. The mechanisms elicited by the typical question encompassed many interesting ideas, including that greenhouse gases create solar radiation (*The greenhouse gases increases the atmospheres temperature because it creates more solar radiation*); conflating infrared radiation and heat transfer through a medium (*The model shows the relationship because the greenhouse gases keep the heat inside the atmosphere*); suggesting that greenhouse gases remove infrared radiation from the atmosphere (*The effect of the running factory is that some of the pollution gets rid of some of the infrared radiation*); and hypothesizing that heat from the factory itself causes climate change (*The energy from the factory produces heat that goes into the air*). These kinds of answers, which demonstrate students' reasoning, are much more informative for assessing knowledge than answers which only state the direction in which the temperature will change. These kinds of answers also provide a rich starting point for discussions. Of workgroups who answered both questions, 14% included mechanisms when responding to the Typical question. Just over twice that, 29%, did so for the New Rule question.

Workgroups demonstrated many creative ideas in their mechanisms for the New Rule question. Some ideas indicated the partial understanding that solar radiation produces heat, but also included non-normative ideas, such as conflating heat within Earth's interior and atmospheric temperature (*The temperature will drop because the solar radiation used to convert to heat*) and failure to distinguish between reflection and absorption (*When the solar radiation hits the surfuse it leaves heat the ground*). Other responses indicated confusion on the relationship between solar and infrared radiation, and the properties of each, such as suggesting that solar radiation in the atmosphere increases atmospheric temperature, as infrared radiation does (*The earths temperature would become colder. and the solar radiation would be only in the atmosphere causing it to become very hot*), and indicating that infrared radiation can be produced without absorption of solar radiation (*No it wouldn't because the infrared radiation would still be made*). Other responses conflated convection currents with temperature measurements (*The temperature will rise because heat rises*) or indicated that the new model would not be accurate (*It will remove heat and infrared radiation, making the model invalid*). Since students may arrive at the same conclusion through different kinds of reasoning, understanding students' reasoning is crucial for accurate assessment and productive instruction. Student responses from these two Global Climate questions suggest that asking students to make predictions about the effect of a new, incorrect rule in a model may elicit more science reasoning than typical questions. This pattern also holds for the Chemical Reactions questions. These results show the value of using a computational lens to probe students' science learning: considering how the relationships are set up in the model led students to think more deeply about the science.

6.4 Chemical Reactions Results

For Chemical Reactions, the Typical question asked students to explain the relationship between the movement of the molecules and the temperature (Open Response). The New Rule question asked, "Suppose a programmer adds a new rule that says, 'If a single hydrogen collides with another hydrogen then they bond together.' How will this rule impact the formation of H2O in the model?" (Multiple Choice + Explain). Adding the new rule would reduce the amount of water produced, since, in the existing model, hydrogens can only bond to oxygen when they are not bound to each other, and it requires energy to break the bonds of two hydrogens. With the same amount of energy added to the system by the spark, the outcome of this rule would be less water forming. We first asked students a multiple choice question about the outcome of the rule, and then we asked them to explain their choice. Open Response answers to the typical question were coded for the relationship (that molecular speed and temperature increase together) and for a mechanism (any reasoning to explain why this happens; the normative explanation is that temperature is a measurement of molecular motion, and both increase because the combustion reaction releases energy). For the New Rule question, the Multiple-Choice answers were used for scoring and analyzing outcomes, and the Explain portion was coded for mechanisms (Table 7).

Of the 238 total workgroups, 198 answered both the Temperature (Typical) and Hydrogens Bond (New Rule) questions; 27 left both blank; 2 workgroups answered Temperature but only the multiple-choice section of Hydrogens Bond; 3 workgroups answered Temperature but did not answer any portion of Hydrogens Bond. The analyses below exclude students who left both questions blank. We note that the key to the chemical reactions model (Figure 2) inadvertently reversed oxygen and hydrogen. Many workgroups appeared not to notice, and we did not penalize the few workgroups who indicated that water was made of two oxygens and one hydrogen. Also, the new rule would have the same outcome for a similar reason if it were applied to oxygen instead of hydrogen. Rubric categories and examples for both questions are shown in Table 7.

6.4.1 Chemical Reactions: Outcomes. As with the Global Climate questions, students determined the outcome for the Typical question more easily than for the New Rule question. Considering the non-blank responses to each question, 117/203 (58%) were correct for Temperature, and 48/200 (24%) were correct on the multiple-choice selection for Hydrogens Bond. Considering only the 200 workgroups who answered both the Temperature question and the Multiple-Choice part of the Hydrogens Bond question, 58% of workgroups were correct for Temperature, while only 24% were correct for Hydrogens Bond (Table 7). A McNemar test shows that the difference in performance between the two questions is significant ($p < .01$).

That the Typical question is easier is not surprising, since students could find the answer by running the model. However, the extent of students' difficulty with determining the outcome of the New Rule question is notable. As with the Global Climate responses, the outcomes in students' answers suggest that they detected a general pattern in the model (the speed of the molecules has a positive relationship with temperature), but did not notice key model behaviors (that water only forms when single hydrogen atoms bind to oxygen, and that oxygen never binds to H_2). Similar to the results with Global Climate, even though it was harder for students to determine the outcome for the New Rule question, it elicited more scientific reasoning.

6.4.2 Chemical Reactions: Mechanisms. While the Hydrogens Bond question was more difficult, workgroups were more likely to propose a mechanism with that question compared to the Temperature question. Considering all of the non-blank responses for each question, 105/198 (53%) of the Hydrogen Bond question included a mechanism, while only 9/203 (4%) did so for Temperature. Considering only workgroups who answered both questions, only 5% of workgroups included mechanisms for the Temperature question, while 53% did so for the Hydrogen Bond question (Table 7). A McNemar test shows that the difference in the inclusion of mechanisms is statistically significant ($p < .01$).

Table 7. Categories of responses to the Temperature (typical) and Two Hydrogens Bond (new rule) questions for the chemical reactions model. Percentages are out of the whole sample, N = 230.

Categories and Sample Responses for 1) Temperature (Typical) and 2) Two Hydrogens Bond (New Rule)	Typical	New Rule
Total Explanations with Mechanisms	**4% (9)**	**46% (105)**
Fully Correct, Typical: the formation of water is an exothermic reaction (explaining molecular movement and temperature as an identity relationship would also have been scored as fully correct)	1	2
(1) When the hydrogen atoms connect with the oxygen molecules it causes a chemical reaction that results in H2O. My theory is that the connection of the three atoms releases some heat energy on contact.		
Fully Correct, New Rule: less water will be produced because oxygen atoms can only bond to single hydrogen atoms		
(2) Oxygen molecules cannot be on Binding molecules only on separate molecules.		

Continued on next page

Table 7:Categories and Sample Responses for Chemical Reactions, continued from previous page	Typical	New Rule
Correct outcome with non-normative or partial mechanism, OR, Correct mechanism with incorrect outcome	8	36
(1) The movement of the molecules causes it to make friction, and heats the temperature in the other graph up. When the movement of the molecules moves faster, it heats up more. (2) It will produce less water because the more mass the slower, so the molecules would not go rapid and will slowly produce less water.		
Missing or non-normative outcome with non-normative mechanism:	0	67
(2) if a hydrogen bonds with another hydrogen it's easier for them to bond with oxygen and make more water (2) If 2 hydrogen collide is will most likey make a bigger hydrogen. There fore producing more water.		
Total Outcomes without Mechanisms	**51% (118)**	**13% (30)**
Correct outcome:	108	3
(1) The temperature goes up as the molecules make H2O. (1) The more they move the hotter it gets. (1) When the temperature increases the molecules move faster. (2) If there i s less water molecules there will be less water.		
Incorrect outcome:	10	27
(1) when the tempture is higher the molucules move slower. (2) I think more water will be produced.		
Total Other Substantive Responses	**22% (51)**	**7% (17)**
Observations:	51	1
(1) The temperature goes up when the spark hits, and the molecules move around fastly. (2) The graph increases.		
Misunderstanding the new rule:	N/A	15
(2) There is not enough information because it doesn't tell what happens with the oxygen. (2) Because hydrogen and oxygen create water, so if you add more hydrogen, it will join with the oxygen and create more water.(H20)		

Continued on next page

Table 7:Categories and Sample Responses for Chemical Reactions, continued from previous page	Typical	New Rule
Vague, nonsensical, or incomplete More water = better chance of an organism to survive.	**11% (25)**	**20% (46)**
Blank	**12% (27)**	**14% (32)**

6.4.3 Chemical Reactions: Student Ideas. Since only 9 workgroups included mechanisms for the Temperature question, there were only five distinct ideas for why molecular movement has a relationship with temperature. Two ideas apply everyday reasoning to an atomic scale (each proposed by more than one workgroup): that heat is caused by friction when molecules move (*If the molecules are fast then it will be warm because of the friction*), and that atoms get warmer when they move, like humans do (*The more movement heats them up like a human gets warm exercising not staying still*). The remaining ideas for Temperature proposed that heat is created when molecules break (*As one hydrogen molecule breaks apart, the temperature goes up*), form (*The molecules touch and it will create molecules that will heat it*), or collide (*the temperature rises when te molecules hit each other*). Each of these three ideas express partial understanding of why the combustion reaction produces heat. The ideas in these mechanisms lend themselves to further discussion, and provide a much fuller picture of student thinking than answers which only include outcomes (*Molecules move faster when the temp is high*).

Proposed mechanisms for the New Rule question revealed many interesting student ideas on the formation of water in the model. Many responses indicated partial understanding of science concepts, including that water is made from two hydrogens and one oxygen, but without recognition of how those atoms are connected (*two hydrogen molecules are needed to create a water molecule so the two hydrogen molecules collide, it is a step closer to becoming a water molecule*). These responses focus on the number of atoms and ignore how they are bonded. Other responses recognized that two hydrogens are necessary to form water, and ignored the oxygen (*When the two fuse, it creates a H20 molecule making it, water*).

Other responses reveal confusion around the ways that atoms and molecule combine, suggesting that when two of the same atom bond, nothing changes (*If a hydrogen collides with another hydrogen it will still be a hydrogen*), or it creates a larger atom (*If they bond together then it will be just like adding two together so they will be working together as a bigger size so it will be the same*), or that these larger atoms can then form larger molecules (*We think it will produce more water because it will combine and make one big H2O. We think this because if you add two things together that can bond together, then it will probably make it bigger.*) These ideas seem based on students' normative ideas about the physical world, and express confusion about which ideas apply to the atomic scale. For example, when two drops of water combine, they form one large drop of water; it is not unreasonable for students to apply this thinking to individual atoms. Some responses liken the chemical reaction to cell division, indicating that water molecules can make copies of themselves (*It will produce less water because the double hydrogen will take up more room and the H2O in the model will have less room to duplicate*).

Other responses with mechanisms include correct ideas that are not elaborated fully: that after two hydrogens bond, there will be fewer single hydrogen atoms (*When they combine there would be less hydrogen*), that H_2 is not the same as water (*I think it will produce less water because if two hydrogen molecules bond they will not create water but another object*), and that the way oxygen binds to hydrogen is important (*The oxygen cant conect to hydrogen*).

The proposed mechanisms for the New Rule question reveal much more about students' thinking than the outcomes alone (*If a single hydrogen collides with another hydrogen they bond together, they will create more water*). Further, many workgroups arrived at the same outcomes for different reasons. When answers propose mechanisms, either normative or non-normative, rather than just outcomes, they provide a better opportunity to understand what a student is thinking, and offer clearer paths for instruction suited to the student's needs. While the mechanisms that students proposed for each question were interesting, workgroups were much more likely to propose them at all in response to the New Rule question.

6.5 Explaining Mechanisms: Discussion

Comparing the responses to the Typical and New Rule questions shows that focusing on computational ideas can elicit deeper ideas about the science. Asking students to make predictions about the implementation of incorrect rules led to more scientific reasoning about mechanisms than simply asking students about a causal relationship portrayed in a correct model. This pattern was evident for both model contexts, with twice as many workgroups proposing mechanisms with the New Rule question compared to the Typical question for Global Climate (29% vs. 14%) and ten times as many workgroups doing so for Chemical Reactions (53% vs. 5%). This result was not because the New Rule questions were easier. Rather, it was harder for students to determine the correct relationship or outcome for the New Rule questions compared to the Typical questions (32% vs. 57% correct for Global Climate, and 24% vs. 58% for Chemical Reactions).

Students' mechanistic ideas are valuable because they reveal students' thinking on why a phenomenon happens, in terms that make sense to them [41]. When a student proposes a mechanism, even for an idea that is not fully correct, a teacher can leverage that thinking to help the student better understand the correct idea [41]. Eliciting ideas, normative or not, is important for learning, since ideas which are not elicited cannot be inspected or challenged (a key tenet of the Knowledge Integration Framework; [23]). Further, encouraging mechanistic reasoning is important for enculturating students in the practice of science [41].

While mechanisms are valuable, prior work has demonstrated the difficulty in designing questions that elicit them from middle school students. Work on a similar (but slightly more complex) climate model found that, over three tasks (all Typical), 45% of responses consisted of correct outcomes without mechanisms, while only 11% of responses included mechanisms (students were in middle and high school; 13% of responses were misinterpretations of the model and were not differentiated between outcomes and mechanisms) [34].

Why did the New Rule questions elicit more mechanistic reasoning? The New Rule questions differed from the Typical questions in two main ways: they directed students' attention to individual rules rather than overall patterns, and they asked students to image a new model that did not exist. First, the underlying model rules determine which mechanisms are present in a model and how they are represented, so focusing on rules may help students think about mechanisms more generally. Second, since students were asked about a hypothetical model, they could not test or observe the relationship directly, as they could with the Typical questions. Once students found the outcomes for the Typical questions, they may not have thought it necessary to identify the intermediate actions that linked the initial cause and final effect. However, with the New Rule questions, students needed to reason through what the effect of the rule would be in order to decide how the new rule would affect the outcome. These two hypotheses have different implications for instructional design. The first implies that any activity that focuses students on individual rules will be beneficial. The second implies that students should not be able to test outcomes directly until they have made detailed predictions.

Responses to the New Rule questions suggest many benefits for instruction that allows students to engage with incorrect model rules. Focusing on individual rules may help students think about underlying mechanisms rather than overall cause and effect. Making predictions about incorrect rules illustrates students' thinking about the underlying science. Testing those predictions with a model that implements incorrect rules may solidify students' understanding of the science. Further, students' answers to the New Rule questions revealed some understanding that model rules can create indirect effects (e.g., changing the behavior of solar radiation will affect temperature, even though a direct effect on temperature is not specified). Helping students distinguish between direct and indirect effects may provide a foundation from which to build understanding of emergent patterns. These results show that middle school students can integrate their understanding of science with their computational thinking, although they will need support for each knowledge base to do so correctly. Students' responses also show how they are thinking about computation. In particular, students' predictions of what would happen when a rule was changed seemed to be based in their understanding of the science, not the other rules in the model.

7 IMPLICATIONS FOR INSTRUCTIONAL DESIGN

Questions in the Computational Modeling Inventory linked computational thinking and science content knowledge through model rules, a strategy that yielded rich student ideas. The assessment revealed that exploring pre-built models is a fruitful context for computational thinking, and that questions with a computational lens elicit important student thinking about the science. With pre-built models, students may dive in to exploring the science without explicitly noticing the technology. Not needing to think about the technology is generally a goal for good design [33, 55]. However, for students to understand the science content and the models' limitations, they need think of the models as computational tools, and they need to know the basics of how computational tools work and what they can and cannot do. The goal of instruction on computational thinking with models should be help students approach computational models with a computational lens.

Students with a computational lens would use their knowledge of how technology works to help them understand a new model. They would approach the model as working from programmed, logical rules which can be guessed by observing the behavior of the model under different conditions. They would recognize that the model cannot implement two contradictory rules. They would identify which variables are represented in the model and which are left out. They would attribute the model's behavior to the programmed rules, and would not reference unoperationalized concepts in their explanations of what the model was doing. On this assessment, students engaged with the concept of model rules, but revealed misconceptions and gaps in their thinking about how computational tools work. This work shows that using computational thinking to think explicitly about computational tools is important when students use technology, not just when they make it.

7.1 Teaching Computational Thinking with Models

Our questions on the Computational Modeling Inventory revealed that middle school students hold many non-normative ideas about algorithms and abstractions, and that they struggle with decomposition and systems thinking. Instruction in computational thinking should target these areas. The types of assessment presented here could be augmented with automated feedback and serve as instructional activities. For example, in a Rule Sorting task, students can be alerted if they select contradictory rules. Critique activities, such as identifying which variables are represented in the model, and which relevant variables are not included, may also help students recognize which rules a model is following. Finally, instruction should also allow students to engage with incorrect rules, which can elicit student reasoning and provide a contrasting case that draws attention to the correct rule. Rule Sorting may also offer possible interfaces for students to create and modify models, by offering students a set of rules which could be turned on or off. Such interfaces could allow students to test incorrect rules without programming.

In particular, these results indicate that middle school students do not recognize emergent patterns in computer models, but they demonstrate many skills that suggest preparedness for doing so. Students considered: that the data generated by the model was consistent with the proposed rule, how the model should behave under different conditions (e.g., a different starting temperature), and that sub-zero temperatures exist in the real world. All of these ideas show that students are making connections among model rules, model behaviors, the purpose and function of models, and real-world experiences related to the phenomenon. Instruction on computational thinking in science can build on this foundation and introduce more sophisticated concepts.

Questions in the Computational Modeling Inventory also showed that students could thoughtfully answer questions about model rules. For middle-school science, rules should be written at a level that emphasizes the key computational or scientific goals of the lesson. Therefore, the rule *When the light is on, total glucose made increases* will be acceptable for lessons on the basics of model rules and glucose production, but a lesson on operationalization may ask students to compare more detailed rules, such as *For every day the light is on, total glucose made increases by X* where X is a natural number. Designers may infer rules from a model's behavior when the designer does not have access to the underlying code. Inferring rules is complicated when dealing with emergent patterns. However, lessons can teach the concept of emergent patterns and propose several low-level rules that would produce the same behavior as a high-level rule regardless of how the model was built.

Finally, the Computational Modeling Inventory elicited rich student ideas using only pre-built models. This demonstrates that specialized environments are not necessary to engage students in computational thinking. While programming and model-building environments allow for a greater range of activities, designing instruction around pre-built models is a promising approach for integrating computational thinking into science classrooms.

7.2 Enhancing Science Models with Computational Thinking

Results from the Computational Modeling Inventory show the benefit of a computational perspective when students work with science models. The Rule-Sorting questions reveal what students think the model is doing in a more precise way than the science-only, open-response items. With Rule Sorting, it is clearer which behaviors a student is observing in a model. Students must observe behaviors and recognize them as functions of the model before students can synthesize those behaviors and reflect on what they mean. When creating Rule Sorting items, instructional designers should consider which model behaviors are most important and which are likely to be overlooked or incorrectly inferred. The New Rule questions also revealed science ideas that the science-only explanation items did not. In particular, the New Rule questions showed that students held many non-normative science ideas even after interacting with the correct model (e.g., conflating solar radiation and heat). The New Rule questions prompted students to think about a counterfactual scenario, showing the value of having students engage with incorrect ideas. The New Rule questions were successful at revealing students' thinking because they elicited mechanisms. Framing models as tools which follow a set of rules may help students distinguish between low-level rules and overall outcomes, which is important for explaining how an overarching relationship results from intermediate mechanisms. The questions from the Computational Modeling Inventory shows that computational thinking can help students engage more deeply with the science, and that students do not require decontextualized, non-science lessons on computational thinking to do so.

Since models are representations of scientific thinking, we argue that understanding model rules is an important aspect of *metarepresentational competence*: the ability to implement, use, evaluate, and revise representations [9]. Just as students often come to science class with prior conceptions of the target content, diSessa and Sherin [9] found that students have prior conceptions about representations. Importantly, they argue that students' knowledge about representations interacts with their sense-making, in particular by allowing students to separate social conventions from logical constraints (e.g., it is a social convention for higher y-values on a graph to indicate "more," but that is not an inherent requirement of graphs) [9]. Questions in the Computational Modeling Inventory indicated that students have varied prior conceptions of the logical constraints of computer models, and that students' poor abilities to identify model rules may hamper meaningful critique.

Prior work has found that when middle-school students create models in the context of a curriculum focused on metarepresentational competence (also called *metamodeling knowledge* in this context), they can develop sophisticated knowledge about what models are for, how models are constructed, and meaningful ways in which models can be critiqued [21, 44, 45]. Therefore, we hypothesize that instruction focused on model rules can enhance students' metarepresentational competence, computational thinking, and science learning.

A skill related to metarepresentational competence is *representational fluency*: "the ability to reason with and among multiple representations" [31, p. 3]. An important distinction among types of representations are those that show a single instance (e.g., a row in a table, a point on a graph) versus representations that are holistic (e.g., an equation that shows a relationship, a graph of a function) [31]. Holistic representations, like computer models, are necessary for fully representing complex relationships. However, Nathan et al. show that for middle school students, learning to reason with instance-based representations precedes reasoning with holistic representations [31]. We hypothesize that model rules are useful instance-based representations of computer models, and that we can help students reason about computer models as a whole by helping them reason about their component parts.

7.3 Limitations and Future Work

Results from questions on the Computational Modeling Inventory show promise for integrating computational thinking and science. While the Computational Modeling Inventory explored the assumptions students make about the computation behind models, it did not assess learning. Future research should explore the relationships among instruction in computational thinking, the ability to analyze the rules behind a model, and learning of the science content. Future work should also examine different instructional approaches, including those based on the item types in the Inventory.

This paper presents results from 7th grade students at two schools. Future work should include a broader range of schools and grade levels, as students' level of experience with models may affect how they respond. Some student responses indicated misunderstanding of the questions, so future work should explore how to word these types of items. While the assessment was done as part of students' science classes, it was not graded, and students may not have been trying their best. Therefore, the results may underestimate students' knowledge.

Results from the Computational Modeling Inventory also point to new questions for instructional design at the intersection of science and computational thinking: How should instruction convey to students that models are governed by rules? How should instruction support students in thinking about these model rules as they explore pre-built models and begin to create their own? How can instruction position students' varied ideas as opportunities for learning more about the science content and about computation? Future work could explore ways to redesign model interfaces to strengthen connections between a model's rules and its observable behaviors. For example, a model could show which behaviors result from a specific rule, and instruction could contrast models governed by different rules. Additionally, the rules selected for study in this paper were designed at a fairly abstract level. Further research is needed to compare methods for communicating rules, including using a programming language to articulate the rules.

8 CONCLUSIONS

The Rule Sorting, Emergent Patterns, and New Rule items from the Computational Modeling Inventory successfully elicited student ideas on the connections between model rules and science content, and offer ways to measure computational thinking with science models. These items revealed that 7th grade students are not fluent in the computational thinking skills that are required to fully interpret a model. Only 14% of workgroups correctly identified all of the observable rules in the Rule Sorting tasks; 65% indicated that a model could follow contradictory rules; 57% thought a model took into account more variables than it did; 68% could not distinguish between a science concept and the operationalization of that concept in a model; and 98% could not distinguish between behaviors that directly result from model rules and those that are emergent patterns. While 7th graders are not fluent in these skills, they engaged that tasks meaningfully and were moderately successful on some of them (with average scores of 58% and 66% correct on the two Rule Sorting items). Therefore, the tasks seem appropriate for middle school. We anticipate that students would improve with instructional support.

The New Rule questions demonstrated advantages for asking students to consider hypothetical, inaccurate model rules. Such questions can elicit deeper science reasoning than Typical questions that focus on cause-and-effect relationships portrayed by the model. Consistent with the knowledge integration framework, New Rule items often elicited intriguing student ideas about the scientific phenomena, including ideas that were not supported by the pre-built (correct) model. Students often used the pre-built model to test their ideas yet failed to reconcile the differences between their ideas and the ones depicted in the model. They need additional guidance to reflect on their tests with the model and to use this information to revise their insights. These results offer promise for instruction that focuses on distinguishing between accurate and inaccurate rules in the context of pre-built models. Instruction featuring New Rule questions can reveal gaps in student understanding and may enhance students' science learning.

Computational thinking skills with model rules are essential for students to analyze the implications of dynamic, scientific models. Specifically, students need to use *decomposition* when they examine each element in the model individually and determine how it behaves in relationship to other elements. Students need to use *abstraction* when mapping simplified elements and behaviors in the model to the complex real-world entities and actions they represent. Students need to mentally analyze the *algorithms* that could generate the outcomes shown in the model. Student success in these areas is reflected in their ability to identify a non-contradictory set of rules, to determine if the model follows or violates a rule, and to realize that the computer model is a simplification of the scientific phenomena. Furthermore, computational thinking expertise is reflected in student success in recognizing when relevant variables are not included in a model and in distinguishing between a correct science concept and the implementation of that concept in a model. The questions in the Computational Modeling Inventory show how these computational thinking skills manifest in the context of pre-built science models, and demonstrates the importance of these skills in using and interpreting computer models of science phenomena.

ACKNOWLEDGMENTS

We thank our partner schools and teachers. The work was supported by the National Science Foundation under Grant No. DRL-1418423 (GRIDS: Graphing Research on Inquiry with Data in Science) and INT-1451604 (PLANS: Project Learning with Automated, Networked Supports).

REFERENCES

[1] 2018. PhET Interactive Simulations. https://phet.colorado.edu/

[2] Tanya N. Beran, Alejandro Ramirez-Serrano, Roman Kuzyk, Meghann Fior, and Sarah Nugent. 2011. Understanding how children understand robots: Perceived animism in child-robot interaction. *International Journal of Human Computer Studies* 69, 7-8 (2011), 539–550. https://doi.org/10.1016/j.ijhcs.2011.04.003

[3] Elizabeth L Bjork and Robert Bjork. 2009. Making Things Hard on Yourself, but in a Good Way: Creating Desirable Difficulties to Enhance Learning. In *Psychology and the Real World: Essays Illustrating Fundamental Contributions to Society*, Morton Ann Gernsbacher, Richart W. Pew, Leatta M. Hough, and James R. Pomerantz (Eds.). Worth Publishers, New York, NY, 55–64.

[4] Hsin-yi Chang and Marcia C Linn. 2013. Scaffolding Learning From Molecular Visualizations. *Journal of Research in Science Teaching* 50, 7 (2013), 858–886. https://doi.org/10.1002/tea.21089

[5] The Concord Consortium. 2013. Molecular Workbench. http://mw.concord.org/modeler/showcase/

[6] Alan Cooper. 1996. Three models of computer software. *Technical Communication* 43, 3 (1996), 229–236.

[7] Maria Cutumisu, Cathy Adams, and Chang Lu. 2019. A Scoping Review of Empirical Research on Recent Computational Thinking Assessments. *Journal of Science Education and Technology* 28, 6 (2019), 651–676. https://doi.org/10.1007/s10956-019-09799-3

[8] Janet Davis and Samuel A Rebelsky. 2007. Food-first computer science: starting the first course right with PB&J. *SIGCSE '07: Proceedings of the 38th SIGCSE technical symposium on Computer science education* (2007), 372–376. https://doi.org/10.1145/1227310.1227440

[9] Andrea A DiSessa and Bruce L Sherin. 2000. Meta-representation: an introduction. *Journal of Mathematical Behavior* 19 (2000), 385–398.

[10] Benedict Du Boulay. 1986. Some Difficulties of Learning to Program. *Journal of Educational Computing Research* 2, 1 (1986), 57–73.

[11] Ilenia Fronza, Nabil El Ioini, and Luis Corral. 2017. Teaching Computational Thinking Using Agile Software Engineering Methods. *ACM Transactions on Computing Education* 17, 4 (2017), 1–28. https://doi.org/10.1145/3055258

[12] David J. Gilmore. 1995. Interface Design: Have We got it wrong? In *Human-Computer Interaction: Interact '95*, Knut Nordby, Per Helmerson, David J Gilmore, and Svein A Arnesan (Eds.). Springer US, Boston, MA, 173–178. https://doi.org/10.1007/978-1-5041-2896-4_29

[13] Janice D. Gobert and Amy Pallant. 2004. Fostering Students' Epistemologies of Models via Authentic Model-Based Tasks. *Journal of Science Education and Technology* 13, 1 (2004), 7–22. https://doi.org/10.1023/B:JOST.0000019635.70068.6f

[14] Ashok K Goel and David A Joyner. 2014. Computational Ideation in Scientific Discovery: Interactive Construction, Evaluation and Revision of Conceptual Models. In *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*. Quebec City, Canada, 27–34.

[15] David Golightly. 1996. Harnessing the interface for domain learning. In *Conference Companion on Human Factors in Computing Systems (CHI '96)*. ACM, 37–38. https://doi.org/10.1145/257089.257121

[16] Shuchi Grover and Roy Pea. 2013. Computational Thinking in K-12: A Review of the State of the Field. *Educational Researcher* 42, 38 (2013), 38–43. https://doi.org/10.3102/0013189X12463051

[17] F. Heider and M. Simmel. 1944. An experimental study of social behavior. *The American Journal of Psychology* 57, 2 (1944), 243–259.

[18] Benjamin Herold. 2016. Technolog in Education: An Overview. https://www.edweek.org/ew/issues/technology-in-education/

[19] Daiki Isayama, Masaki Ishiyama, Raissa Relator, and Koichi Yamazaki. 2016. Computer Science Education for Primary and Lower Secondary School Students. *ACM Transactions on Computing Education* 17, 1 (2016), 1–28. https://doi.org/10.1145/2940331

[20] David Klahr and Sharon Carver. 1988. Cognitive Objectives in a LOGO Debugging Curriculum: Instruction, Learning, and Transfer. *Cognitive Psychology* 20, 3 (1988), 362–404.

[21] Joseph Krajcik and Joi Merritt. 2012. Engaging Students in Scientific Practices: What does constructing and revising models look like in the science classroom? *Science and Children* 49, 7 (2012), 10–13.

[22] Marcia C Linn and Bat-Sheva Eylon. 2011. *Science Learning and Instruction: Taking Advantage of Technology to Promote Knowledge Integration*. Routledge. https://books.google.com/books?id=IWmpAgAAQBAJ

[23] Marcia C Linn, H S. Lee, R Tinker, F Husic, and J L Chiu. 2006. Teaching and Assessing Knowledge Integration in Science. *Science* 313 (2006), 1049–1050. https://doi.org/10.1126/science.1131408

[24] Ou Lydia Liu, Hee Sun Lee, Carolyn Hofstetter, and Marcia C. Linn. 2008. Assessing knowledge integration in science: Construct, measures, and evidence. *Educational Assessment* 13, 1 (2008), 33–55. https://doi.org/10.1080/10627190801968224

[25] Yanjin Long and Vincent Aleven. 2017. Educational Game and Intelligent Tutoring System: A Classroom Study and Comparative Design Analysis. *ACM Transactions on Computer-Human Interaction* 24, 3 (2017), 1–27. https:

//doi.org/10.1145/3057889

[26] Loucas T Louca and Zacharias C Zacharia. 2012. Modeling-based learning in science education: cognitive, metacognitive, social, material and epistemological contributions. *Educational Review* 54, 4 (2012), 471–492. https://doi.org/10.1080/00131911.2011.628748

[27] Betti Marenko. 2014. Neo-Animism and Design. *Design and Culture* 6, 2 (2014), 219–242. https://doi.org/10.2752/175470814x14031924627185

[28] Santosh Mathan and Kenneth R. Koedinger. 2005. Fostering the Intelligent Novice: Learning From Errors With Metacognitive Tutoring. *Educational Psychologist* 40, 4 (dec 2005), 257–265. https://doi.org/10.1207/s15326985ep4004_7

[29] Kevin W. McElhaney, Hsin-Yi Chang, Jennifer L. Chiu, and Marcia C. Linn. 2014. Evidence for effective uses of dynamic visualisations in science curriculum materials. *Studies in Science Education* 51, 1 (2014), 49–85. https://doi.org/10.1080/03057267.2014.984506

[30] Mitchell J Nathan. 1998. Knowledge and Situational Feedback in a Learning Environment for Algebra Story Problem Solving. *Interactive Learning Environments* 5, 1 (1998), 135–159.

[31] Mitchell J Nathan, Martha W Alibali, Kate Masarik, Ana C Stephens, and Kenneth R. Koedinger. 2010. Enhancing middle school students' representational fluency: A classroom-based study. (2010), 29 pages. https://wcer.wisc.edu/docs/working-papers/Working{_}Paper{_}No{_}2010{_}09.pdf

[32] NGSS Lead States. 2013. *Next Generation Science Standards: For States, by States.* Technical Report. The National Academies Press, Washington, DC. https://doi.org/10.17226/18290

[33] Donald A. Norman. 1990. *The Design of Everyday Things.* The MIT Press, London, England.

[34] Amy Pallant and Hee Sun Lee. 2015. Constructing Scientific Arguments Using Evidence from Dynamic Computational Climate Models. *Journal of Science Education and Technology* 24, 2-3 (2015), 378–395. https://doi.org/10.1007/s10956-014-9499-3

[35] Roy D Pea. 1986. Language-independent conceptual "bugs" in novice programming. *Journal of Educational Computing Research* 2, 1 (1986), 25–36.

[36] Edys S Quellmalz, Michael J Timms, Matt D Silberglitt, and Barbara C Buckley. 2012. Science Assessments for All: Integrating Science Simulations Into Balanced State Science Assessment Systems. *Journal of Research in Science Teaching* 49, 3 (2012), 363–393. https://doi.org/10.1002/tea.21005

[37] Noel Rappin, Mark Guzdial, Matthew Realff, and Pete Ludovice. 1997. Balancing usability and learning in an interface. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. 479–486. https://doi.org/10.1145/258549.258995

[38] Alexander Repenning, David C Webb, Kyu Han Koh, Hilarie Nickerson, Susan B Miller, Catharine Brand, Ian Her Many Horses, Ashok Basawapatna, Fred Gluck, Ryan Grover, Kris Gutiérrez, and Nadia Repenning. 2015. Scalable Game Design: A strategy to bring systemic computer science education to schools through game design and simulation creation. *ACM Transactions on Computing Education* 15, 2 (2015), 1–31. https://doi.org/10.1145/2700517

[39] Jeremy Roschelle. 1995. Learning in Interactive Environments: Prior Knowledge and New Experience. In *Public institutions for personal learning: Establishing a research agenda*, J.H Falk and L.D Dierking (Eds.). American Association of Museums, Washington, DC, 37–51.

[40] Marco Rozendaal. 2016. Objects with Intent: A New Paradigm for Interaction Design. *Interactions* 23, 3 (2016), 62–65.

[41] Rosemary S. Russ, Janet E. Coffey, David Hammer, and Paul Hutchison. 2009. Making classroom assessment more accountable to scientific reasoning: A case for attending to mechanistic thinking. *Science Education* 93, 5 (2009), 875–891. https://doi.org/10.1002/sce.20320

[42] Kihyun Ryoo and Marcia C Linn. 2012. Can Dynamic Visualizations Improve Middle School Students' Understanding of Energy in Photosynthesis? *Journal of Research in Science Teaching* 49, 2 (2012), 218–243. https://doi.org/10.1002/tea.21003

[43] Brian J. Scholl and Patrice D. Tremoulet. 2000. Perceptual causality and animacy. *Trends in Cognitive Sciences* 4, 8 (2000), 299–309. https://doi.org/10.1016/S1364-6613(00)01506-0

[44] Christina V Schwarz, Brian J Reiser, Elizabeth A Davis, Lisa Kenyon, Andres Acher, David Fortus, Yael Shwartz, Barbara Hug, and Joe Krajcik. 2009. Developing a Learning Progression for Scientific Modeling: Making Scientific Modeling Accessible and Meaningful for Learners. *Journal of Research in Science Teaching* 46, 6 (2009), 632–654. https://doi.org/10.1002/tea.20311

[45] Christina V Schwarz and Barbara Y White. 2005. Metamodeling Knowledge: Developing Students' Understanding of Scientific Modeling. *Cognition and Instruction* 23, 2 (2005), 165–205.

[46] Pratim Sengupta and Amy Voss Farris. 2012. Learning kinematics in elementary grades using agent-based computational modeling. In *Proceedings of the 11th International Conference on Interaction Design and Children - IDC '12*. 78–87. https://doi.org/10.1145/2307096.2307106

[47] Pratim Sengupta, John S. Kinnebrew, Satabdi Basu, Gautam Biswas, and Douglas Clark. 2013. Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and*

Information Technologies 18, 2 (2013), 351–380. https://doi.org/10.1007/s10639-012-9240-x

[48] Pratim Sengupta, Amy Voss Farris, and Mason Wright. 2013. From agents to continuous change via aesthetics: Learning mechanics with visual agent-based computational modeling. *Technology, Knowledge and Learning* 17, 1-2 (2013), 23–42. https://doi.org/10.1007/s10758-012-9190-9

[49] Juha Sorva. 2013. Notional machines and introductory programming education. *ACM Transactions on Computing Education* 13, 2 (2013). https://doi.org/10.1145/2483710.2483713

[50] John Sweller. 1994. Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction* 4, 4 (1994), 295–312. https://doi.org/10.1016/0959-4752(94)90003-5

[51] Kurt VanLehn, Greg Chung, Sachin Grover, Ayesha Madni, and Jon Wetzel. 2016. Learning Science by Constructing Models: Can Dragoon Increase Learning without Increasing the Time Required? *International Journal of Artificial Intelligence in Education* 26, 4 (2016), 1033–1068. https://doi.org/10.1007/s40593-015-0093-5

[52] Jonathan M Vitale, Elizabeth Mcbride, and Marcia C Linn. 2016. Distinguishing complex ideas about climate change: knowledge integration vs. specific guidance. *International Journal of Science Education* 38, 9 (2016), 1548–1569. https://doi.org/10.1080/09500693.2016.1198969

[53] Kevin P. Waterman, Lynn Goldsmith, and Marian Pasquale. 2020. Integrating Computational Thinking into Elementary Science Curriculum: an Examination of Activities that Support Students' Computational Thinking in the Service of Disciplinary Learning. *Journal of Science Education and Technology* 29, 1 (2020), 53–64. https://doi.org/10.1007/s10956-019-09801-y

[54] David Weintrop, Elham Beheshti, Michael Horn, Kai Orton, Kemi Jona, Laura Trouille, and Uri Wilensky. 2016. Defining Computational Thinking for Mathematics and Science Classrooms. *Journal of Science Education and Technology* 25, 1 (2016), 127–147. https://doi.org/10.1007/s10956-015-9581-5

[55] Mark Weiser. 1991. The Computer for the 21 st Century. *Scientific American* 265, 3 (1991), 94–105. https://doi.org/10.2307/24938718

[56] Astrid Weiss, Daniela Wurhofer, and Manfred Tscheligi. 2009. "I love this dog" - children's emotional attachment to the robotic dog AIBO. *International Journal of Social Robotics* 1, 3 (2009), 243–248. https://doi.org/10.1007/s12369-009-0024-4

[57] Eliane Stampfer Wiese, Hannah Gogel, Libby F Gerard, Jonathan M Vitale, and Marcia C Linn. 2017. *Probing Middle-School Students' Understanding of Computer Models*. Presented at the 2017 Annual Meeting of the American Educational Research Association, San Antonio, TX.

[58] Uri Wilensky and Kenneth Reisman. 2006. Thinking Like a Wolf, a Sheep, or a Firefly: Learning Biology Through Constructing and Testing Computational Theories—An Embodied Modeling Approach. *Cognition and Instruction* 24, 2 (2006), 171–209. https://doi.org/10.1207/s1532690xci2402_1

[59] Michelle H. Wilkerson-Jerde, Brian E. Gravel, and Christopher A. Macrander. 2015. Exploring Shifts in Middle School Learners' Modeling Activity While Generating Drawings, Animations, and Computational Simulations of Molecular Diffusion. *Journal of Science Education and Technology* 24, 2-3 (2015), 396–415. https://doi.org/10.1007/s10956-014-9497-5

[60] Jeannette M Wing. 2006. Computational Thinking. *Commun. ACM* 49, 3 (2006), 33–35. https://doi.org/10.1145/1118178.1118215

[61] Eben B. Witherspoon, Ross M. Higashi, Christian D. Schunn, Emily C. Baehr, and Robin Shoop. 2017. Developing Computational Thinking through a Virtual Robotics Programming Curriculum. *ACM Transactions on Computing Education* 18, 1 (2017), 1–20. https://doi.org/10.1145/3104982