Received: 25 May 2020 | Revised: 16 September 2020 | Accepted: 2 October 2020

DOI: 10.1002/tea.21674

RESEARCH ARTICLE

JRST | WILEY

Impact of computer modeling on learning and teaching systems thinking

Ha Nguyen | **Rossella Santagata**

School of Education, University of California, Irvine, California

Correspondence
Ha Nguyen, School of Education, University of California, 3200 Education Bldg., Irvine, CA 92617.
Email: thicn@uci.edu.

Abstract

Researchers have found that computer modeling fosters the learning of causal mechanisms in systems, an important crosscutting concept in science that many novice learners find challenging. Despite the research that highlights the role of teacher's instructional practices in enacting computer tools, few studies have considered teachers' use of computer modeling and its implications for student learning in classroom interactions, compared to interactions without computer tools. In this study, we examine (a) the impact of computer modeling on students' understanding of causal links in decomposition and (b) classroom interactions with use of computer modeling. We employed a quasi-experimental design with eight middle school science classes that served predominately Latinx students. The random treatment was at the class level (computer modeling; $n = 60$, four classes) and control (paper modeling; $n = 59$, four classes). Analyses incorporated student preassessment and postassessment, classroom observations, and audio-recorded modeling instruction. Results indicate that compared to paper modeling, computer modeling enriched systems thinking, particularly students' ability to provide causally coherent statements in explaining scientific ideas and evidence. Enactment of computer modeling may be associated with a shift in classroom interactions to include more invitation for students' elaboration of causal systems. We discuss aspects of computer modeling that may foster systems thinking, with implications for the future design of tools and curricula.

KEYWORDS

computer modeling, instruction, quasi-experiment, systems thinking

1 | INTRODUCTION

How do microorganisms impact plant growth? Many young learners would initially answer this question with linear causal links, such as "microorganisms create nutrients." These answers may not demonstrate what current science standards consider high-level systems thinking—reasoning about how systems components interact (NGSS, 2013). For example, students with more elaborate systems thinking would articulate that "decomposers return the carbon in organisms to the atmosphere, continuing the carbon cycle." Such complex thinking is key for learners to be able to engage in scientific reasoning, such as understanding patterns, cause and effect, and interdependence among phenomena, across grade levels and beyond.

Computer modeling may help enrich systems thinking by highlighting the central phenomena of a system (Damelin, Krajcik, Mcintyre, & Bielik, 2017; Dickes, Sengupta, Farris, & Basu, 2016; Jordan et al., 2018) and deepening students' argumentation around systems mechanisms (Damelin et al., 2017; Rosenberg & Lawson, 2019). Computer tools, such as data logging systems, graphing tools, simulation, and modeling environments, may also influence learning by affording changes in classroom interactions (Kim, Hannafin, & Bryan, 2007; Maor & Taylor, 1995). However, few studies examine how teachers' and students' talk shifts when using modeling tools, compared to pedagogical enactments of similar lessons without tool use (for a review, see Yoon, Goh, & Park, 2018). Educators' approaches to using the same tools may differ, leading to questions about how uses of computer tools relate to student learning (Puntambekar, Stylianou, & Goldstein, 2007).

The purpose of our study is to examine the extent to which use of computer modeling deepened middle school students' understanding of interconnected systems and facilitated classroom instructions. The study was conducted with middle school students because researchers have found a strong correlation between baseline understanding of systems and later development in system thinking, and thus recommended introduction to systems thinking in early grades (Assaraf & Orion, 2005). In this study, we conducted a quasi-experimental study with eight classes that served predominately low-income, Latinx students. The classes were randomly divided into treatment (computer modeling; n students = 60) and control (paper modeling; n = 59). This study design allowed us to evaluate the effect of computer modeling on systems thinking development, as measured by pretest and posttest. Involving the same teachers in enacting two different modeling approaches allowed us to examine differences in classroom interactions between conditions, through use of observation and audio recording.

2 | THEORETICAL FRAMEWORK

2.1 | Systems thinking about decomposition

Science instruction rarely involves causal relations between organisms, behaviors, and mechanisms when introducing systems concepts such as ecosystems to young learners (Hayes, Plumley, Smith, & Esch, 2017; Hmelo-Silver, Jordan, Eberbach, & Sinha, 2017; Hmelo-Silver, Marathe, & Liu, 2007; Wilensky & Resnick, 1999). Complex systems are often broken down into linear chains, as opposed to more sophisticated causal patterns such as cyclic (i.e., events occur in a circle), domino-like (i.e., multiple effects result from causes), or reciprocal (i.e., causes become effects and vice versa; Hayes et al., 2017). Younger learners thus tend to lean toward sequential causality: they might say, "more predators decrease prey populations," instead of reasoning about reciprocal feedback loop, "predators will eventually decrease if preys die out" (Reiner & Eilam, 2001).

Consider an example from decomposition, our study's core idea. Few students between the age of 5 and 16 could link knowledge about photosynthesis and feeding to the decomposition process (Leach, Driver, Scott, & Wood-Robinson, 1995). Challenges in understanding decomposition rest on the inability to identify invisible causes and effects that are distant in time and space, such as the activity of microorganisms (Sweeney & Sterman, 2007). When introduced to the larger system that entails decomposition, producers, and consumers, students have difficulties understanding cyclical causal links (Barman, 1994). They can articulate linear links, such as producers create food, but may not recognize how decomposers break down dead matter into nutrients for producers.

Such naïve understanding of systems affects development of complex systems thinking in later grades. More complex systems thinking at the middle school level is associated with higher-level approaches to problem solving in science in later grades (Assaraf & Orion, 2009). Researchers have thus examined the promise of integrating systems in middle grades, pointing to suggestive evidence that explicitly teaching about systems models helped improve student reasoning (Jacobson & Wilensky, 2006; Plate, 2010). In particular, modeling practices can help students to express scientific ideas as physical or diagrammatic representations to generate predictions and explanations of systems (Schwarz et al., 2009).

2.2 | Modeling and systems thinking

Engaging learners in modeling-centered inquiries goes beyond making sense of existing models; it invites learners to construct representations of scientific ideas, develop hypotheses and explanations, and revise their models with new data and understanding (Schwarz et al., 2009). Researchers have used paper-and-pencil modeling as instructional and assessment activities in systems thinking curricula, to reveal growth in students' understanding of systems interactions over time (Hmelo-Silver et al., 2007; Hmelo-Silver & Pfeffer, 2004). For example, Eberbach, Hmelo-Silver, Jordan, Sinha, and Goel (2012) examine students' paper modeling artifacts at two time points (prior to and following an instructional unit), and find that the artifacts illustrate students' enhanced knowledge in macrointeractions and biotic dimensions of systems. Paper models, however, may be difficult for students to revise (Hwang, Wu, & Kuo, 2013). In addition, two-dimensional models on paper may not represent elements and processes that are distant in time and space. Students more often focus on individual components, rather than the interrelations between microinteraction and macrointeraction within systems (Komis, Ergazaki, & Zogza, 2007).

Researchers have thus explored the affordances of computer modeling in introducing systems. Computer modeling presents an opportunity for learners to articulate and simulate the systemic concepts and ideas that are too abstract to comprehend (Jacobson & Wilensky, 2006; Wilensky & Resnick, 1999). Computer interfaces can have built-in data analysis and simulation tools for students to test hypotheses and solutions, assess models' predictions against real-world data or expert simulations, and design conceptual links (Basu, Dickes, Kinnebrew, Sengupta, & Biswas, 2013; Louca & Zacharia, 2012; Sengupta, Kinnebrew, Basu, Biswas, & Clark, 2013; Weintrop et al., 2016; Xiang & Passmore, 2015). Through these activities, learners begin to articulate phenomena not in terms of simple, linear links, but as causal feedback loops (Hmelo-Silver et al., 2017).

Educators have leveraged computer modeling to develop students' representational skills, content knowledge, and engagement in scientific practices to enhance causal thinking (Damelin et al., 2017; Jordan et al., 2018). To explore the affordances of computer modeling, researchers have examined students' written work, think-aloud protocols when interacting with tools, and interviews (Damelin et al., 2017; Novak & Krajcik, 2019; Rosenberg & Lawson, 2019; Wilensky & Resnick, 1999). Researchers have also compared learning outcomes in instruction with computer modeling tools and instruction without the tools (e.g., Hmelo-Silver et al., 2017; Plate, 2010; van Borkulo, van Joolingen, Savelsbergh, & de Jong, 2012). Findings from these studies suggest that computer modeling may facilitate students' iterative construction of evidence-based and causally coherent explanations of systems components (Novak & Krajcik, 2019; Rosenberg & Lawson, 2019).

While prior studies have mostly focused on student learning outcomes, emergent research has noted that modeling may also influence learning through teachers' use of tools during instruction and facilitation of classroom interactions (Hmelo-Silver, Liu, Gray, & Jordan, 2015). There remains a need for research that explores the differences in classroom interactions when using computer tools compared to typical instruction (Kim et al., 2007), and the learning opportunities that follow teachers' different tool enactments (Hmelo-Silver et al., 2015).

2.3 | Tool use and classroom interactions

Prior work on classroom interactions between teachers and students with use of computer tools suggests two propositions. First, teachers' initial enactment of educational technology relates to their existing pedagogical beliefs about structuring classroom participation as more procedural or more inquiry-driven (Looi, Sun, Seow, & Chia, 2014; Sun, Looi, & Xie, 2014). Second, engagement with computer tools may deepen teachers' understanding of the subject matter and ways that they notice, reason about, and support student ideas (Kim et al., 2007; Maor & Taylor, 1995). Evolving engagement with tool use in turn motivates teachers to modify classroom interactions, moving from direct instruction to student-driven exploration of concepts (Levin & Wadmany, 2006; Murphy & Rodriguez-Manzanares, 2008).

In this study, we conceptualize classroom interactions as teacher and student discourse when using computational tools. We build on the Vygotskian conception that discourse serves as a means of knowledge construction (Mercer, 2007). Discourse that emphasizes student elaboration is particularly important in our study context to foster development of systems thinking through modeling activities (Hmelo-Silver et al., 2015; Puntambekar et al., 2007). This is because teacher discourse that creates opportunity for student elaboration has been associated with productive student engagement in modeling practices (Alonzo & Gotwals, 2012). "How" and "why" questions create opportunities for students to reason about evidence and causal links and develop deeper understanding of scientific phenomena (Otero & Graesser, 2001). In addition, teacher discourse that is conversational, as opposed to limited to facts-oriented questioning, produces more complex student interaction patterns in scientific sensemaking (Ahern, Peck, & Laycock, 1992). Students are more likely to engage in scientific practices—iteratively generating questions, collecting data, and refining their explanations—when they view the modeling activity as a sensemaking process (Alonzo & Gotwals, 2012).

The role of classroom interactions in science learning holds true in technology-mediated environments. Puntambekar et al. (2007), for example, observed classrooms that used the same scientific concept mapping tool, and found that students whose teacher spent more time prompting for elaboration on how the tools connected to scientific principles showed significantly better learning, compared to those whose teacher focused on procedural talk.

To summarize, when using computer tools, teachers may deepen knowledge of interconnected systems and guide classroom interactions toward elaboration of scientific phenomena. However, most studies on classroom interactions are structured as case studies of different teachers using tools in instruction (Hmelo-Silver et al., 2015; Looi et al., 2014; Puntambekar et al., 2007; Sun et al., 2014), rather than examining how teachers' and students' talk differs with and without tool use. Thus, it is difficult to draw conclusions on how engagement with technology interacts with teachers' existing pedagogical beliefs about effective ways to structure classroom interactions. Finally, few studies have examined teachers' use of computer tools and their facilitation of student participation in the context of instruction centered on systems thinking. We turn to our conceptual framework to address this gap.

2.4 | Conceptual framework: Understanding causal links in systems

This study builds on two bodies of research: (a) frameworks on systems thinking and (b) assessment of causal reasoning. Systems thinking frameworks attend to components and mechanisms, while those on causal reasoning conceptualize how to understand the depth of explanations. Incorporating both bodies of work provides a more comprehensive view of how students and teachers come to understand systems dynamics and apply evidence to support their reasoning and classroom interactions.

A common theme among systems thinking frameworks is the focus on systems components and the links among them (Chandler & Boutilier, 1992; Hmelo-Silver et al., 2007; Wilensky & Resnick, 1999). For example, Hmelo-Silver et al. (2007) propose the structure–behavior–function (SBF) framework. Structure describes the visible, physical components of the systems (e.g., plants, fungus). Behavior encompasses the actions in a system to achieve a goal or outcome. Behaviors include both visible (e.g., feeding) and invisible (e.g., decomposition) mechanisms. These behaviors serve a set of functions—the purpose of the system. For example, decomposition functions in reproduction of organisms.

Frameworks on causal reasoning in scientific explanations (Braaten & Windschitl, 2011; Kang, Thompson, & Windschitl, 2014) are complementary to systems thinking frameworks in that they underscore the depth of student understanding. We choose to integrate causal reasoning models into our analysis because decomposition involves dynamic causal mechanisms. Attending to causal reasoning highlights the role of evidence and causal coherence, which demonstrates a logical chain that connects student explanations, evidence, and ideas (Kang et al., 2014). Combining views on systems component and causal coherence, our framework includes the following dimensions: element, evidence, and causal coherence. In addition to student learning, our conceptual framework can be linked to the teaching of systems thinking about decomposition.

Element encompasses the system components and processes that students can identify, and is equivalent to Structure and Behavior in the SBF framework (Hmelo-Silver et al., 2017; Hmelo-Silver & Pfeffer, 2004). Elements can be living organisms (e.g., animal), nonliving components (e.g., sun, water), or mechanistic processes (e.g., water cycle, decomposition). Students who score low in the elements dimension may focus only on observable elements, whereas students with a high score would describe the elements that are not visible or processes that take longer to unfold. Classroom interactions that emphasize element would highlight the components and processes in the systems.

Evidence refers to the extent to which students use evidence from data. This criterion is reflected in Kang et al.'s (2014) framework, pertaining to the extent to which students use data. The evidence dimension can be characterized as (a) explanations with no evidence use, (b) explanations that mention some forms of evidence, but the evidence is inappropriate or insufficient, and (c) explanations supported by data that are directly linked to scientific phenomena (Kang et al., 2014). Teachers may attend to evidence by guiding students to use their observations or the data they collect to support their hypotheses.

Finally, causal coherence describes the extent to which causal explanations reveals a coherent and logical chain of reasoning that connects observable elements and unobservable processes or disciplinary scientific ideas. Examples of ideas in learning about decomposition include understanding of the interdependent relationships in ecosystems, of cycles of matter and energy transfer, and of ecosystem dynamics and functioning (NGSS, 2013). Causal coherence captures whether explanations are consistent with the data (internally valid) and with scientific ideas (externally valid). The criterion aligns with the interactions between functions and behaviors in the SBF framework (Hmelo-Silver et al., 2017). Causal coherence can be divided into (a) explanations with no causal claims, (b) explanations that show partially coherent causal links, but inconsistent connection to scientific ideas, and (c) coherent explanations that link explanations, evidence, and scientific concepts (Kang et al., 2014). Classroom interactions that underscore causal coherence would involve teachers' scaffolding or modeling to students how to make explicit and complex connections between element and evidence.

3 | RESEARCH QUESTIONS

A limitation to prior work on the benefits of computer modeling is that most do not involve a comparison group when examining learning outcomes or classroom interactions (Yoon et al., 2018). We address these gaps in the extant research by employing a quasi-experimental design, exploring the following questions:

> **RQ1.** How does computer modeling influence students' causal thinking of interconnected systems, compared to paper modeling?

> **RQ2.** To what extent does computer modeling afford different classroom interactions that support systems thinking, compared to paper modeling?

4 | METHODS

4.1 | Participants

Participants were 119 sixth graders (11–12 years old; 91% Hispanic, 92% on Free and Reduced Lunch, 43% Female; 44% English learners) and two teachers (Annie and Peter; pseudonyms) in eight classes. The classes practiced one-to-one laptop, and students were familiar with using the computers to research and write scientific reports.

Both teachers volunteered to teach the curriculum for the first time. Annie majored in law and switched from a business career to teaching sciences. She had been a science teacher for 19 years. Peter had been teaching science subjects for 3 years as his first job. Peter taught five participating classes, while Annie taught three classes.

4.2 | Attrition

Data were collected from students who consented and finished both assessments. Attrition was because the posttest was administered right before the summer break, and many students either missed school or did not want to complete the test amidst other year-end activities. In total, 169 students finished either or both tests, and 119 finished both (attrition for control: 31.40%; treatment: 27.71%).

We conducted sensitivity analyses to understand the required effect size, given the probability of false positive, power, and sample size. Sensitivity analyses of the actual sample yielded $d = .51$. The analyses would have enough power to detect medium to large effects, but would be underpowered to detect small effects (Hill, Bloom, Black, & Lipsey, 2008). Notably, t test revealed that the students who did not remain in the sample between pretest and posttest differed from the final sample at pretest, element, $t(169) = 3.54$, $p < .001$; evidence: $t(169) = 2.62$, $p = .01$; causal coherence, $t(169) = 2.93$, $p = .004$. Limitations about attrition were noted in Section 6.

4.3 | Curriculum

4.3.1 | Collaborative curriculum design

This study is part of a larger project, Project Crystal Code, structured as a research–practice partnership between state park educators, a school district, education researchers, and environmental biology scientists to develop a middle school participatory science curriculum (McKinley et al., 2017). The study took place in the Southwestern U.S. The partnership engaged participants in distributed knowledge and scientific practices while contributing to plant restoration efforts. By anchoring scientific concepts in a local context, the project aimed to engage students in authentic, student-driven learning.

The curriculum was designed around combining data literacy (i.e., understand, gather, and analyze data) with computer modeling to study plant restoration. The students experimented with how different mulch conditions (no mulch, "woody"; i.e., bark, and "straw-like" mulch; i.e., dry mustard) affected decomposition rate and soil moisture. Teachers were involved as curriculum developers, alongside scientists, park educators, and researchers. At the first codesign session, the teachers reflected on their students' and their own limited exposure to technologies. The group consequently revised the curriculum to include brainstorming paper diagrams as a scaffolding activity before transitioning to modeling. Teachers received professional development to familiarize themselves with the activities. Park educators and researchers were in the classrooms to assist teachers with the modeling lessons. The curriculum lasted 3 weeks, and the experiment (the modeling activity) was over two consecutive days.

4.3.2 | The modeling activity

Prior to the modeling activity, students in both conditions received the same lessons on the park ecosystem, the mulch experiment, and hypothesis generation about decomposition rate and soil moisture in each mulch condition. The eight classes were then randomly divided (using a number generator) into treatment and control groups at the class level (n treatment students = 60, four classes; n control = 59, four classes). The order was as follows: treatment (Peter's period 1 and 7; Annie's period 2 and 4); control (Peter's period 3, 4, and 6; Annie's period 3). The period was the consecutive class that the teachers taught. Figure 1 shows examples of the conditions.

Brainstorming activity (20 min)

All students in both conditions engaged in a brainstorming activity, where they worked in groups of three to four to generate a list of the components and processes of ecosystems. Student groups then selected 7–10 components and processes from the list to create an initial paper model (10 min). The model responded to the question: How do different mulch conditions affect the decomposition rate of native leaf?

Paper modeling (45 min)

Students worked on the modeling activity in small groups, which were formed at their tables by the teachers prior to the curriculum. Groups were heterogeneous in terms of English proficiency, science proficiency, and gender. The control group students brainstormed and sketched the systems on paper in groups. Student groups then transferred their initial models to poster paper, with instruction to use arrow sizes and directions to highlight the extent to which the components and processes were related.

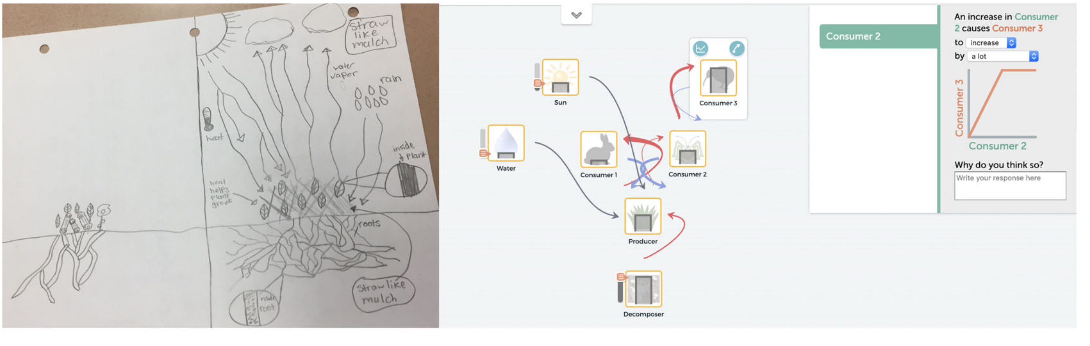


FIGURE 1 Experimental conditions: Paper modeling (left) and computer modeling (right) [Color figure can be viewed at wileyonlinelibrary.com]

Computer modeling (45 min)

Similar to the control group, the treatment group brainstormed the system elements on paper in their preexisting small groups, based on where students were seated. Then, they formed pairs to model the dynamics on computer, using the web-based tool SageModeler (Damelin et al., 2017). The treatment group students defined components and links that showed how components may influence one another using a computer tool. Two added features of computer modeling compared to paper modeling are scaffolds for relation specification and simulation. The interface included scaffolds for students to select the direction of the relation, for instance, whether an increase in soil moisture would cause the number of microorganisms to increase, decrease, or stay the same, and by how much. The specified magnitude and direction of the relations were also reflected in the diagram in colors and sizes. Blue arrows indicate decreasing relationships and red arrows show positive ones. Meanwhile, thicker arrows indicate larger associations. The platform included a Simulation function for students to test their models against theories or existing data. For example, students could simulate how an increase in sunlight would affect the rest of the systems based on the direction and scale they have specified.

4.4 | Data sources

4.4.1 | RQ1. Pretest and posttest

The pretest and posttest were identical. Each was completed in one class period of 45 min. The pretest was administered before students were introduced to the curriculum. The posttest was delivered after students had finished the whole curriculum. Content validation was established through consulting with a panel of experts in science education (one advanced doctoral student in Biology, two science education researchers, and three park educators).

Each test consisted of three questions. The first question asked students to predict what would happen to plant growth under different experimental conditions and provide an explanation for their hypotheses. The second question asked students to fill in the blank to complete a food web and explain how an increase in predators would affect the quantity of nutrients in the soil. The last question provided students with a graph of changes to soil moisture and decomposition rate in different mulch conditions and asked them to recommend to the state park which mulch to use. Each question received separate scores for the three dimensions (element, evidence, and causal coherence), and the scores were added up for all questions to create a composite score for each dimension.

4.4.2 | RQ2. Classroom interactions

We collected observation field notes and audio recording of three paper modeling classroom sessions and three computer modeling sessions (approximately 6 hr of audio; 20 pages of transcripts). The audio, which followed the teachers' talk, was transcribed following data collection. The field notes documented classroom events in chronological order. The observers included the first author and a research assistant trained on the study's coding scheme. The observers also took note of small group interactions of two to three groups that were randomly chosen in each session when students were creating their models. The activities that students carried out in all of the classes were the same for the respective experimental condition, including engaging in brainstorming of the original hypotheses, exploring the SageModeler (treatment condition), creating their models, and discussing their hypotheses.

We provided illustrative examples for the readers about teacher and student talk from two lessons (one computer, one paper modeling) from both teachers, Peter and Annie. The examples illuminate the differences in student–teacher interactions between the two modeling conditions. The presence of an observer in at least two iterations of each condition by the same teachers helped verify that the activities in different classes were parallel.

Our illustrative examples particularly focused on Peter because he taught more classes in total and thus had more opportunities to rehearse his instruction. This helped us to verify whether there was a sustained distinction between computer and paper modeling within teachers. The data that we presented were from Peter's second iteration, which was the second time that Peter taught the paper/computer modeling lesson within the same day, with the computer modeling lesson preceding paper modeling.

4.5 | Analytic strategy

4.5.1 | RQ1. Development of students' systems thinking

Pretest and posttest were coded for elements, evidence, and causal coherence (Table 1). The first author and a trained research assistant coded 25% of the data and reached acceptable agreement on all three dimensions after one round of training and two rounds of reliability check, Cohen's $\kappa = .73, .92, .88$, respectively. The first author coded the rest of the data.

We first examined whether random assignment of classrooms to treatment and control conditions resulted in equivalent groups. We conducted t-test to check group difference at pretest for each dimension of systems thinking (element, evidence, causal coherence). In addition, to check that classrooms were equivalent at pretest, independent of treatment conditions, we conducted analysis of variance (ANOVA) across the eight classrooms on pretest scores.

Next, we examined whether the order that teachers taught the lessons was associated with the posttest scores. We conducted ANOVA for classrooms within treatment (four classes) and control conditions (four classes) to test for differences at posttest. Finally, we computed a series of t-test to check potential differences at posttest by teacher, gender, and English learner status.

Upon checking the assumptions for the randomization and potential predictors of posttest outcomes, we turned to our research question about the difference between treatment groups in systems thinking. We conducted a repeated measure multivariate ANOVA (MANOVA) to assess changes in systems thinking from pretest to posttest, with scores in element, evidence, and causal coherence as the dependent variables. The treatment conditions (computer and paper modeling) served as the between-subjects variable and time (pretest and posttest) was the within-subject variable.

The data met the MANOVA assumptions for absence of univariate and multivariate outliers (assessed by box plots and Mahalanobis distance), normality (univariate: by examining QQ plots; multivariate: Mardia's skew $p = .055$, kurtosis $p = .054$), linearity (scatterplots of dependent variables), absence of multicollinearity (no Pearson's correlation above .80; element-evidence $r = .27$, element-coherence $r = .37$, evidence-coherence $r = .40$; $p < .001$), and homogeneity of variance and covariance (variance by treatment conditions: Levene's test for element: $p = .27$; evidence: $p = .35$, coherence: $p = .54$; covariance: Box's M-test $p = .18$).

Note that students were nested within classrooms and thus violated the assumption of independent observation. However, ANOVA on student baseline systems understanding (pretests)

TABLE 1 Coding scheme for element, evidence, and causal coherence, student assessment

Score		0	1	2
Element	The number (counts) of elements that students list. Elements include those that are biotic (living), abiotic (nonliving), processes (e.g., photosynthesis), or human activities (e.g., fire, pollution).	Few elements 0–1 element identified	Some elements 2–4 factors identified	Many elements 5 or more factors identified
Evidence	The extent to which students provide evidence from direct observation or data to support claims	No evidence I think so because it sounds right.	Provide evidence, without linking to student explanations ... because the graph shows that bark mulch has more soil moisture.	Provide evidence and link to explanations ... because the graph shows that bark mulch has more soil moisture and decomposes faster. Moisture helps plants grow.
Causal coherence	The depth and coherence of connections that students provide between elements and causal mechanisms	No causality Plants need water to grow.	Linear causality We should choose woody mulch because woody mulch will keep the soil cool and moist, will attract more microorganisms and help break down mulch.	Coherent/nonlinear causality ... would have more microorganisms to break down nutrients. The decomposition will increase the nutrients [, and nutrients] would go to the soil helping plants grow. More animals will be attracted to the plants. Animals will die and get decomposed.

Note: Examples come from student responses. We keep the responses as is, and add brackets for grammar to aid readability.

indicated that there were no significant differences across classrooms. Furthermore, the small number of clusters did not guarantee unbiased estimates at the classroom level in mixed-level model (Maas & Hox, 2005). As robustness check, we fitted linear mixed models separately for element, evidence, and causal coherence, with random intercepts and classrooms as the cluster variable to predict the posttest scores, controlling for pretests scores. The positive associations between treatment and evidence and causal coherence posttest scores remain unchanged. Intraclass correlations range is [.00–.05], suggesting that the ratio of between-class variance to total variance was small and the simpler analyses of variance sufficed.

4.5.2 | RQ2. Difference in classroom interactions

Coding process occurred in two stages: generating a codebook and coding. The codebook was generated based on prior literature on systems thinking and causal reasoning (Hmelo-Silver et al., 2017; Kang et al., 2014), as well as analysis of a sample of audio transcripts to examine how the initial codes aligned with teacher and student talk. The first dimension of the code pertains to the three categories in our conceptual framework: element, evidence, and causal coherence.

An additional code dimension was created based on prior work on the relation between student learning and teachers' prompting for student explanation when using computer tools (Hmelo-Silver et al., 2015; Puntambekar et al., 2007). In particular, we coded for whether the teacher's talk created an opportunity for students' elaboration, as these talk moves have been associated with productive students' engagement in modeling (Alonzo & Gotwals, 2012). Table 2 outlines the codes from both dimensions with illustrative examples.

The next phase involved coding the audio transcripts. Ninety-two episodes of teachers' talk units were identified from the audio as pertaining to the causal thinking dimensions. A unit of talk was counted as each turn the teachers took in modeling activities or facilitating discussion. The field notes provided the context of the activities the teachers were engaged in at the point of talk. Each unit received one code for each of the causal dimension and one code for student-driven elaboration (0: code is absent; 1: code is present). Following initial coding of all the data by the first author, a trained research assistant coded a randomly selected subset of 20% of the transcripts. High agreement was reached, Cohen's $\kappa = 1.00$, .93, .88, 1.00 for element, evidence, causal coherence, and invitation for student explanations.

5 | FINDINGS

5.1 | RQ1. Development of students' Systems thinking

We first examined whether random assignment of classrooms to treatment and control conditions resulted in equivalent groups at baseline, between conditions and across classrooms. *t*-Test indicated no significant difference at pretests between treatment and control groups for each dimension (element, evidence, causal coherence; Table 3, *t*-test column, Rows 3–5). ANOVA suggested group equivalence at pretests across classrooms, element: $F(7,111) = .54$, $p = .80$; evidence: $F(7,111) = 1.29$, $p = .26$; causal coherence: $F(7,111) = 1.47$, $p = .19$.

We then examined whether students in both treatment and control conditions improved in their systems thinking. The t test revealed significant difference between pretests and posttest

TABLE 2 Coding scheme for element, evidence, and causal coherence, teacher talk moves

		Close-ended (not inviting elaboration; "what" statements)	Open-ended (inviting elaboration; "why/how" statements)
Element	The extent to which teacher mentioned or prompted students to name system elements, including biotic, abiotic, or social activities.	What is the small thing that help with that [decomposition]?	You said that the soil is getting darker. Why?
Evidence	The extent to which teacher mentioned or prompted students to use evidence from their observation or data to support claims	Sunlight will decrease because mulch blocks sunlight.	From the video, why do you think there is more nutrients in the soil?
Causal coherence	The extent to which teacher mentioned or prompted students to provide causal links, that is, connections between elements and causal mechanisms	If sunlight increases temperature goes up by how much, a lot or a little?	Increase in temperature causes soil moisture to do what, and why?

Note: Example sentences come from recorded teacher talk.

for each condition (Table 3, Rows 9–11). For both conditions, the posttest scores were significantly higher than pretests (paper modeling: evidence: $t(57) = 5.23$, $p < .001$, evidence: $t(57) = 2.78$, $p < .001$, causal coherence: $t(57) = 2.15$, $p < .05$; computer modeling: evidence: $t(58) = 6.61$, $p < .001$, evidence: $t(58) = 6.49$, $p < .001$, causal coherence: $t(58) = 4.15$, $p < .001$). Table 3 reports descriptive statistics and results from the t test.

Next, we examined potential predictors of posttest scores that were not included in the main analysis (MANOVA). We found that the order that teachers taught the lessons was not associated with the posttest scores. ANOVA for classrooms within treatment (four classes) and control conditions (four classes) found no difference in scores across all treatment or all control conditions at posttest, control: element: $F(3,55) = .06$, $p = .48$; evidence: $F(3,55) = .043$, $p = .99$; causal coherence: $F(3,55) = 2.25$, $p = .09$; treatment: element: $F(3,56) = 1.06$, $p = .40$; evidence: $F(3,56) = .86$, $p = .47$; causal coherence: $F(3,56) = 1.51$, $p = .22$. Additionally, at posttest, we found no significant differences by teacher, gender, and English learner status within each treatment and control conditions (Table 4). For example, the statistics for the teacher variable can be interpreted as follows: In the Control (paper modeling) group, there was no significant difference in posttest scores for element, evidence, and causal coherence for students taught by Peter versus Annie, $t(57) = -0.37$, $p = .71$; $t(57) = 0.33$, $p = .74$; and $t(57) = 0.70$, $p = .49$. There was also no significant difference in posttest scores between teachers for the treatment group, $t(58) = 0.15$, $p = .88$; $t(58) = 0.90$, $p = .38$; and $t(58) = 0.21$, $p = .83$.

To answer the main research question about development of systems thinking between conditions, we compared treatment-control group differences from pre to posttest. MANOVA revealed a significant interaction between the computer modeling treatment and time (pretest, posttest) for the combined scores of element, evidence, and causal coherence, $F(3,115) = 2.95$, $p = .04$, $\eta_p^2 = .071$ (medium effect size, Tabachnick, Fidell, & Ullman, 2007). Overall, there was a significant difference in the change from pretests to posttest in systems thinking between modeling groups.

TABLE 3 Descriptive statistics of student pretest and posttest by treatment conditions

		Control (n = 59)		Treatment (n = 60)		t-Test
Control-treatment difference	**Range**	***M***	***SD***	***M***	***SD***	**Treat control**
1. Female (%)	[0%, 100%]	50.850		55.000		
2. English learner status (%)	[0%, 100%]	45.763		38.700		
3. Pretests: element	[0, 6]	2.508	.989	2.517	.873	.048
4. Pretests: evidence	[0, 6]	.915	.934	.867	.747	−.313
5. Pretests: causal coherence	[0, 6]	1.814	1.210	1.950	1.227	.611
6. Posttest: element	[0, 6]	3.525	1.120	3.700	1.078	.866
7. Posttest: evidence	[0, 6]	1.475	1.237	2.050	1.199	2.576*
8. Posttest: causal coherence	[0, 6]	2.339	1.434	3.000	1.529	2.433*
t-Test for pretest–posttest difference, Within each condition		Control (n = 59) t test		Treatment (n = 60) t test		
9. Element		5.229***		6.606***		
10. Evidence		2.772**		6.487***		
11. Causal coherence		2.151*		4.1478***		

Note: *$p < .05$, **$p < .01$, ***$p < .001$. Each question in the pre/posttest received separate scores for the three dimensions (element, evidence, and causal coherence), and the scores for each dimension were added up.

TABLE 4 t-Test for difference by demographics at posttest, within each condition

	Control (n = 59) t test (p)	**Treatment (n = 60) t test (p)**
Teacher: Peter (coded as 1); Annie (coded as 0)		
Posttest: element	−.37 (.71)	.15 (.88)
Posttest: evidence	.33 (.74)	.90 (.38)
Posttest: causal coherence	.70 (.49)	.21 (.83)
Gender: female (1); male (0)		
Posttest: element	.41 (.68)	.74 (.46)
Posttest: evidence	−.16 (.87)	1.46 (.15)
Posttest: causal coherence	.76 (.45)	1.27 (.21)
English learner: English learner (1); non-English learner (0)		
Posttest: element	.36 (.72)	.45 (.65)
Posttest: evidence	−.51 (.61)	.75 (.46)
Posttest: causal coherence	.73 (.47)	.31 (.76)

Note: None of the t-test was significant at $p = .05$.

Given the significant result from the MANOVA, we conducted follow-up ANOVA to examine which dimension (element, evidence, or causal coherence) the two groups differed in. To account for concurrent testing of three hypotheses, we applied the Bonferroni correction to reach the critical value of .017. Follow-up analyses indicated that the interactions between treatment and time were significant for evidence, $F(1,117) = 8.00$, $p = .006$, $\eta_p^2 = .064$ (medium effect size), and causal coherence, $F(1,117) = 6.17$, $p = .014$, $\eta_p^2 = .049$ (small effect size), but not for element, $F(1,117) = .49$, $p = .49$, $\eta_p^2 = .004$ (small effect size).

Finally, to determine the nature and magnitude of differences between treatment groups, we ran a simple effect analysis (i.e., ANOVA between conditions at pretests and posttest). Although there was no difference at pretests, at posttest, students in the computer modeling group cited more evidence and causal links, compared to the paper modeling group, evidence: $F(1,117) = 6.64$, $p = .01$, $\eta_p^2 = .054$ (small effect size), and causal coherence: $F(1,117) = 5.92$, $p = .02$, $\eta_p^2 = .048$ (small effect size; for descriptive statistics, see Table 3, Rows 6–8).

To help contextualize the findings about changes from pre to posttest between conditions, Table 5 presents illustrative student responses in the paper (Pearl, pseudonym) and computer modeling (Jan, pseudonym). Pearl and Jan scored the same at pretests (3 for element, 1 for evidence, 2 for causal coherence) and had posttest scores similar to the average performance of all responses in their respective condition (Pearl/paper modeling: 3 for element, 1 for evidence, 2 for causal coherence; Jan/computer modeling: 3 for element, 2 for evidence, 3 for causal coherence). At pretests, both students did not provide details of the mechanisms behind decomposition. For example, in the third question, both Pearl and Jan used the graph data to support their answer (e.g., "bark has more soil moisture," "plants need water to grow"), but did not explain how decomposition rate could be linked to soil moisture.

At posttest, Pearl's answers remained largely unchanged. Meanwhile, Jan was able to use evidence to support her ideas more substantially (Table 5, Q3). She drew from both soil moisture and decomposition graphs to support her explanations, connecting different rates of water retention in mulch to decomposition. She wrote:

> [If] more moisture the bark mulch to the soil, there would be more microorganisms and plants have water to grow. The reason I think this is the best strategy is that bark mulch is taking a long time to decompose [decomposition graph], but is giving lots of moisture to the soil [moisture graph], unlike the leaf mulch.

This answer suggested emergent understanding that microorganisms need moisture to decompose mulch into nutrients. We also note that Jan showed enhanced causal coherence in her explanation of the food web (Table 5, Q2). Her posttest answer that "all of the food in its waste will give nutrients to the soil" hinted at the carbon cycle, which was absent from her pretests.

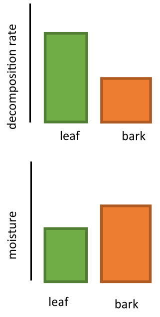
5.2 | RQ2. Difference in classroom interactions

Analyses of classroom observations and audio recording revealed two insights. First, teachers interacted differently with students depending on the tool they used. Both teachers focused more on the systems interactivity in the computer modeling lesson, while they mostly emphasized procedural tasks in the paper modeling lesson. Second, we observed variation in students' interactions between conditions, particularly in the way students problematized the systems phenomena and engaged with modeling.

TABLE 5 Illustrative student responses from pretest to posttest [Color table can be viewed at wileyonlinelibrary.com]

		Pearl (paper/control)	Jan (computer/treatment)
Q1: Predict what would happen to plant growth under different mulch condition.	Pretests	*The mulch in condition* A *will help plant grow more after 1 year because* I think that the rates have more nutrience [nutrients]. The mulch size will help plant grow. Also the mulch of leaves has less leaves but it has more nutrience [nutrients] than the bark. The microorganism will give the nutrience [nutrients] to the roots and that will also help plants grow.	*The mulch in condition* A *will help plant grow more after 1 year because* the decomposition rate will give the soil nutrients after it decomposes, which will then give plants nutrients to help it [them] grow.
	Posttest	*The mulch in condition* A *will help plant grow more after 1 year* because it is more thin.	*The mulch in condition* A *will help plant grow more after 1 year* because leaves [mulch] has more leftover nutrients and moisture and can transfer them to the soil, which then go into the plants for it [them] to grow stronger and healthier.
Q2: Explain the food web: How an increase in predators would affect quantity of nutrients in the soil.	Pretests	*If predators increase*, there could not be that much nutrience [nutrients] in plants. I think so because predator would eat plants and there would not be more nutrience [nutrients].	*If predators increase*, there will be less herbivores to eat plants which will increase the nutrients in the soil. I think so because the herbivores will not be eating so many plants.
	Posttest	*If predators increase*, there could not be that much nutrience [nutrients] in plants because they might eat the plants.	*If predators increase*, then the amount of nutrients in the soil will increase because all of the food in its waste [herbivores] will give nutrients to the soil.
Q3: Recommend which mulch to use, based on graphs of changes to soil moisture and decomposition rate with different mulch.	Pretests	*My recommendation is to* use bark mulch. *The evidence that supports this* is that that bark has more soil moisture and it does not decompose as fast as leaf mulch does. *The reason I think this is the best strategy is that* it has more soil moisture than the leaves mulch.	*My recommendation is to* use bark mulch. *The evidence that supports this* is plants need lots of water to live and the bark mulch has more moisture than the leaf mulch. *The reason I think this is the best strategy is that* the plant will be able to grow because of the amount of moisture in the bark mulch.

TABLE 5 (Continued)

		Pearl (paper/control)	Jan (computer/treatment)
	Posttest	*My recommendation is to* use leaf mulch. *The evidence that supports this* is it decomposes faster than bark mulch. *The reason I think this is the best strategy is that* if there is a drought, I think the leaf mulch has more water.	*My recommendation is to* use bark mulch. *The evidence that supports this* is plants need use water to grow and become healthy and strong. He should also use it because the more moisture the bark mulch to the soil there would be more microorganisms and plants have water to grow. *The reason I think this is the best strategy is that* since the bark mulch is taking a long time to decompose but is giving lots of moisture to the soil, unlike the leaf mulch. The bark mulch is giving nutrients and water to the plant and soil.

Note: Italic text is scaffolding sentence starters in the assessment.

5.2.1 | Focus on components and processes in paper modeling

To explain the impact of computer modeling on student learning, we explored how computer modeling, versus paper, was incorporated into classroom interactions. Overall, observations indicated that the combination of teacher talk related to causal thinking in the computer modeling markedly differed from that in the paper condition for both teachers. Figure 2 outlines the talk sequences in each condition. The figure illustrates that talk about the elements of the system was the most prevalent in the paper condition (83.79% occurrences of elements were close-ended/not inviting for student elaboration). Meanwhile, the teachers talked about causal coherence most frequently in the computer modeling condition. Notably, 80.77% of the talk about causal links in the computer condition was prompting for student elaboration.

Two excerpts (one from each condition) from Peter's classroom illustrate the context and development of teacher talk. In the paper modeling lesson, Peter first reiterated the decomposition process to students through a video that illustrates the process of leaves being broken down. The teacher then facilitated a whole-class discussion for students to break down the components involved in decomposition. Here, most of the questions focused on specific elements and processes, and students' answers contained little elaboration:

> Teacher: **What** do you notice?
> Student 1: I see worms.
> Teacher: **What** are they doing?
> Student 1: Eating leaves?
> Teacher: **What** happens there? **What** are decomposers?

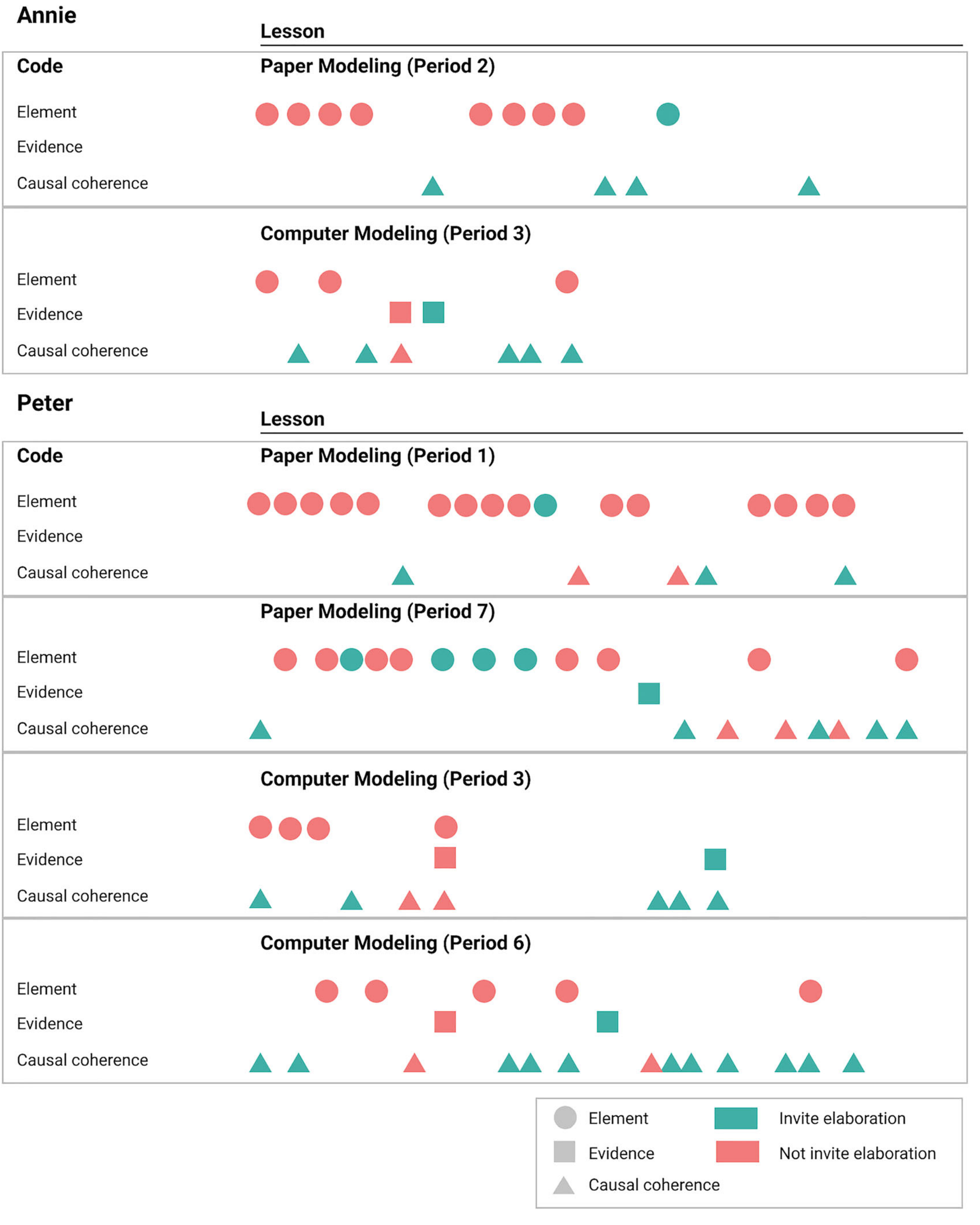


FIGURE 2 Instructional talk moves from Annie and Peter, paper versus computer modeling [Color figure can be viewed at wileyonlinelibrary.com]

In this example, while probing for students' ideas, Peter focused on reinstating the components and made sure students understand the role of microorganisms in decomposition. Peter then modeled to the whole class how to develop their paper model, focusing on the relationships in the diagram and prompting students to highlight those interactions with arrows. Peter explained the relations in linear terms, with no invitation for student elaboration, "If you have more sunlight, what will it do to temperature? Do we have more or less soil moisture?"

The teacher then modeled how to show connections on paper with direct instruction, for example, "We will draw bigger arrows for sunlight here." Peter immediately distributed the paper diagrams students had worked on from the brainstorming activity and walked around as student groups continued working on their paper models. As he circled around the room, the teacher used verbal instruction to draw the students' attention to specific steps they should take. The teacher explicitly told students to "go down the line [to include all systems components in their model]: light, temperature, soil moisture, animals, microorganisms, decomposition, nutrients, and growth."

5.2.2 | Focus on causal interactivity and elaboration in computer modeling

Peter spent significantly more time in the computer modeling lesson to explain the different features of the interface and facilitate whole-class discussion around the connections between the components (Figure 2). Notably, the teacher leveraged the tool's interactivity to ask students to play with the Simulate function and highlighted the connections among components.

Similar to the paper modeling session, the lesson started with the teacher prompting students to think about the decomposition process in the video. The teacher then modeled to students how to create a computer diagram, with similar focus on connections and relationships. Because this was the first time the students had seen the computer modeling interface, Peter spent significantly more time explaining the tasks than he did in the paper model. He demonstrated to students how to look up images to represent the different components and to connect the two elements and show relationships. While he was working with the computer program that was projected to the whole class, he continued to pose questions to all students about the nature of the relationships to create their hypotheses. For example:

> Teacher: On my poster I think the sun will affect the straw-like mulch more. **Why**? You can add in the explanation here.
> Student 1: Straw-like will let sun go through more.
> Teacher: Okay, what about sun and bark mulch here? If sun increases, let's say mulch goes down by a lot. I don't know. **Why** is that?

Peter further guided students' attention to the Simulation function, which was not included in the original lesson plan but briefly introduced to teacher during the professional development:

> Teacher: Now to test this we go to Simulate. Do we want more or less mulch? **Why**? So we [Peter slid the scale to less mulch]. Decrease in mulch leads to increase in sunlight. What happens to soil moisture? Temperature? Increase in straw-like mulch will get it [soil moisture] to do what, and **why**?
> Student: Straw-like will let sun pass through, temperature will increase and soil moisture will decrease because water evaporates.

Here, the teacher was framing questions as "why," instead of "what," which was predominant in the paper modeling session. As the teacher circled around the room, he asked students to think about which other relations they may also include. For example:

> Teacher: I have a challenge for you. Is it just nutrients that affect growth?
> Student: I have more. [Student pointed to the link between water and plant growth] Like here plants will grow more because woody mulch retains more water, and water helps plants directly too. [Students continued to draw multiple links connecting items]

Notice how Peter was eliciting students' ideas and prior knowledge about dynamic links in the systems, which was not observed in the paper modeling condition.

5.2.3 | Similar distinctions between modeling conditions in Annie's classes

Similar distinctions between paper and computer modeling—focus on procedures versus causal explanation—can be found in Annie's classrooms (Figure 2). Consider the instruction that she gave to the class prior to the modeling task in each condition. During the paper modeling lesson, she mostly focused on the procedures:

> Teacher: We are listing 6 things on your poster. **What** are microorganisms? Do we need them? Do we say "Eww fugus" or "Yes fugus"? [...] We want to see how wood and straw are changing overtime. How do we show that? Like if I want to show the difference in soil moisture? I can draw droplets. This side may have 20 while the other may have 3.

In the computer modeling lesson, Annie still focused on instructing students how to show the relationship between mulch, soil moisture, and other factors in the model. However, she integrated into instruction questions about causal mechanisms, such as "**How** is moisture affecting decomposition? **Why** do you think so? Is it an increase, decease, or the same?" These types of questions were not observed in the paper modeling lesson.

5.2.4 | Classroom prompts facilitated student interactions

We observed that the type of questions the teacher asked may affect students' responses as they reasoned about the connections within systems. Here, we highlight two episodes—one from each condition in Peter's classes. All of the paper modeling groups we observed went down the component list and drew connections sequentially. Consider the following talk from group A [emphasis added for systems components (bolded) and nonlinear relationships (italicized)]:

> Student 1: The **wood** would have more **sun**.
> Student 2: The straw has more, remember? [...]

Student 3: We are drawing only one **microorganism** here because this one has more sunlight so it'll have less moisture and fewer microorganisms. **Plant growth** will stay the same because plants take **nutrients** and this also take nutrients.

Here, students showed understanding of the relation among sunlight, microorganism, and plant growth (which was reflected in the teacher's component list). The relations they described were mostly linear and aligned with the list of components that their teacher provided.

Meanwhile, a unique dissonance appeared in the computer modeling section when students started thinking about the directions, magnitude, and interactivity among those components. Discussion emerged among student groups about which component the cause-effect arrows went toward. For example, several students observed that if they specified a link from sunlight to mulch (e.g., more sunlight means more mulch), there was no way to close the "loop" to show how different mulch conditions may affect temperature and moisture retention. There were also more conversations about "what else," which were absent from the paper modeling. Finally, students began to highlight dynamic scenarios about component interactivity by experimenting with the Simulation. A student observed:

Student 1: The **woody mulch** makes the **sunlight** go down because the woody mulch blocks the sunlight. If I **put the woody mulch up**, sunlight goes down because it blocks the sun. Then the sun increases the **temperature**, so if there's more sunlight temperature goes up, which decreases **soil moisture** a little.

Another student in the group jumped in:

Student 2: ***If we don't have a lot of*** temperature [sliding the simulation scale], moisture goes back up. And **microorganisms** like moisture so they will come if more **moisture** is present. And **microorganisms** decompose like scraps and stuff, that turns into **nutrients** and nutrients affect **plant growth. Soil moisture and sunlight *also*** affect growth.

In this discussion, while the first student reasoned about plant growth in linear terms, the second highlighted how multiple components may affect growth beyond mulch choice. They justified both scenarios (high versus low temperature) with the simulation. This example indicated emergent practice of students exploring dynamics within systems, articulating how a slight change in one component may create a domino-like effect for the other components.

6 | DISCUSSION

6.1 | Computer modeling influenced development of systems thinking

We explored the differences in systems thinking development in two conditions, computer and paper modeling. At posttest, the two groups did not differ significantly in the number of system elements that they cited, but the computer modeling group demonstrated a substantially enhanced use of evidence and causal links. The results align with prior research providing evidence that computer modeling helps students gain deeper understanding of the causal mechanisms behind scientific phenomena (Wilensky & Reisman, 2006).

The lack of evidence about differences in how students cited system elements could be because instruction and the brainstorming activity in both conditions already prompted students to include these elements in their models. The finding echoes prior research on the difference in systems thinking between experts and novice learners (Hmelo-Silver & Pfeffer, 2004). Experts may not significantly differ from novices (i.e., middle school students) in naming elements of aquatic systems, but are more likely to integrate behavioral and functional perspectives of the systems (Hmelo-Silver & Pfeffer, 2004).

6.2 | Computer modeling as mediator of teaching and learning

Analyses of audio-recorded lessons and classroom observations revealed that each condition (paper and computer) afforded different classroom interactions. We illustrated a shift in the teachers' talk, from an emphasis on systems elements and procedural talk in the paper modeling condition to causal links in the computer modeling lesson. The teachers' talk specific to causal links showed noticeable differences between the two conditions. Half of the teacher discourse in the control condition did not initiate student explanation, whereas it was predominately inviting for students' elaboration in the treatment condition. While we provided examples from both teachers' classrooms, we analyzed in depth Peter's second iteration in each condition (computer modeling; period 6 preceded paper modeling; period 7) to limit concerns about possible influence of rehearsal on teacher talk. We observed that Peter's talk moves shifted back and forth between paper and computer modeling (Figure 2), such that the focus was on elements in paper modeling and causal coherence in computer modeling.

The finding echoes prior research on the link between positive student learning outcomes and teachers' ability to support student-driven inquiries with computer tools (Puntambekar et al., 2007). The shift toward questions about the systems' causal links ("how") and evidence ("why") may have encouraged students to engage in deeper understanding of scientific phenomena (Otero & Graesser, 2001). Prior work has examined how teachers in different classrooms enacted computer tools differently in instruction, distinguishing teachers' procedural talk from instruction that invites for student explanation (Hmelo-Silver et al., 2015; Puntambekar et al., 2007). Our results illustrated that even within the same teachers, different tools may facilitate variation in teacher talk. Recognition of computer modeling affordances may have led to shifts in teacher discourse (from components to interactivity), and consequently students' responses to the teacher's prompts.

In particular, the two features that teachers attended to in the computer modeling lesson aligned with the additional affordances of the modeling interface, compared to paper modeling. We observed that the teachers and students in the computer modeling lessons focused more on relation specification (e.g., draw arrows to specify links among components, specify magnitude of relation). Observation notes revealed that the teachers and students more frequently discussed the "why" behind their hypothesized relations in the computer modeling lesson.

Interestingly, teacher Peter commented to the researcher about the utility of the Simulation feature that allowed students to iteratively test hypotheses about how systems function, and integrated this feature into his lesson even though the activity was not included in the original lesson plan. The focus on simulating systems dynamics might have encouraged students to develop alternative hypotheses and explore more complex causal mechanisms, which was observed in student participation patterns in Peter's computer modeling lesson.

6.3 | Implications

6.3.1 | Affordances of computer modeling for teaching systems thinking

We highlight the design features of the computer modeling environment, compared to paper modeling, that may foster student understanding of systems dynamics. First, in this interface, causal links are made salient through prompts for explanations and visual cues such as colors (increase/decrease relations) and sizes (magnitude). Second, the reversible drag-and-drop elements in the modeling tool enable students to experiment with nonlinear relationships (e.g., domino-like or cyclical). Third, students who experiment with simulation iteratively test alternative hypotheses, which may help to deepen their understanding of systems dynamics. Explicitly stating, visualizing, and testing interconnected systems may support systematic explorations of the hypothesized models, compared to drawing a paper model linearly.

Our conjectures about students' experimentation with simulation were mostly based on our observations with teacher Peter, who initiated the use of Simulation on his own. It is possible that students in Annie's class used Simulation on their own, but we were not able to document that. We plan to incorporate simulation in the next iteration of the curriculum for all classrooms, to further explore the affordances of this feature for developing students' systems thinking.

6.3.2 | Environments for knowledge building

Teachers' use of technology and discourse facilitate construction of learning environments. We observed two distinct focuses in the teachers' approaches between the paper modeling and computer modeling lessons: the former focused on "homogeneous understanding" where students rehearse content, and the latter focused on "heterogeneous understanding" where students develop alternative hypotheses (Brown et al., 1993). These expectations for knowledge construction may have been associated with distinct patterns of student participation—a focus on information retrieval in paper modeling and exploration of mechanisms in computer modeling. Fostering heterogeneous understanding that involved student agency may support development of conceptual understanding and scientific practices in inquiry-driven classroom interactions, particularly in the context of teaching about systems (Hmelo-Silver et al., 2015; Krajcik et al., 1998).

6.3.3 | Responsivity in computer tools

Teachers play a critical role in enacting computer modeling activities: teachers introduce students to the interface, relate students' existing knowledge to scientific models, and facilitate classroom discourse for students to engage in modeling as a scientific practice (Doerr, 2007). Wilkerson-Jerde, Wagh, and Wilensky (2015) propose to reconceptualize modeling tools as flexible and interconnected with instruction. The authors put forward the notion of *responsivity* in designing tools. Rather than being a static fixture of the curricula, modeling tools could respond to student needs as the lessons unfold—facilitating student questions about model assumptions by displaying underlying code, or expanding student interests by adding instructional modules.

Prior work has mostly examined the relation between interface and student learning. Our study expands the idea of responsive tools in relation to teachers' needs: the teachers naturally adopted certain aspects of the tool without the researcher's prompting. We note that opportunities to experiment with data and causal links, as opposed to imposing formal structures around tool use, may help ease assess to and use of tools and scientific concepts.

6.4 | Limitations and future research

Several limitations should be taken into consideration when interpreting findings. First, findings are specific to our study setting—classrooms with predominately Latinx, English learners. We note that this student sample was intentionally chosen by the state park educators, to create opportunities for Latinx low-income youth who may have limited access to scientific practices. Additionally, on average, our final sample scored better on the pretests than the students who were excluded from analyses for not completing the posttests. Thus, results should be interpreted in light of our analysis sample. However, our findings echo other research that employs samples of varied age, locations, and socioeconomic status (see Yoon et al., 2018).

Second, the study mainly focused on the content of student assessments. Future research can explore changes in student thinking over time, as students become more familiar with the content and interface and how learning progress vary across gender, motivational beliefs, and baseline content knowledge. Another topic for ongoing investigation is how students with varied baseline knowledge interact with computer modeling tools, and how tools can be integrated into instruction to respond to levels of interactions.

Third, randomization occurred at the classroom level while the unit of analysis was student. This limits the internal validity of the findings, as there may exist differences between classrooms unaccounted for in our analyses. However, our checks for between-class differences at pretests and linear mixed models for classrooms as the cluster variable suggested that the differences between the treatment and control groups were not likely due to classroom effects. Future work can replicate the study design with student-level randomization to improve the generalizability of findings.

Fourth, researchers have argued that learning evolves from complex learning contexts, being embedded in social interactions and environmental artifacts (Engeström, 1999). Another strand of research may incorporate other data sources, such as clickstreams on the modeling interface, to account for students' engagement and explore the mechanisms through which computer modeling may facilitate different modeling approaches and understanding.

7 | CONCLUSIONS

This study bridges educational research and practice by exploring the impact of computer modeling in classroom contexts. There is a need to contextualize student learning in the environments (tools) and agents (students, teachers) of learning (Cobb, McClain, de Silva

Lamberg, & Dean, 2003). However, there is limited research on teachers' enactment of computer modeling tools in classroom interactions and how it may influence student learning outcomes (Yoon et al., 2018). Our study attempts to explore aspects of classroom interactions in relation to student learning and modeling tools.

The quasi-experimental design allowed us to conclude about potential differences in learning gains between the two conditions. Findings provide evidence of using computer modeling to enrich systems thinking, particularly in evidence and causal coherence. This result is heartening, given that the students did not have prior exposure to the computer modeling tool, but were still able to create their models within the 45-min session.

Results show that enactment of computer tools may afford different types of classroom interactions. Whereas the teachers we observed mostly employed open-ended references to components in the paper condition, their discourse invited for students' elaboration of causal links more often in the computer condition. These results illustrate the promise of tools to be designed for responsivity and natural adoption in instruction to facilitate student inquiries.

ACKNOWLEDGMENTS

We want to thank Sara Ludovise, Kaitlin Magliano, Erick Valdez, and other STEM Education staff from the Crystal Cove Conservancy, Dr. Jennifer Long, Kendra Walters, and the teachers and student participants for collaborating with us to develop and pilot the curriculum. We also want to acknowledge Dr. Sandra Simpkins, Dr. Hosun Kang, Dr. June Ahn, and the editors and reviewers for providing feedback on earlier drafts of this work.

ORCID

Ha Nguyen https://orcid.org/0000-0001-7138-1427

REFERENCES

Ahern, T. C., Peck, K., & Laycock, M. (1992). The effects of teacher discourse in computer-mediated discussion. *Journal of Educational Computing Research*, *8*(3), 291–309. https://doi.org/10.2190/hfpw-jyr3-ymbe-0j7d

Alonzo, A. C., & Gotwals, A. W. (2012). *Learning progressions in science: Current challenges and future directions*. Rotterdam: Sense Publishers.

Assaraf, O. B., & Orion, N. (2005). Development of system thinking skills in the context of earth system education. *Journal of Research in Science Teaching*, *42*(5), 518–560. https://doi.org/10.1002/tea.20061

Assaraf, O. B.-Z., & Orion, N. (2009). System thinking skills at the elementary school level. *Journal of Research in Science Teaching*, *47*(5), 540–563. https://doi.org/10.1002/tea.20351

Barman, C. R. (1994). An analysis of high school students' concepts & textbook presentations of food chains & food webs. *The American Biology Teacher*, *56*(3), 160–163. https://doi.org/10.2307/4449780

Basu, S., Dickes, A., Kinnebrew, J. S., Sengupta, P., & Biswas, G. (2013). CTSiM: A computational thinking environment for learning science through simulation and Modeling. In: *Proceedings of the 5th International Conference on Computer Supported Education*, *1*, 369–378. https://doi.org/10.5220/0004390103690378

Braaten, M., & Windschitl, M. (2011). Working toward a stronger conceptualization of scientific explanation for science education. *Science Education*, *95*(4), 639–669. https://doi.org/10.1002/sce.20449

Brown, A. L., Ash, D., Rutherford, M., Nakagawa, K., Gordon, A., & Campione, J. C. (1993). Distributed expertise in the classroom. In G. Salomon (Ed.), *Distributed cognitions: Psychological and educational considerations* (pp. 188–228). Cambridge, UK: Cambridge University Press.

Chandler, M. J., & Boutilier, R. G. (1992). The development of dynamic system reasoning. *Human Development*, *35*(3), 121–137. https://doi.org/10.1159/000277143

Cobb, P., McClain, K., de Silva Lamberg, T., & Dean, C. (2003). Situating teachers' instructional practices in the institutional setting of the school and district. *Educational Researcher*, *32*(6), 13–24. https://doi.org/10.3102/0013189X032006013

Damelin, D., Krajcik, J. S., Mcintyre, C., & Bielik, T. (2017). Students making systems models an accessible approach. *Science Scope*, *40*(5), 78–83. https://search.proquest.com/docview/1853261028?pq-origsite=gscholar

Dickes, A. C., Sengupta, P., Farris, A. V., & Basu, S. (2016). Development of mechanistic reasoning and multi-level explanations of ecology in third grade using agent-based models. *Science Education*, *100*(4), 734–776. https://doi.org/10.1002/sce.21217

Doerr, H. M. (2007). What knowledge do teachers need for teaching mathematics through applications and modeling? In W. Blum, P. L. Galbraith, H.-W. Henn, & M. Niss (Eds.), *Modeling and applications in mathematics education* (pp. 69–78). New York: Springer.

Eberbach, C., Hmelo-Silver, C., Jordan, R., Sinha, S., & Goel, A. (2012). Multiple trajectories for understanding ecosystems. In J. van Aalst, K. Thompson, M. J. Jacobson, & P. Reimann (Eds.), The future of learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012) – Vol. 1(pp. 411–418). Sydney, NSW, Australia: International Society of the Learning Sciences.

Engeström, Y. (1999). Activity theory and individual and social transformation. In Y. Engeström, M. Miettinen, & R. Punamaki (Eds.), *Perspectives on Activity Theory*, (pp. 19–38). Cambridge, UK: Cambridge University Press.

Hayes, M. L., Plumley, C. L., Smith, P. S., & Esch, R. K. (2017). A review of the research literature on teaching about interdependent relationships in ecosystems to elementary students. *Horizon Research, Inc*. 1–37. https://files.eric.ed.gov/fulltext/ED587207.pdf.

Hill, C. J., Bloom, H. S., Black, A. R., & Lipsey, M. W. (2008). Empirical benchmarks for interpreting effect sizes in research. *Child Development Perspectives*, *2*(3), 172–177. https://doi.org/10.1111/j.1750-8606.2008.00061.x

Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: A quasi-experimental study. *Instructional Science*, *45*(1), 53–72. https://doi.org/10.1007/s11251-016-9392-y

Hmelo-Silver, C. E., Liu, L., Gray, S., & Jordan, R. (2015). Using representational tools to learn about complex systems: A tale of two classrooms. *Journal of Research in Science Teaching*, *52*(1), 6–35. https://doi.org/10.1002/tea.21187

Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *Journal of the Learning Sciences*, *16*(3), 307–331. https://doi.org/10.1080/10508400701413401

Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, *28*(1), 127–138. https://doi.org/10.1207/s15516709cog2801_7

Hwang, G. J., Wu, C. H., & Kuo, F. R. (2013). Effects of touch technology-based concept mapping on students' learning attitudes and perceptions. *Journal of Educational Technology & Society*, *16*(3), 274–285 Retrieved from https://www.jstor.org/stable/jeductechsoci.16.3.274

Jacobson, M. J., & Wilensky, U. (2006). Complex systems in education: Scientific and educational importance and implications for the learning sciences. *Journal of the Learning Sciences*, *15*(1), 11–34. https://doi.org/10.1207/s15327809jls1501_4

Jordan, R., Crall, A., Hmelo-Silver, C. E., Gray, S. A., Newman, G., & Sorensen, A. (2018). Developing model-building as a scientific practice in collaborative citizen science. *Natural Sciences Education*, *47*(1), 1–7. https://doi.org/10.4195/nse2018.07.0013

Kang, H., Thompson, J., & Windschitl, M. (2014). Creating opportunities for students to show what they know: The role of scaffolding in assessment tasks. *Science Education*, *98*(4), 674–704. https://doi.org/10.1002/sce.21123

Kim, M. C., Hannafin, M. J., & Bryan, L. A. (2007). Technology-enhanced inquiry tools in science education: An emerging pedagogical framework for classroom practice. *Science Education*, *91*(6), 1010–1030. https://doi.org/10.1002/sce.20219

Komis, V., Ergazaki, M., & Zogza, V. (2007). Comparing computer-supported dynamic modeling and 'paper & pencil' concept mapping technique in students' collaborative activity. *Computers & Education*, *49*(4), 991–1017. https://doi.org/10.1016/j.compedu.2005.12.007

Krajcik, J., Blumenfeld, P. C., Marx, R. W., Bass, K. M., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences*, *7* (3–4), 313–350. https://doi.org/10.1080/10508406.1998.9672057

Leach, J., Driver, R., Scott, P., & Wood-Robinson, C. (1995). Children's ideas about ecology 1: Theoretical background, design and methodology. *International Journal of Science Education*, *17*(6), 721–732. https://doi.org/10.1080/0950069950170604

Levin, T., & Wadmany, R. (2006). Teachers' beliefs and practices in technology-based classrooms: A developmental view. *Journal of Research on Technology in Education*, *39*(2), 157–181. https://doi.org/10.1080/15391523.2006.10782478

Looi, C.-K., Sun, D., Seow, P., & Chia, G. (2014). Enacting a technology-based science curriculum across a grade level: The journey of teachers' appropriation. *Computers & Education*, *71*, 222–236. https://doi.org/10.1016/j.compedu.2013.10.006

Louca, L. T., & Zacharia, Z. C. (2012). Modeling-based learning in science education: Cognitive, metacognitive, social, material and epistemological contributions. *Educational Review*, *64*(4), 471–492. https://doi.org/10.1080/00131911.2011.628748

Maas, C. J. M., & Hox, J. J. (2005). Sufficient sample sizes for multilevel modeling. *Methodology*, *1*(3), 86–92. https://doi.org/10.1027/1614-2241.1.3.86

Maor, D., & Taylor, P. C. (1995). Teacher epistemology and scientific inquiry in computerized classroom environments. *Journal of Research in Science Teaching*, *32*(8), 839–854. https://doi.org/10.1002/tea.3660320807

McKinley, D. C., Miller-Rushing, A. J., Ballard, H. L., Bonney, R., Brown, H., Cook-Patton, S. C., … Phillips, T. B. (2017). Citizen science can improve conservation science, natural resource management, and environmental protection. *Biological Conservation*, *208*, 15–28. https://doi.org/10.1016/j.biocon.2016.05.015

Mercer, N. (2007). Sociocultural discourse analysis: Analysing classroom talk as a social mode of thinking. *Journal of Applied Linguistics and Professional Practice*, *1*(2), 137–168. https://doi.org/10.1558/japl.v1.i2.137

Murphy, E., & Rodriguez-Manzanares, M. A. (2008). Using activity theory and its principle of contradictions to guide research in educational technology. *Australasian Journal of Educational Technology*, *24*(4), 442–457. https://doi.org/10.14742/ajet.1203

NGSS Lead States. (2013). *Next generation science standards: For states, by states*. Washington, DC: National Academies Press.

Novak, A. M., & Krajcik, J. S. (2019). A case study of project-based learning of middle school students exploring water quality. In M. Moallem, W. Hung, & N. Dabbagh (Eds.), *The Wiley handbook of problem-based learning* (pp. 551–572). Hoboken, NJ: Wiley. https://doi.org/10.1002/9781119173243.ch24

Otero, J., & Graesser, A. C. (2001). PREG: Elements of a model of question asking. *Cognition and Instruction*, *19* (2), 143–175. https://doi.org/10.1207/S1532690XCI1902_01

Plate, R. (2010). Assessing individuals' understanding of nonlinear causal structures in complex systems. *System Dynamics Review*, *26*(1), 19–33. https://doi.org/10.1002/sdr.432

Puntambekar, S., Stylianou, A., & Goldstein, J. (2007). Comparing classroom enactments of an inquiry curriculum: Lessons learned from two teachers. *The Journal of the Learning Sciences*, *16*(1), 81–130. https://doi.org/10.1080/10508400709336943

Reiner, M., & Eilam, B. (2001). Conceptual classroom environment - a system view of learning. *International Journal of Science Education*, *23*(6), 551–568. https://doi.org/10.1080/095006901300172458

Rosenberg, J., & Lawson, M. (2019). An investigation of students' use of a computational science simulation in an online high school physics class. *Education Sciences*, *9*(1), 49. https://doi.org/10.3390/educsci9010049

Schwarz, C. V., Reiser, B. J., Davis, E. A., Kenyon, L., Achér, A., Fortus, D., … Krajcik, J. (2009). Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners. *Journal of Research in Science Teaching*, *46*(6), 632–654. https://doi.org/10.1002/tea.20311

Sengupta, P., Kinnebrew, J. S., Basu, S., Biswas, G., & Clark, D. (2013). Integrating computational thinking with K-12 science education using agent-based computation: A theoretical framework. *Education and Information Technologies*, *18*(2), 351–380. https://doi.org/10.1007/s10639-012-9240-x

Sun, D., Looi, C.-K., & Xie, W. (2014). Collaborative inquiry with a web-based science learning environment: When teachers enact it differently. *Journal of Educational Technology & Society*, *17*(4), 390–403 Retrieved from https://www.jstor.org/stable/jeductechsoci.17.4.390

Sweeney, L. B., & Sterman, J. D. (2007). Thinking about systems: Student and teacher conceptions of natural and social systems. *System Dynamics Review*, *23*(2–3), 285–311. https://doi.org/10.1002/sdr.366

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics* (Vol. 5, pp. 481–498). Boston, MA: Pearson.

van Borkulo, S. P., van Joolingen, W. R., Savelsbergh, E. R., & de Jong, T. (2012). What can be learned from computer modeling? Comparing expository and modeling approaches to teaching dynamic systems behavior. *Journal of Science Education and Technology*, *21*(2), 267–275. https://doi.org/10.1007/s10956-011-9314-3

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147. https://doi.org/10.1007/s10956-015-9581-5

Wilensky, U., & Reisman, K. (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), 171–209. https://doi.org/10.1207/s1532690xci2402_1

Wilensky, U., & Resnick, M. (1999). Thinking in levels: A dynamic systems approach to making sense of the world. *Journal of Science Education and Technology*, *8*(1), 3–19. https://doi.org/10.1023/A:1009421303064

Wilkerson-Jerde, M., Wagh, A., & Wilensky, U. (2015). Balancing curricular and pedagogical needs in computational construction kits: Lessons from the DeltaTick project. *Science Education*, *99*(3), 465–499. https://doi.org/10.1002/sce.21157

Xiang, L., & Passmore, C. (2015). A framework for model-based inquiry through agent-based programming. *Journal of Science Education and Technology*, *24*(2–3), 311–329. https://doi.org/10.1007/s10956-014-9534-4

Yoon, S. A., Goh, S.-E., & Park, M. (2018). Teaching and learning about complex systems in K–12 science education: A review of empirical studies 1995–2015. *Review of Educational Research*, *88*(2), 285–325. https://doi.org/10.3102/0034654317746090

How to cite this article: Nguyen H, Santagata R. Impact of computer modeling on learning and teaching systems thinking. *J Res Sci Teach*. 2020;1–28. https://doi.org/10.1002/tea.21674