Bowers and Eidin *Disciplinary and Interdisciplinary Science Education Research* (2024) 6:24
https://doi.org/10.1186/s43031-024-00115-7

Disciplinary and Interdisciplinary Science Education Research

RESEARCH

Open Access

Analyzing students' systems thinking in-situ through screencasts in the context of computational modeling: a case study

Jonathan Bowers[1*] and Emil Eidin[2]

Abstract

In our interconnected world, Systems Thinking (ST) is increasingly being recognized as a key learning goal for science education to help students make sense of complex phenomena. To support students in mastering ST, educators are advocating for using computational modeling programs. However, studies suggest that students often have challenges with using ST in the context of computational modeling. While previous studies have suggested that students have challenges modeling change over time through collector and flow structures and representing iterative processes through feedback loops, most of these studies investigated student ST through pre and post tests or through interviews. As such there is a gap in the literature regarding how student ST approaches develop and change throughout a computational modeling unit. In this case study, we aimed to determine which aspects of ST students found challenging during a computational modeling unit, how their approaches to ST changed over time, and how the learning environment was supporting students with ST. Building on prior frameworks, we developed a seven-category analysis tool that enabled us to use a mixture of student discourse, writing, and screen actions to categorize seven ST behaviors in real time. Through using this semi-quantitative tool and subsequent narrative analysis, we found evidence for all seven behavior categories, but not all categories were equally represented. Meanwhile our results suggest that opportunities for students to engage in discourse with both their peers and their teacher supported them with ST. Overall, this study demonstrates how student discourse and student writing can be important evidence of ST and serve as a potential factor to evaluate ST application as part of students' learning progression. The case study also provides evidence for the positive impact that the implementation of a social constructivist approach has in the context of constructing computational system models.

Keywords Systems thinking, Computational modeling, Modeling, High school, Science education, Case study

*Correspondence:
Jonathan Bowers
Bowersj8@msu.edu
[1]Michigan State University, East Lansing, MI, USA
[2]University of Wyoming, Laramie, WY 82071, USA

Introduction

The natural world is full of complex and interconnected systems, many of which resist explanation through simple linear "cause and effect" relationships. From the interplay of energy and matter in evaporative cooling to the perilous specter of runaway climate change, countless science phenomena involve multiple interacting elements that together can be defined as distinct systems (Gilissen et al., 2020; Meadows, 2008; Zangori et al., 2017). These complex systems-based phenomena often have behaviors that are difficult to predict from binary interactions between individual pairs of elements within the system, thus limiting the effectiveness of using linear "cause and effect" reasoning (Arnold & Wade, 2015; Assaraf & Orion, 2010; Sweeney & Sterman, 2000). However, K-12 Science Education overly emphasizes linear causal reasoning, which limits the opportunities students have to try and make sense of a wide range of phenomena that have an impact on their everyday lives (Fisher & Systems Thinking Association, 2023; Lee et al., 2019). Hence, ST scholars advocate for science education to move from linear causal explanations towards considering behavior that emerges from cumulative interactions within a system (Easterbrook, 2014; Forrester, 1994; Gilissen et al., 2020; Hmelo-Silver et al., 2017). Moreover, there is a growing consensus to equip students with the skills of a systems thinking mindset, preparing them as citizens who can apply those thinking skills to understand and solve systems level problems (Michalopoulou et al., 2019; Reynolds et al., 2018).

Systems thinking (ST) is a framework for approaching and understanding phenomena as a series of interconnected elements, in which the summative interactions between those results in a behavior that could not be otherwise explained by the mechanism of just one interaction within the system (Arnold & Wade, 2015; Sweeney & Sterman, 2007; Meadows, 2008). Systems thinking is increasingly recognized as a key learning goal for science education across many countries including Germany, the United States, and Australia (ACARA, 2017; KMK, 2020; NRC, 2012). In the United States, ST is embedded in the Next Generations Science Standards (NGSS) as the crosscutting concept of systems and systems modeling (NGSS Lead States, 2013). By integrating ST within a modeling framework, the NGSS, along with other scholars, advocate for the use of computational modeling software as a promising avenue for supporting students' ST (Haas et al., 2020; Metcalf et al., 2000; Shin et al., 2022; Wilensky & Reisman, 2006). Through building and revising computational models, students have opportunities to consider the impact of individual elements and relationships on system behavior, examine how system behavior changes as the relative amount of certain variables fluctuate over time, and explore how feedback loops affect system behavior (Bielik et al., 2020; Bowers et al., 2023; Eidin et al., 2023a; Fretz et al., 2002).

Although systems and system modeling have been discussed in the educational literature for decades, students still have challenges understanding phenomena from a ST perspective (Cronin et al., 2009; Hmelo-Silver et al., 2017; Plate, 2010). Despite being a part of official policy documents, ST remains sidelined in many K-12 science classrooms. Even in classrooms that have shifted towards integrating systems thinking, students tend to apply a reductionist approach, simplifying circular and feedback mechanisms into simple linear relationships (Assaraf & Orion, 2010; Bowers et al., 2023). Several scholars have demonstrated that students often struggle with making sense of dynamic systems, where the behavior of the system changes over time, and representations of dynamic relationships within these systems (Cronin et al., 2009; Hopper & Stave, 2008; Pallant & Lee, 2017; Zuckerman & Resnick, 2005).

It is evident from the literature that there is a need to gain a deeper understanding and characterization of students' systems thinking approaches while engaging in computational systems modeling. If we fully grasp the challenges students face with ST, we can design better learning environments. Additionally, there is also a dearth of literature on how different aspects of the learning environment can assist students with applying ST principles during computational modeling. Although earlier studies have explored student ST through pre-post tests, post unit interviews, and student final models, relatively few studies have looked at how students engage with ST throughout a computational modeling unit (Hmelo-Silver et al., 2007; Khajeloo & Siegel, 2022; Riess & Mischo, 2010; Taylor et al., 2020). Because student learning occurs within a specific context, being heavily influenced by peer-peer interactions and student prior knowledge (Driver, 2012), it is often beneficial to take a holistic approach towards evaluating student learning. Seeing how student ST evolves organically throughout a learning unit can help reveal how students build competency in ST and underscore the inherent challenges with certain aspects of ST in similar contexts. While pre and post tests can reveal if a particular unit is effective at supporting student ST, relying solely on pre and post tests can obfuscate which specific aspects of the unit (be it specific teacher supports, student discourse questions, or technological scaffolds) are beneficial for student learning. As such, having a more in-depth approach that looks at how students are engaging in ST during a unit is important for finding better strategies for supporting ST in science classrooms. Only examining student models at predetermined checkpoints also obscures the rationale behind key modeling decisions and hides the key learning moments. Therefore, we acknowledge that an in-depth investigation that looks at how students build computational models in-situ (including their discourse practices) can help educational researchers better understand how students build competency with ST and the specific aspects of the learning environment that assist with this process. Because a gap in the literature exists with regards to how to use student discourse during the process of modeling to analyze student ST, this study also aims to explore techniques for analyzing student discourse in this learning context. As such we set out to address the following questions.

Research questions

1. How do students apply ST as they build and revise computational systems models in this unit?
2. How do student ST behaviors change over the course of this unit?
3. What supports from the learning environment assist students with applying ST in this unit?

Literature review

Defining systems thinking

Systems thinking (ST) describes a set of cognitive processes whereby one examines a phenomenon as a collection of individual elements that form a system with complex and often unexpected behavioral outcomes (Arnold & Wade, 2015; Assaraf & Orion, 2005; Shin et al., 2022; Stave & Hopper, 2007; Whitehead et al., 2014). Emerging as an alternative to the reductionist approaches to science, science education, and other fields of knowledge, ST scholars have long advocated for approaching phenomena and problems in a holistic manner that examines how behavior emerges not only from simple causal patterns, but from the complex interactions between different elements within a system (Dominici, 2012; Fang & Casadevall, 2011; Forrester, 1994; MacInnis, 1995; Orgill et al., 2019). We next describe some ST principles and skills that are based on the research consensus, while also being useful for the context of computational systems modeling: evaluating system variables, analyzing single causal relationships and linear causal chains, choosing collector variables and discussing collector and flow structures, discussing feedback loops and circular chains, and interpreting graphical model output.

One major aspect of systems thinking is *evaluating system variables* to determine what elements need to be included to adequately represent how a system functions (Arnold & Wade, 2015; Stave & Hopper, 2007; Sweeney & Sterman, 2007). We define *element* as a core aspect of a system that can be described independently and interacts with other elements in the system (Arnold & Wade, 2015; Meadows, 2008; Riess & Mischo, 2010; Shin et al., 2022). For example, an element in a model of a forest ecosystem would be the number of wolves in the forest as it can be independently described and impacts other elements in the ecosystem, such as the number of rabbits or number of deer. In the context of computational systems modeling, elements need to be described in a manner that can be recognized by the software program and thus transform into quantitative or semi-quantitative variables. Evaluating system variables often requires considering system boundaries and figuring out which variables are superfluous to the model (Arnold & Wade, 2017; Assaraf & Orion, 2005; Stave & Hopper, 2007). While the number of fish in a stream does have an impact on a forest ecosystem, if the focus is on predator/prey relationships between wolves, rabbits, and deer, it might not be a necessary element to include in a model and can be considered to be outside of the boundaries of the system.

Analyzing causal relationships refers to describing and unpacking the *interactions* between elements, or the various types of relationships that can exist between elements in a system (Arnold & Wade, 2015; Cabrera et al., 2008; Plate, 2010). ST literature includes a common nomenclature to describe the different types of interactions that can occur between elements (Arnold & Wade, 2015; Cronin et al., 2009; Hopper & Stave, 2008; Monat & Gannon, 2015). At a base level are direct or *single causal relationships*, where one element has a direct impact on another element (i.e., A affects B). Such direct causal relationships can have varying *magnitudes* or rates defining these relationships; some elements can cause another element to increase exponentially while others cause another element to decrease at a steady rate. When several single causal relationships are arranged in a linear pattern, they are often considered to form a *linear causal chain* where A affects B, B affects C, and so on (Plate, 2010; Stephens et al., 2023). As students develop more familiarity with systems thinking principles, they often shift from analyzing single causal relationships towards examining linear causal chains (Mambrey et al., 2022; Mehren et al., 2018) and ultimately consider how external mediating factors (elements outside of the causal chain) impact the linear causal chain (Stephens et al., 2023).

Another type of relationship found in ST literature are *collectors and flows*, also referred to as stock and flow systems (Cronin et al., 2009; Eidin et al., 2023a; Sweeney & Sterman, 2000). A collector (or stock) represents a variable that can accumulate and deplete in one form (collector) and be transferred via a "flow" to another form thereby showing change over time. For example, in a forest ecosystem, the number of living deer can be represented by a collector as the population of living deer can accumulate or deplete. In such a model it would also be possible to show how the number of living deer are transferred to the number of dead deer (represented by a second collector) as predators, disease, and other factors cause living deer to die, directly adding to the number of dead deer (Fig. 1). Collector and flow structures are often useful for representing how systems can change over time or remain in a state of dynamic equilibrium (Eidin et al., 2023a). Thinking in terms of change over time refers to the consideration of the dynamic nature of systems, delays, and reaction to change (Arnold & Wade, 2017; Gotwals & Songer, 2010; Sweeney & Sterman, 2007). An array of phenomena and problems require the application or framing in terms of change over time. Evolution, the creation of a Canyon, and climate change are just a few examples of phenomena that require the addressing of dynamic processes, changes, and time delays to explain them and make predictions (Petrosino et al., 2015; Roychoudhury et al., 2017).

Discussing feedback loops and circular chains involves the recognition of recursive or circular relationships patterns and their impacts on broader system behavior. Feedback loops describe recursive relationship patterns whereby the output of a process is in turn used as an input in a subsequent cycle, thereby creating recursion in a system (Cox et al., 2019; Hopper & Stave, 2008; Richmond, 1993). Positive/reinforcing feedback loops, such as the relationship between atmospheric carbon dioxide and global temperatures, amplify the effects of an initial input variable, leading to an exponential increase overtime (Betley et al., 2021; Danish et al., 2017; York & Orgill, 2020). Conversely, negative/stabilizing feedback loops counteract or dampen the impact of an initial variable, helping a system maintain equilibrium (Flood, 2010; Sweeney & Sterman, 2000). While students can include recursive or circular patterns in their models by accident, understanding feedback mechanisms and their broader impact on the system is evidence of more sophisticated systems thinking.

A final aspect of systems thinking that is largely unique to the context of computational modeling is *interpreting graphical model output*. Computational modeling allows students to visualize relationships that exist between the variables of the system and generate graphical model output that can be used to see if the system is behaving according to their expectations and understanding of the underlying phenomenon (Haas et al., 2020; Nguyen & Santagata, 2021; Pierson & Brady, 2020). By interpreting graphical model output students can examine how various relationship patterns, such as collector and flow/stock and flow structures and feedback loops, influence model behavior and represent a dynamic system. As such this aspect of ST can overlap and support students with the previously mentioned ST aspects. In many instances, students can also use computational modeling programs to compare their model output to real world data to see how their conceptualization of system structure reflects the actual behavior of the system (Abar et al., 2017; Bowers et al., 2023; Campbell & Oh, 2015; Grapin et al., 2022). As the type of graphical model output varies across computational modeling programs, the exact mechanics of *interpreting graphical model output* are defined by the software one uses.

Using computational modeling to support students with ST

There are two main approaches to computational modeling that are commonly used to support students with systems thinking: agent-based modeling and icon-based modeling. Agent based modeling focuses on having students program the behavior of specific agents that can interact with other agents in a system (Goldstone & Janssen, 2005; Goldstone & Wilensky, 2008; Wilensky & Rand, 2015; Yoon et al., 2016). For example, in a simplified model of an ecosystem, a student might program an herbivore "agent" to consume plants continuously, to duplicate if it has procured enough food, and for individual herbivore agents to die if they have not procured food. Through the programming of various agents in the ecosystem, such as carnivores, decomposers, and producers, students can create a complex system that can produce dynamic behavior, but key relationships between various agents can be obscured as they are not directly represented in a visual format. In icon-based modeling, students input key elements as variables on a modeling canvas (Damelin et al., 2017; Metcalf et al., 2000; Nguyen & Santagata, 2021; Zhang et al., 2006). The students then define specific relationships between these variables in a quantitative or semi-quantitative manner. These relationships between individual variables are then visually represented on the modeling canvas. Students can then

Fig. 1 Example of a collector and flow system

manipulate the relative amount of each input variable to see how their model behaves under different initial conditions (Bielik et al., 2020; Bowers et al., 2023; Fretz et al., 2002). In some icon-based modeling programs, students can create collector and flow relationships and make feedback loops to generate models with dynamic behavior. (Bielik et al., 2020; Eidin et al., 2023a, b).

One icon-based modeling program that has the potential to allow students to build dynamic systems and create opportunities for students to develop sophisticated ST is SageModeler. SageModeler is a free open-source computational modeling software that allows students to place individual variables onto a canvas and set relationships between these variables in a semi-quantitative manner via a drop-down menu (Fig. 2A) (Bielik et al., 2020; Bowers et al., 2023; Damelin et al., 2017; Nguyen & Santagata, 2021). Students can also make collector and flow relationships and feedback loops (Fig. 2B). Once students have created their models, they can use the simulate feature to explore how the model behaves under different initial starting conditions and/or how the behavior of the model changes over time (Fig. 2C). Students also can input external data into SageModeler and compare these data to model output to validate their computational models (Fig. 2D). These features enable students to construct computational models of dynamic systems and have the potential to support them in building competency with ST (Eidin et al., 2023a; Shin et al., 2022). Although SageModeler has been shown to be a useful tool to support students with systems thinking, these earlier studies primarily focused on having students work with constructing static-equilibrium models (Damelin et al., 2017; Eidin et al., 2023b; Nguyen & Santagata, 2021). In these static equilibrium models, dynamic representational features such as collector and flow structures are lacking in scope, making it more difficult for students to represent changes in the behavior of a system over time. Given SageModeler's potential to support students in ST, we are interested in investigating how students use ST as they construct models of dynamic systems with this program.

Despite the affordances of computational modeling tools in supporting users' ST, there are evident challenges of users' ability to utilize the full potential of these tools. For example, students and adult learners tend to primarily think in terms of linear causality, without considering more complex relationship dynamics (Cronin et al., 2009; Driver et al., 1985; Hmelo-Silver et al., 2017; Plate, 2010). Indeed, the focus on linear cause and effect relationships,

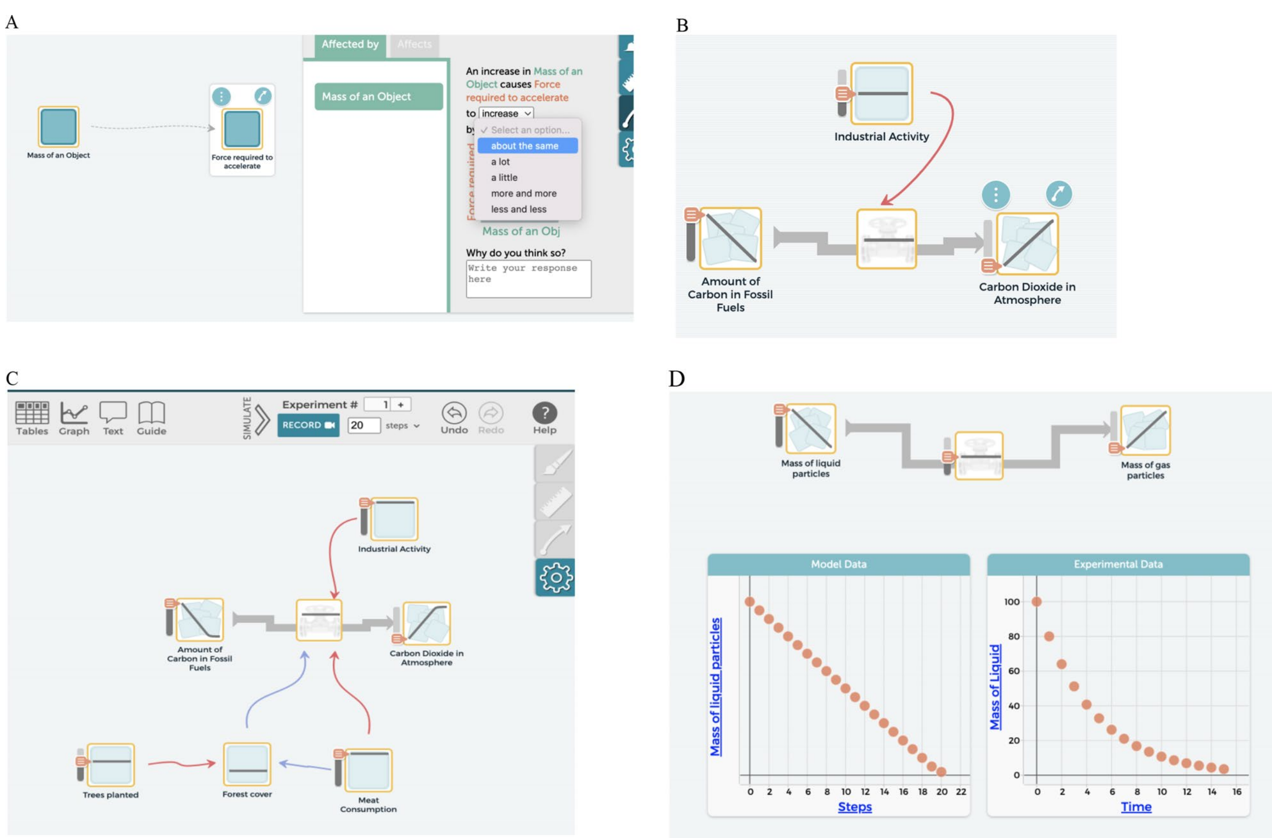

Fig. 2 SageModeler. **A** Setting a relationship in SageModeler. **B** collector and flow relationships in SageModeler. **C** Simulation features in SageModeler. **D** Comparing experimental data to model output data

that is often reinforced by standard approaches to science education (Gilissen et al., 2020; Raia, 2005), itself appears to be a hindrance for conceptualizing how systems can change over time (Eidin et al., 2023b). As such, many students have challenges with understanding dynamic systems, and its model representation such as collector/stock and flow structures and generated graphs simulating how the behavior of a system changes over time (Cronin et al., 2009; Hopper & Stave, 2008; Pallant & Lee, 2017; Zuckerman & Resnick, 2005). This is reflected in studies where students have difficulty utilizing collector and flow/stock and flow relationships to demonstrate how a key element of the system (such as the mass of water) can be transferred from one form to another (in this case from liquid to gas) over time (Cronin et al., 2009; Pallant & Lee, 2017; Zuckerman & Resnick, 2005). Another aspect of dynamic systems that students often find challenging is understanding the behavioral impact of feedback loops and circular behavioral patterns (Assaraf & Orion, 2010; Cox et al., 2019; Hmelo-Silver et al., 2017; Hopper & Stave, 2008). Given that many important phenomena, such as climate change, forest ecosystems, and the human endocrine system, involve dynamic systems that are explained through feedback loop mechanisms, it is important that science educators find ways of supporting students with moving beyond linear causal reasoning.

Methods

Study context

Classroom environment

This study took place at Faraday High School (FHS), a pseudonym for a public magnet school in the Midwestern United States, during November and December of 2022. As a public magnet school, FHS recruits students from across the tri-county "Faraday City" area primarily based on academic merit as determined by student test scores and teacher recommendations. Around 21% of FHS students identify as Non-White and around 54% of students receive free or reduced lunches. FHS runs on a block schedule, meaning that students attend each class twice a week for 80 min. In this work, I collaborated with Mr. H (a middle-aged White Male high school chemistry teacher with approximately 20 years of teaching experience) to implement a unit on evaporative cooling. Mr. H participated in a weekly professional learning community (PLC) prior to implementing the evaporative cooling unit where we went over the evaporative cooling unit in depth and discussed specific strategies for supporting students with ST.

Curriculum

Mr. H implemented a five-week high school chemistry unit on evaporative cooling with his 10th grade students. Evaporative cooling is the process by which high kinetic energy (KE) particles are the first to evaporate from a liquid, causing the liquid to cool. As these particles break their intermolecular bonds with other particles and evaporate as a gas, their kinetic energy is transferred to the potential energy (PE) of the evaporated gas. Because the high kinetic energy particles evaporate first, the average kinetic energy of the liquid (i.e. the temperature of the liquid) decreases, causing the liquid to become colder. Given that the temperature of the liquid helps determine the evaporation rate, the loss of kinetic energy to potential energy during evaporation creates a negative feedback loop, causing the evaporation rate to slow as the liquid cools. The rate of evaporation is also in part determined by the strength of the intermolecular forces (IMFs) of each liquid; liquids with weaker IMFs will evaporate more quickly and therefore feel colder as they evaporate off our skin (Fig. 3).

As a design-based research unit, we deliberately developed specific supports and pacing instructions to benefit students with engaging in ST. Many of these design choices were inspired by the existing literature and generally accepted practices for unit design. As this was the fourth iteration of this unit, several design choices were influenced by our past experiences with implementing earlier iterations of this unit. One early decision based on existing curriculum design literature was to design the evaporative cooling unit based on Project Based Learning (PBL) principles (Krajcik & Shin, 2022). These principles include: having students experience a hands-on scientific phenomenon, guiding student inquiry and allowing students to ask their own questions related to the phenomenon through the use of driving questions and a driving question board, facilitating students in using scientific practices to investigate the phenomenon to address the driving question, and having students generate a knowledge product (i.e. a computational systems model) that demonstrates their learning (Krajcik & Shin, 2022). This unit was also designed to align with official science education standards adapted from the NGSS. In this unit, students were tasked with creating a computational model of evaporative cooling that addressed the following driving question: "Why do I feel colder when I am wet than when I am dry?" The students worked in groups of 2–3 to build, test, and revise their computational models using SageModeler software.

Prior to starting this unit, students were briefly introduced to SageModeler through a short tutorial where they discussed how to create variables, set relationships between these variables, and create collector and flow relationships using this software program. Students began the unit by experiencing the evaporative cooling phenomenon by noticing how three different liquids (water, acetone, and rubbing alcohol) feel as they

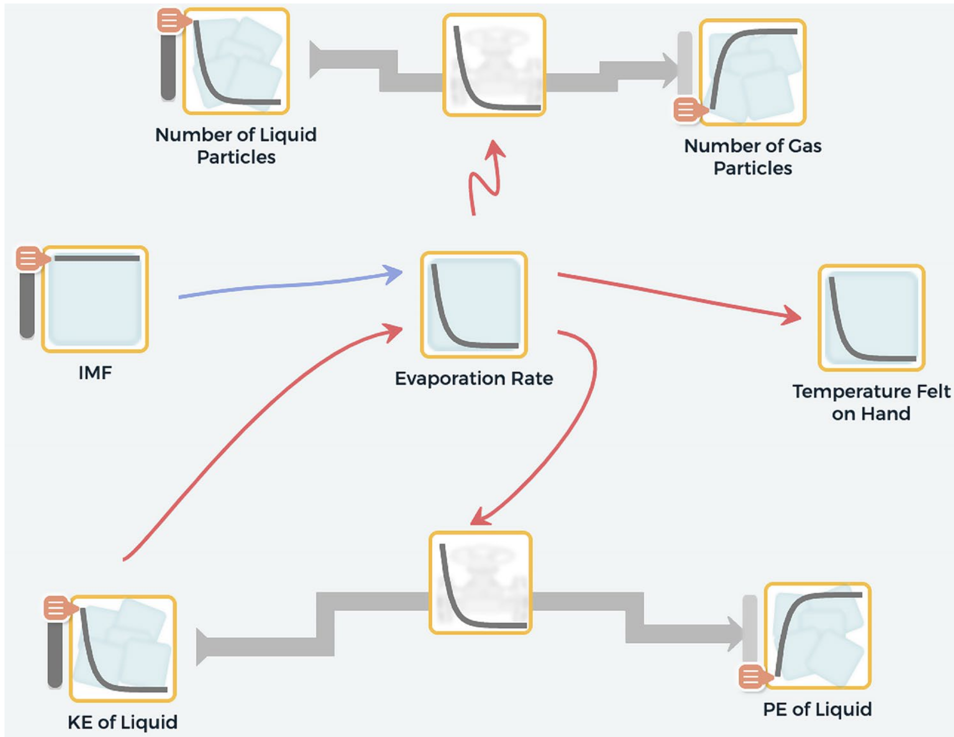

Fig. 3 Evaporative cooling model

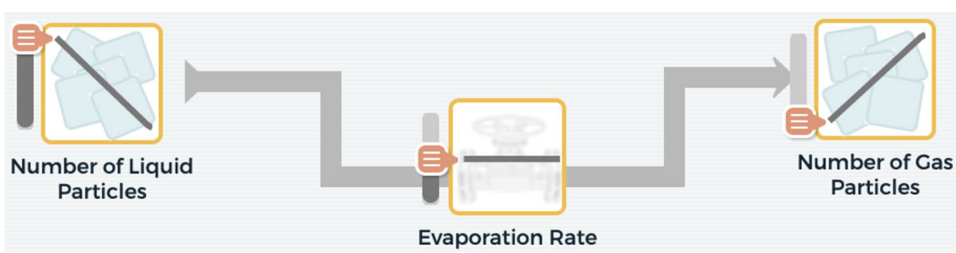

Fig. 4 Backbone of evaporative cooling model

evaporated from their hands. Students were then tasked with making an initial diagrammatic model of this phenomenon and engaging in an embodied modeling task where they acted out the role of liquid molecules evaporating into gas particles. Mr. H then worked with his students to co-construct the initial "backbone" of this model in SageModeler, i.e. the transfer relationship between the number of liquid particles and the number of gas particles (Fig. 4).

After constructing this initial backbone, students were encouraged to add additional variables to their SageModeler models. As the unit progressed, students, with the help of online learning modules, first-hand experiments, and classroom lectures/discussions, were introduced to additional science concepts (i.e. IMF, Kinetic Energy, and Potential Energy). After each learning module, students were tasked with revising their computational models based on their evolving understandings of evaporative cooling. Students were also given multiple opportunities to critique peer models, both through whole class critiques and small group discussions, and receive feedback on their models. Note that we divided the unit up into two main sections for analysis purposes: Pre Potential Energy phase (the first six lessons) and Post Potential Energy phase (the final five lessons). In the Pre Potential Energy phase, students are accumulating knowledge about different elements of evaporative cooling to add on to the unitary backbone relationship showing evaporation as the transformation from liquid to gas. Because students learn about potential energy, the final new conceptual idea/system variable, on lesson seven, we consider this a key turning point in the unit as they have all of the variables necessary to complete a holistic model and can begin model revision and validation. As such students enter the Post Potential Energy phase, where they should include two parallel collector and flow structures

(Fig. 4) to demonstrate that evaporation involves both a change in state (matter being transformed from liquid to gas) and in energy (kinetic energy to potential energy). Towards the end of the unit, students collected experimental temperature vs. time data, which helped validate their final models. Table 1 in the Appendix provides a summary of the science content goals, ST learning goals, and learning activities of each lesson of the unit. More details about design changes made across multiple iterations of this unit can be found in the Appendix.

Data collection

We collected data for this study in November and December of 2022 in partnership with Mr. H who implemented the evaporative cooling unit at FHS. For the purpose of this case study, we collected student data from just one class (29 students) taught by Mr. H. Our primary data source was student screencasts: video and audio recordings that allow researchers to capture student screen actions (how they are building and constructing their models) along with video and audio of student conversations conducted during this process. We were able to track how students built and revised their models along with the discourse students had around ST during the modeling process. These screencasts were collected from five student groups (11 students total) in Mr. H's class, whose demographics and pseudonyms are found in the table below (Table 1). These five student groups were chosen based on convenience sampling as they were the only students who volunteered to participate in the screencast data collection process. While the other 18 students participated in all other classroom activities, they did not have their laptop screens recorded for this study. Given that each class period was 80 min and that this unit lasted for 11 lessons, we collected 880 min of screencasts for each student group or 4400 min for all five groups. Because screencast data was only collected from these five student groups, only their data and class time is included in our data analysis process. We also collected audio from Mr. H to capture his pedagogical moves and dialogue that were not picked up by the student screencasts.

Table 1 Screencast student pseudonyms and demographics

Student group	Student pseudonyms	Demographics
Group 1	Conrad and Zion	South Asian Male, White Male
Group 2	Amy and Leia	White Female, White Female
Group 3	Walter, Larry, and Ivan	White Male, White Male, White Male
Group 4	Brianna and Kate	South Asian Female, White Female
Group 5	Phillip and Robyn	White Male, White Female

Instrument development and validation

To analyze student ST from the screencasts in this case study, we developed the Dynamic Systems Thinking through Modeling Analysis Tool (Dynamic ST Tool). Other efforts to measure student ST in a computational modeling context have either analyzed the structural complexity of student models structures through pre-post tests (Taylor et al., 2020), assessed student knowledge of ST principles through written assessments (Riess & Mischo, 2010), or conducted scaffolded interviews where students unpacked their reasoning behind different modeling decisions (Khajeloo & Siegel, 2022) or tested their generalized ST knowledge (Mambrey et al., 2022). In contrast to these earlier studies, which either assessed student ST in isolation from the modeling process, analyzed student models as final products, or interviewed students after their models were completed, this tool aimed to focus on student discourse and student actions during the whole modeling process and perhaps provides a more authentic view of student ST in situ.

When designing this instrument, we began by reviewing existing ST literature, looking for common themes found across different studies. In our literature review, we focused on studies that involved computational modeling environments or were influential to the conceptualization of ST as described by other computational modeling studies. Some of the major studies that drove our literature review of ST include: Arnold and Wade (2015, 2017); Assaraf and Orion (2005); Bielik et al. (2023); Cronin et al. (2009); Meadows (2008); Mehren (2018); Plate (2010); Richmond (1993); Shin et al. (2022); Stave and Hopper (2007); and Sweeney and Sterman (2000, 2007). In particular, our previous work on developing "A Framework for Computational Systems Modeling" (Shin et al., 2022) heavily influenced our conceptualization of ST in the context of computational systems modeling. From these studies, we listed out different aspects of ST that were important across the ST literature and that were relevant to the context of computational modeling. Aspects of ST that were unique to only one or two papers were excluded at this stage. We then discussed how these aspects of ST could be manifested as students built and revised computational models using SageModeler software. Such discussions aimed to find behavioral indicators that could clearly be identified from student screencast data and easily distinguished from each other. Through these discussions, we identified seven main indicators of student ST. With the initial indicators, we analyzed two hours of student screencasts to test the feasibility of our research instrument. After this initial test run, we made additional revisions, adding helpful details to better describe student ST behaviors. We then created and tested a four-part classification system (from Level 1 to Level 4 in ascending order) for each of the seven indicators to explore the sophistication of observed student discourse and behavior. Please note that while we do not include these level categories in our quantitative analyses in this paper, we did use these level categories to inform our qualitative analysis and to help differentiate between different levels of sophistication in student ST behavior across the course of this unit. The current version of this instrument can be found below (Table 2).

Given the novelty of the Dynamic ST Tool, this instrument was validated by a panel of three expert reviewers. These three reviewers (who have asked to remain anonymous) are all science education professors at R1 universities, have multiple publications and presentations covering ST and Computational Modeling, and are not affiliated with our home institution. The expert review panel provided critical feedback on how the instrument interacts with existing ST literature (content validity) and whether these indicators are valid means of assessing the desired aspects of ST (construct validity). Once this feedback was received, the instrument was revised to bolster its content and construct validity. Finally, both of the authors along with one additional colleague (Linsey Brennan) independently categorized three 30-min segments (one from the beginning of the unit, one from the middle, and one from the end of the unit) of student screencasts using this instrument. For this interrater reliability test, we independently recorded which testing and debugging indicators were present within each five-minute interval. When we compared our analyses, we achieved an initial interrater reliability of 82%, meaning that we agreed on which indicators were present and the level of sophistication for these indicators for 82% of all these five-minute intervals. Upon further discussion we managed to reach a resolution on all of our coding disagreements.

Data analysis

Once we finished validating the Dynamic ST Tool, we began analyzing the data in four distinct phases. In phase 1, we used the Atlas.ti program with the Dynamic ST Tool to conduct an initial analysis of student screencasts. During this process, we put time stamps on moments where students were using ST and categorized them based on the Dynamic ST Tool. For example, when we saw clear moments where students were Evaluating System Variables, we placed a time stamp to record the number of minutes students spent Evaluating System Variables. We then took detailed notes using the descriptive memos tool of Atlas.ti to summarize student talking points during these conversations to help drive later qualitative analysis. In phase 2, we compiled the Dynamic ST Tool categories for each lesson into a summary table for each student group. For each lesson of the unit, we added up the minutes attributed to each student behavior to create a sum total of minutes spent performing activities associated with various ST behaviors for that lesson. For example, if a student had a conversation involving Evaluating System Variables that lasted 2 min and another conversation involving Evaluating System Variables and Analyzing Single Causal Relationships that lasted 5 min, they would be recorded as having spent 7 min Evaluating System Variables and 5 min Analyzing Single Causal Relationships for that lesson. These sum totals enabled us to identify broader trends in how common different ST behaviors were across the unit as a whole as well how student approaches to ST changed as the unit progressed. For the purposes of this study, we created a frequency table, showing how many minutes each student group spent with a specific ST behavior in the whole unit to help address Research Question 1: How do students apply ST as they build and revise computational systems models in this unit? We then used a separate summary table as a timeline to show how the ST behaviors of all five groups changed as the unit progressed (sum of each behavior for all five groups for each lesson of the unit) to address Research Question 2: How do student ST behaviors change over the course of this unit?

After we established the key patterns of student dynamic ST through these quantitative analyses, we conducted a narrative analysis for each ST behavior (phase 3). We began by returning to the descriptive memos of each group and looked for specific episodes that clearly demonstrated students exhibiting specific indicators. We simultaneously looked for patterns and outliers between student groups, so that we could articulate the main approaches students were using for each ST behavior described in the Dynamic ST Tool and create a cohesive narrative for each indicator. For example, when conducting the narrative analysis for interpreting graphical model output, we rewatched all of the episodes we labeled as examples of this behavior. As we rewatched these episodes we reviewed our detailed descriptive memos of how those students were approaching this behavior, using the "level of sophistication" classification system to scaffold this analysis. After reviewing all of these episodes we summarized the different examples we saw of students interpreting graphical model output and noted which sort of examples were more common and which examples were exemplary. From this summary, we began writing our narrative analysis, including prominent student examples that showed both common and exemplary student behaviors. Once we had written these narratives, we compared them to our quantitative analyses to check for internal consistency. Additionally, another colleague, who independently analyzed the same screencasts and other student data from this same

Table 2 Dynamic systems thinking through modeling analysis tool

Indicator	Level descriptions	Additional information
Evaluating System Variables (EV)	Level 1: Students add, remove, or rename variables but do not provide any verbal explanation for doing so. Level 2: Students provide superficial verbalized explanations that neither address how the variable impacts model behavior nor are necessary for understanding or explaining the phenomenon. Level 3A: Students provide a verbalized explanation that addresses how the variable impacts its immediate neighbors. Level 3B: Students provide a verbalized explanation that addresses the variable's importance to explaining one aspect of the phenomenon. (e.g. why the rate of evaporation slows down). Level 4A: Students provide a verbalized explanation that addresses how the variable impacts model behavior on a broader scale. Level 4B: Student verbal explanations address the variable's importance to explaining two or more aspects of the phenomenon.	There are two ways students can provide evidence for ST through this indicator: model based explanations and phenomenon based explanations. Model based explanations focus on impact of element and model behavior; phenomenon based explanations look at how element is necessary to explain the phenomenon.
Analyzing Single Causal Relationships (SC)	Level 1: Students create, propose, or modify a causal or correlational relationship between two adjacent elements but provide no verbalized or written explanation for this relationship. (Explanations must address "why"). Level 2: Students provide a verbalized or written explanation for relationship **causality** but not for directionality or magnitude. Level 3: Students provide a verbalized or written explanation that addresses the **directionality** of the relationship. Level 4: Students provide a verbalized or written explanation that addresses the **magnitude** of the relationship.	There are three main ways a relationship can be defined: Causality, Directionality, and Magnitude. Causality refers to the order in which two elements interact in a causal chain (does A impact B). Directionality addresses if the two elements have a positive or negative correlation (does A cause B to increase or decrease). Magnitude examines the nuances of the relationship between two elements (is the relationship linear, exponential, logarithmic, etc.)
Analyzing Linear Causal Chains (LC)	Level 1: Students verbally walkthrough a linear causal chain of three or more elements but provide no verbalized reasoning explaining the rationale for any of the relationships in this causal chain. Level 2: As students walk through a linear causal chain, they provide verbal reasoning or critique for one or more individual relationships in this causal chain but do not address the net relationship of the causal chain. Level 3: Student provide a verbalized explanation that addresses the net relationship of the causal chain Level 4: Student verbal explanation addresses the net relationship and explains how one or more mediating variables impacts the net relationship of the causal chain	A linear causal chain is a series of causal relationships involving three or more elements found within a systems model. (A effects B effects C… effects X). A linear causal chain can be said to be composed of individual relationships (A effects B), and an overall net relationship (A effects X). Within linear causal chains are one or more intermediate variables that impact the system. Mediating variables are external elements that have a meaningful impact on the causal chain.
Interpreting Graphical Model Output (MO)	Level 1: Students generate graphical model output but do not verbally (or through writing) interpret the output of their models. Level 2: Students interpret graphical model output in terms of effect (cause and effect or correlation, e.g. when X increases, Y decreases) <u>**but**</u> do not address how the system might change (or remain constant) overtime or the cumulative effects of multiple input variables on model output. Level 3: When students interpret graphical model output they either discuss how the model output changes overtime (or remains constant) <u>**or**</u> the cumulative effects of multiple input variables on model behavior. Level 4: When students interpret graphical model output they discuss both how the model output changes (or remains constant) overtime <u>**and**</u> the cumulative effects of multiple input variables on model behavior.	Graphical model outputs for SageModeler are primarily generated through the simulation feature/minigraphs (but can include student generated graphs). For students to discuss change over time, they need to discuss how the graphical output of their model is showing change over time. To get cumulative effects, they need to discuss how two or more input variables are impacting model output.

Table 2 (continued)

Indicator	Level descriptions	Additional information
Choosing Collector Variables (CC)	Level 1: Students decide to add a single collector to their models or two "incompatible" collectors to their models without discussing the purpose of adding collectors to their models or why these variables specifically should be collector variables. Level 2: Students either add a single collector variable or two incompatible collector variables to their models and discuss either the purpose of adding collectors to their models or why these variables specifically should be collector variables **OR** students add two compatible collector variables to their models (or modify their models so there is now a new set of compatible collectors) with no discussion on the purpose of adding collectors to their models or why these variables specifically should be collector variables **OR** students remove/modify an inappropriate collector variable without giving a verbal explanation. Level 3: Students add two compatible collector variables to their models (or modify their models so there is now a new set of compatible collector variables) and discuss either the purpose of adding collectors to their models **or** why these variables specifically should be collector variables **OR** students discuss why a collector variable shouldn't be present in their model or shouldn't be a collector and delete/modify this variable Level 4: When adding two compatible collector variables to their models (or modifying their models so there is now a new set of compatible collector variables), students discuss both the purpose of adding collectors to their models **and** why these variables specifically should be collector variables.	Collector "compatibility" refers to a pair of collectors that describe substances that can logically transform or transfer between each other within the confines of the phenomenon being modeled. For example, Potential Energy and Kinetic Energy are compatible collectors whereas Kinetic Energy and Temperature are not compatible collectors. The term "Collector" also refers to "sources" and "sinks" which also need to be compatible with an existing collector or set of collectors.
Constructing and Interpreting Collector and Flow Structures (CF)	Level 1: Students construct, delete, or modify a collector and flow structure but provide no verbal reasoning or interpretation of said structure or their interpretation discusses the collector and flow in traditional causal terms (X causes Y). Level 2: When students discuss a collector and flow structure, their discussion correctly interprets these structures as transfer relationships (or otherwise indicate that the valve represents a transformation process) or discusses how the relative amount of the collector variables are changing over time. However, this interpretation does not get into the rationale behind this relationship or discuss how it interacts with other aspects of system behavior. Level 3: When students discuss a collector and flow system they either provide a rationale for including (or removing)this structure in this model ***or*** discuss how the collector and flow structure interacts with other aspects of model or system behavior. (Students must still interpret the collector and flow structure correctly to receive this score). Level 4: When students discuss a collector and flow system they provide both a rationale for including this structure in their model ***and*** discuss how the collector and flow structure interacts with other aspects of model behavior. (Students must still interpret the collector and flow structure correctly to receive this score).	Collectors and flows are complex types of relationships found within dynamic modeling. They represent how one element or variable is transforming or is being transferred from one state into another. This allows students to model how a system can change over time. In the case of evaporative cooling, the Kinetic Energy of Liquid Molecules is being transformed into the Potential Energy of Gas Molecules. Note that "interactions" include discussions on how other variables are affecting the flow rate in the collector and flow system.
Constructing and Interpreting Feedback Loops and Circular Causal Chains (FL)	Level 1: Students create but do not correctly identify a feedback structure or students incorrectly declare a structure to be a loop Level 2: Students recognize or propose a circular/feedback structure in their model but do not discuss its function or effect on their model Level 3: Students provide a verbal explanation of a circular structure but only examine its effect on local behavior (feedback loop structure and immediately adjacent elements) Level 4: Students discuss the effect of feedback loop or circular structure on overall model behavior.	For these systems, we are anticipating that the feedback loop is involved with the collector and flow structures students have built in their models.

implementation was consulted as a form of member checking. This narrative analysis along with our previous quantitative analysis allowed us to fully address Research Question 1: How do students apply ST as they build and revise computational systems models in this unit?

In parallel with this qualitative analysis of each ST behavior, we conducted a thematic analysis of the screencasts to determine aspects of the learning environment that seemed to support students in ST (phase 4). Through our initial examination of student screencasts, we identified two major categories of support from the learning environment: peer sharing/reviews and teacher supports. We then went through all of the screencasts and the classroom videos (which primarily captured Mr. H's pedagogical moves), highlighting key moments where these two aspects of the learning environment were helping students with their emerging ST skills. From this thematic analysis, we were able to write a cohesive narrative for both of these aspects of the learning environment. As with our other qualitative analysis, we had another colleague who was also present during the data collection and data analysis process review our findings. As such, we were able to address Research Question 3: What supports from the learning environment assist students with applying ST in this unit?

Results

Research question 1: How do students apply ST as they build and revise computational systems models in this unit?

Based on our analysis of student screencasts, there is clear evidence of student behaviors that correlate with all seven indicators of the Dynamic ST Tool and that the frequency of these seven indicators is not uniform (Table 3). In the following paragraphs, we will explore how these student behaviors manifested during the process of creating and revising computational models as well as the relative frequency of each behavior.

Evaluating system variables

Evaluating system variables is a common ST behavior students exhibit as they build and revise computational models. Students need to input variables into SageModeler before they can set relationships between these variables, making this an unavoidable part of the modeling process. However, students often set these variables into their models without meaningful discussions as 54% of all instances of evaluating system variables were not accompanied by a verbal explanation (Level 1) and an additional 18% of instances involved a superficial explanation (Level 2). These silent additions of variables make it difficult to ascertain student reasoning behind these additions. When students are less certain about the variables they want to add to their model, they often list out possible variables and briefly discuss their merits. For example, when Amy and Leia are trying to decide what to add to their model Amy asks:

> *Amy: What should we put? Density? IMF strength?*
> *Leia: Doesn't IMF strength affect the speed of particle evaporation?*
> *Amy: Maybe? Let's ask Mr. H.*

This example highlights how Amy and Leia are considering what variables to add to their models alongside how these variables interact with other parts of their models, suggesting an overlap between *evaluating system elements* and *analyzing single causal relationships*.

Students can also evaluate system variables when they are renaming or recontextualizing existing variables. When Robyn and Phillip are testing their model, Robyn reconsiders the variable "temperature of the hand", which is their main outcome variable.

> *Robyn: I feel like we should name this something different because it is not really temperature of the hand itself, it is how it feels. So, this should be "temperature of liquid felt"?*
> *Phillip: How about "Temperature Felt".*

Table 3 Summary table of student ST behaviors

Category	Group 1	Group 2	Group 3	Group 4	Group 5	Total	% of Total Time
EV	39.75	23.5	16	24.75	20.5	124.5	2.8%
SC	72.25	37.25	35	36.25	24	204.75	4.7%
LC	10	7	1	2.5	3.75	24.25	0.6%
MO	51.5	23.5	18.75	51.75	29.5	175	4.0%
CC	8.75	13.75	10.75	7.75	5	46	1.0%
CF	26.5	17.75	22.75	11.25	9	87.25	2.0%
FL	6.25	8	0.5	2.5	7.5	24.75	0.6%

Note that "Percent of total time" is calculated from eleven 80 minute class periods (880 minutes) multiplied by the number of screencast groups (five) to reach 4400 min of total class time

Robyn then changes the “temperature of the hand” variable to “temperature felt”.

Students also discuss where to add variables based on their impact on model behavior. As Robyn adds temperature to their model based on feedback from another group, Phillip questions its position in their model (Fig. 5).

> *Phillip: And why does temperature not affect the model the way he (the other group) had it? (Points to the transfer valve between # of liquid particles and # of gas particles)*
> *Robyn: Because he had it affect evaporation, and this (points to particle evaporation speed) is our evaporation.*

This example shows that the students are not only considering which variable to add to their model, but how the variable should be positioned in relation to other aspects of their model, once again highlighting the interconnectedness between different ST behaviors that occurs during the modeling process. It also shows a more mature understanding of what their “particle evaporation speed” variable represents in this system.

Analyzing single causal relationships

Throughout the modeling process, students frequently need to reconsider and discuss the relationships that exist between variables in their models. Whenever students are setting, modifying, or discussing a relationship between two variables, they are analyzing single causal relationships. As such, this ST behavior is common (205 total minutes) across the entire unit. While students can silently set relationships between variables, they also frequently write out written explanations for these single causal relationships in the box provided by SageModeler. For example, Zion justifies the “more and more” relationship between temperature and evaporation by writing “Higher temperature speeds up evaporation. The higher the temp, the faster the molecules will move, allowing more to escape”, drawing on his growing knowledge to provide a high-level explanation for why an increase in temperature would have an exponential impact on the evaporation rate. Students also often discuss individual

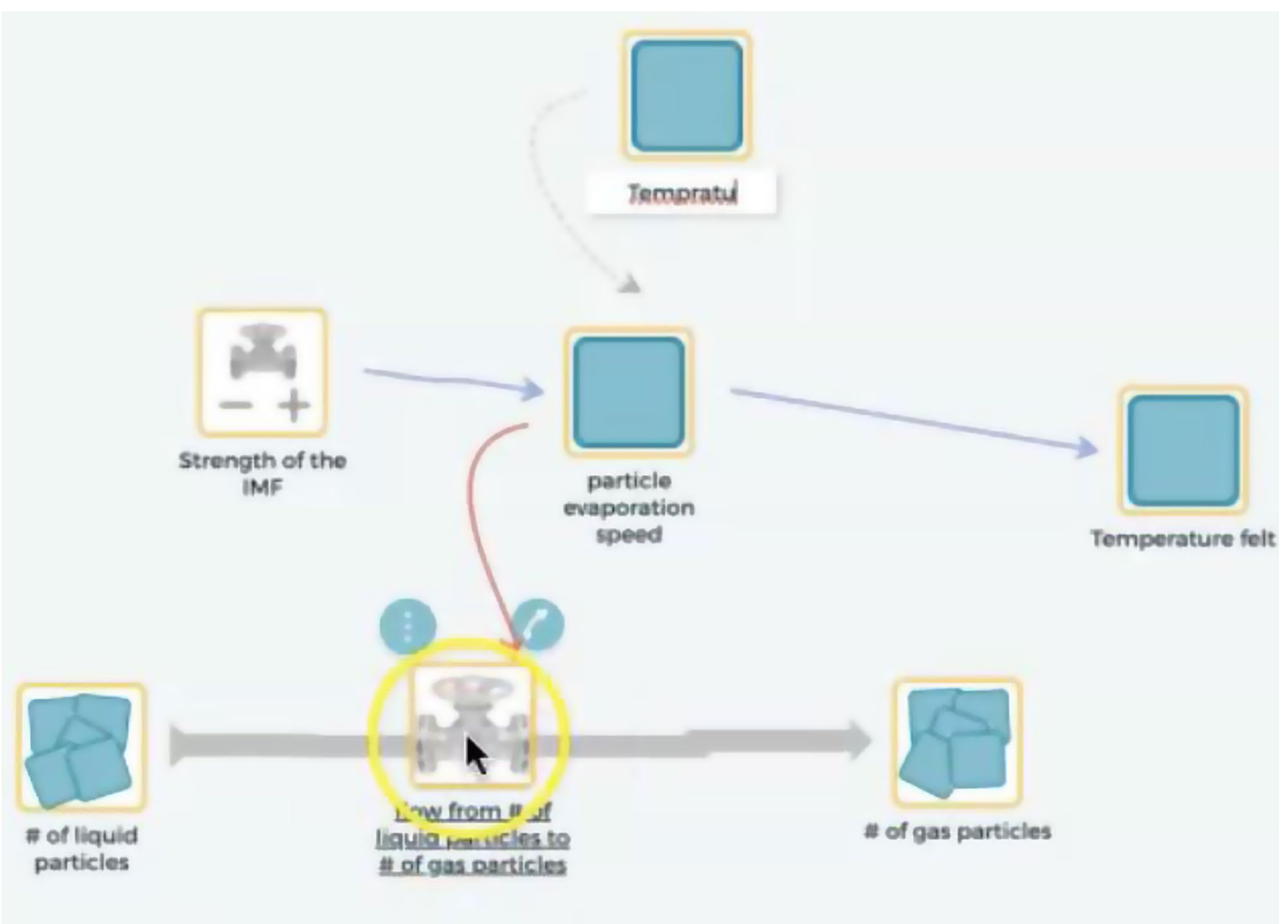

Fig. 5 Phillip and Robyn discuss temperature as a variable

causal relationships when they are using the simulation features present in SageModeler, creating a clear overlap between these two practices.

Analyzing linear causal chains

In contrast to the previous two ST behaviors, analyzing linear causal chains is fairly uncommon across all five groups, with only 24 min recorded across the entire course. When working with their partners, students seldom stop what they are doing and explain their reasoning for an entire causal chain to each other. As such, examples of this behavior tend to occur when students are explaining their models to other people. For example, when Mr. H asked Zion and Conrad to "walk me through your model", Zion responded:

> *Zion: So, we have the temperature of the liquid, which affects the average speed of particles, which affects the number of particles escaping. But for the IMFs, the higher the number, the less particles escape. This affects evaporation and also decreases the temperature which affects the temperature felt on the hand.*

In this example, Zion is going through each linear causal chain in his model, listing individual relationships. While he does not provide reasoning for the individual relationships in these causal chains, nor does he discuss the net relationship, he tacitly acknowledges that multiple pathways work together to influence the outcome by discussing the impact of the IMF on the primary causal chain.

Students also tend to analyze linear causal chains when they are reviewing peer models. When Amy and Leia are looking at another group's model for the first time, they start going through the main causal chain in a linear fashion, listing each relationship, *"So IMF of particles affects the number of particles escaping, which leads to a decrease in temperature… why is this such a mess"* (Fig. 6). In this example, despite Conrad and Zion's model complexity, Amy and Leia impose a linear causal chain explanation to make sense of the phenomenon. In subsequent feedback to Conrad and Zion, Amy and Leia recommend that they "*simplify their model*" and remove unnecessary variables. This example shows the tendency of students to think in linear causal chains and interpret models through that lens even when more complexity is present.

Interpreting graphical model output

In contrast to analyzing linear causal chains, interpreting graphical model output is the second most common form of ST at 175 min across all five groups. As with several other ST behaviors, we can see evidence of students interpreting graphical model output without the need for discourse as students silently move the slider bars up and down. It is important to note that in those cases it is not clear what purpose the simulation is serving, whether it is to make sense of a single relationship, a linear causal chain, or the model as a whole. However, more sophisticated examples of this behavior require that students verbalize their thought processes and therefore reveal how it synergizes with other ST behaviors.

Students also examine model output to make sense of peer models. When looking at Conrad and Zion's model, Phillip and Robyn use the simulate tools to analyze the relationship between IMF and Evaporation (Fig. 7).

> *Robyn: Do we have any questions as to why he included something?*
> *Phillip: Why does the IMF decrease the rate of evaporation?*

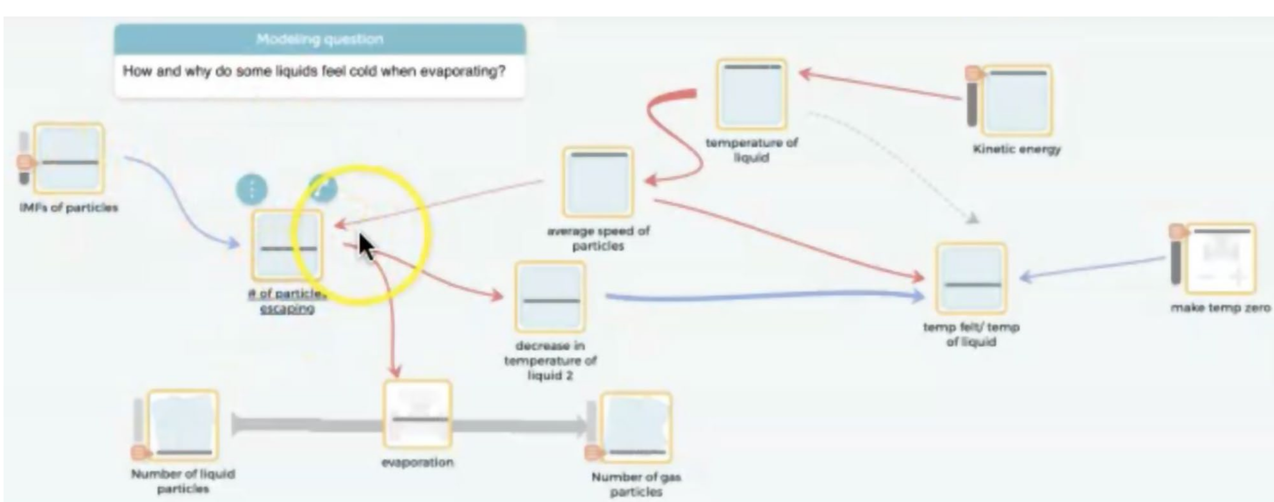

Fig. 6 Amy and Leia's linear causal analysis of Conrad and Zion's model

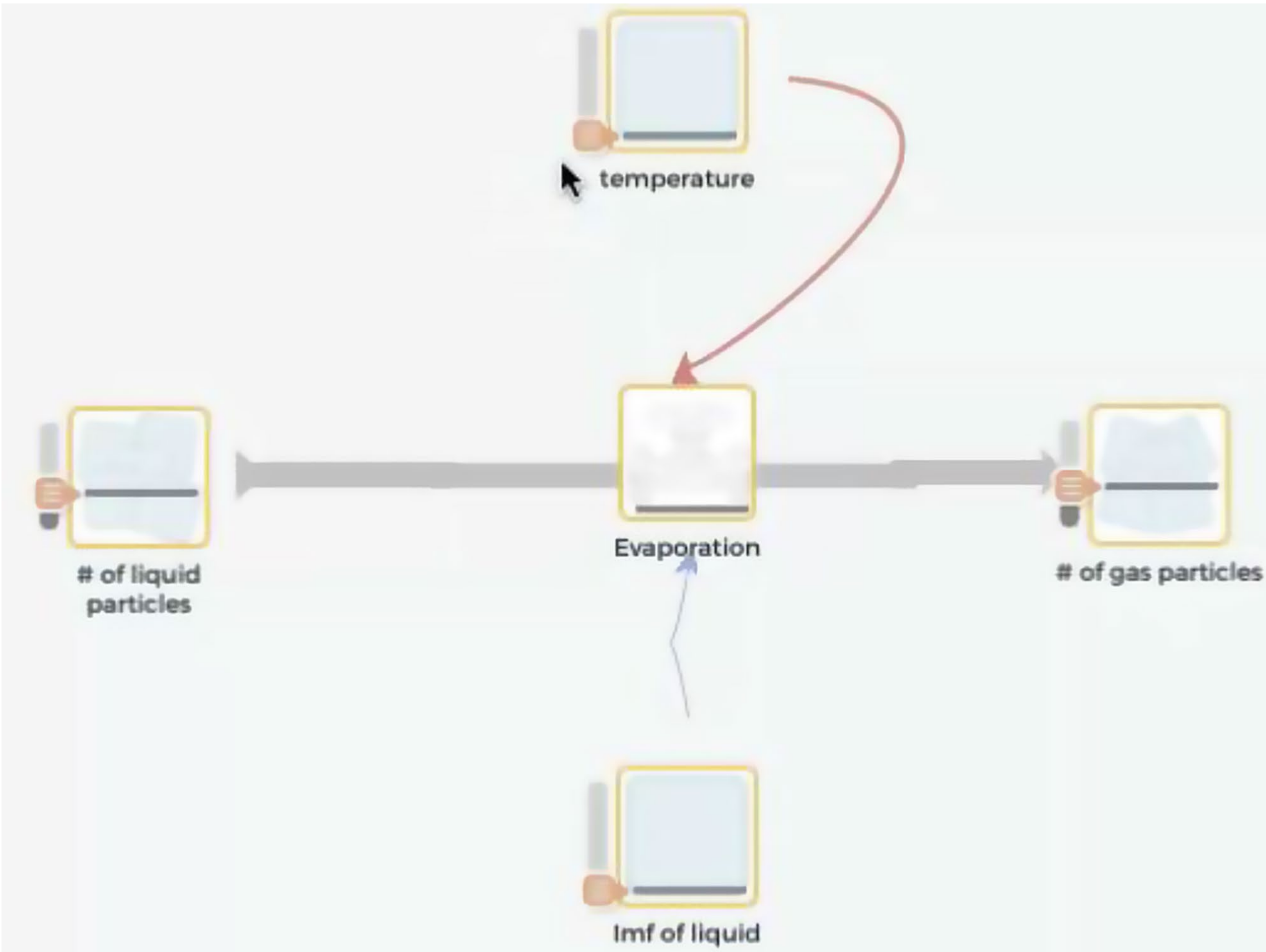

Fig. 7 Robyn and Phillip analyze Conrad and Zion's model

> *Robyn: Oh, fix this (moves the sliders so they are even and then continues to move the IMF slider). It doesn't really change it a lot though.*
> *Phillip: (moves temperature up and down) Temperature does though.*

By using the simulation features, Robyn and Phillip correctly deduce that there is an asymmetric relationship between IMF, Temperature, and Evaporation, with Temperature having a disproportionate impact on the rate of evaporation compared to IMF. When Robyn shares this observation with Zion, Zion is able to defend his design choices.

> *Robyn: (to Zion): I just was wondering why the IMF of the liquid doesn't change it much, but it might not have much to do with it.*
> *Zion: Well, I have temperature to be exponentially increasing (The model shows that as temperature increases, the evaporation rate increases exponentially). So, the lower down the temp is, the less impact the IMF will have.*

In this example, the analysis of graphical model output supports a meaningful discussion on how the convergent impact of these two input variables (IMF and Temperature) impact the overall behavior of this model (as mediated through the rate of evaporation labeled as "Evaporation").

In addition to using the simulate feature some students utilized the graphing feature to analyze their model. After inputting experimental data into SageModeler from the temperature vs. time experiment (where students tested how the temperature of water, acetone, and rubbing alcohol changed over the course of evaporation), Brianna and Kate used the graphing features of SageModeler to look at how their model was measuring temperature over time (Fig. 8). Given that their model output graph showed the temperature increasing over time (which is contrary to their experimental data), the students recognized that they needed to make changes to their models moving forward.

Fig. 8 Brianna and Kate use the graphing features of SageModeler

Choosing collector variables and constructing and interpreting collector and flow structures

As both of these practices deal with student conversations and use of collector and flow structures, they are often deeply intertwined and share a common narrative. All of the student groups started off with an initial "model backbone" of a collector and flow relationship between number of liquid particles and number of gas particles (Fig. 5). Although they were given some instruction on the purpose of collectors and how to set flow relationships, many student groups had difficulty recreating this initial backbone in their models. In the case of Conrad and Zion, they asked Mr. H for help, who showed them how to create a flow relationship between the collector variables. Soon after being shown the mechanics of building a collector and flow relationship, Conrad and Zion successfully recreated the model backbone and labeled the flow as "evaporation", which demonstrates their growing understanding that this collector and flow relationship represents the transfer of liquid particles into gas particles, i.e. evaporation.

As Conrad and Zion began to add on to their backbone, they experimented with adding additional collector variables to their model. In this case, they made temperature a collector that transformed into the number of gas particles. However, once Zion began simulating the model, he noticed that having temperature as a collector did not make sense.

> *Zion: The way you have it set up, temperature is putting liquid into the number of gas particles, I think you have temperature as the wrong kind of thing. I am feeling like temperature would speed up evaporation.*

In this brief comment, Zion recognizes that the way they have their model set up suggests that temperature is somehow being transformed into gas particles. He also makes the observation that temperature is "the wrong kind of thing" and should not be treated as a collector. Finally, Zion suggests that temperature should "speed up" the rate of evaporation. These observations and subsequent revisions to their model demonstrate that Conrad and Zion are able to use this moment as an opportunity to reflect on their model and develop a stronger understanding of how to represent the system through collector and flow structures. This example also shows how students can self-correct and build stronger ST skills through independent practice with SageModeler. It also demonstrates how SageModeler can support students in recognizing which elements can and cannot be represented by collector variables (i.e. transform from one form to another).

Unlike Conard and Zion, other groups often required direct assistance from Mr. H to understand the function of collectors within the model. For instance, early on Amy and Leia had a fairly straightforward model that showed the transfer of liquid to gas being impacted by the temperature (Fig. 9A). While trying to decide how to include the concept of IMF in their model, Leia asked about other variables they should include in their model. Leia asked, "*Should we put speed, like speed of evaporation, or should we put boiling point as well.*" This suggestion to include boiling point, led Amy to replace the collector for "number of gas particles" with "boiling point" and later "how long till boiling point". This single change led to a cascade of other changes until their model, while structurally resembling their earlier example, was largely unrecognizable (Fig. 9B). At this point, they asked Mr. H for guidance who subsequently told them.

> *Mr. H. Remember the backbone with the collectors. The transfer of number of liquid particles to number of gas particles? That shouldn't change throughout the unit as that is the phenomenon we are trying to explain.*

With this reminder to return to the initial collector backbone showing the transfer of number of liquid particles to number of gas particles, Amy and Leia were able to restore their previous model and continue the process of model revision. This example demonstrates the challenges that students faced with making sense of collector and flow relationships in SageModeler and how additional teacher support, in the form of one-on-one conversations, was often critical for students to move forward in the modeling process. It also underscores the importance of a knowledgeable teacher in supporting students with ST in a computational modeling context.

Towards the latter half of the unit, many student groups became increasingly confident in their model backbone of the transfer relationship between the number of liquid particles and the number of gas particles. Therefore, they were less likely to make major structural changes to this model backbone, like Amy and Leia did in the previous examples. However, this did not translate in most cases into students making a parallel collector and flow relationship, showing the transition from kinetic energy to potential energy that also characterizes evaporation and was a learning goal of the curriculum. Instead, students tended to become more conservative with collector and flow relationships in the latter half of the unit (post-potential energy phase) and seldom tried any new approaches to collectors and flows after learning about potential energy. This suggests that these students are still not fully comfortable with independently constructing collector and flow structures in SageModeler. In one notable exception, Conrad and Zion decided to make a secondary flow relationship between gas and liquid particles, this time showing the condensation relationship that

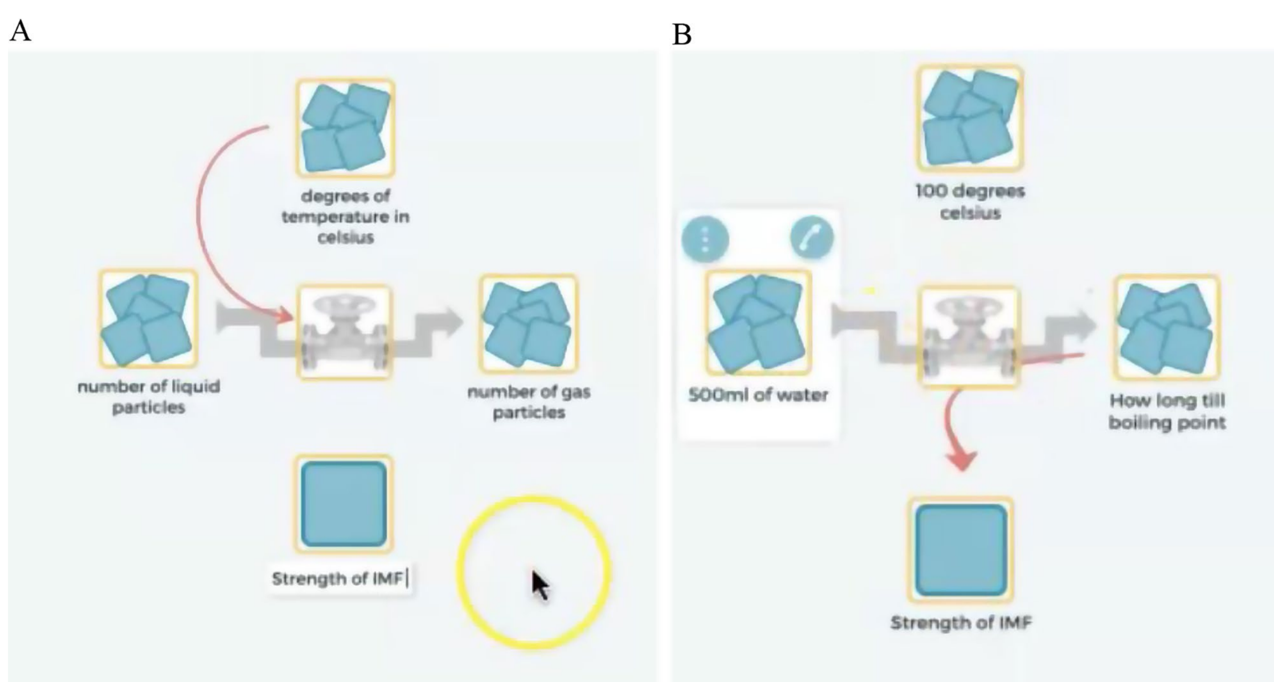

Fig. 9 Amy and Leia's efforts to modify their initial model. **A** Amy and Leia's initial model. **B** Amy and Leia's revised model

can also exist when gas is transferred into a liquid. While this later effort does suggest that this group understood the mechanics and ST concepts underpinning collector and flow relationships, its overall impact on their model was negligible.

Constructing and interpreting feedback loops and circular causal chains

Compared with many of the other aspects of ST analyzed in this study, constructing and interpreting feedback loops and circular causal chains was a fairly uncommon occurrence (25 min). Even when students did create feedback loops, it was often either unintentional or went unaddressed in student discourse. However, there were a few notable instances where students were able to recognize feedback loops in their model. When Amy and Leia were making model revisions, they noticed that they had created a model with two feedback loops (Fig. 10).

> *Amy: Oh, we made two feedback loops.*
> *Leia: Should we have this (points to flow of particles from liquid to gas) affect the IMF?*
> *Amy: Well IMF has to stay out of the feedback loop in order to be controlled, so we can't change it.*

In this example, Amy and Leia correctly identify that there are two separate causal feedback loops in their model (the loop between temperature of the hand, speed of evaporation, and flow of particles and the loop between potential energy, kinetic energy, and flow of particles). While they do not discuss their reasoning behind these feedback loops (which seem to have been created in an ad-hoc fashion rather than being pre-planned structures) nor their impact on model behavior, they do briefly critique a third potential feedback loop. When Leia proposes a third feedback loop involving the strength of IMF, Amy is hesitant to create this loop because she recognizes that IMF is a variable that they want to be able to control and that if it was part of a feedback loop, it would be dependent on other variables and could not be freely manipulated.

In another example, Zion creates a feedback loop and takes note of its impact on model behavior (Fig. 11). When he uses the simulation features to test his model, Zion remarks, "*Oh. we should be able to set this (points to Kinetic Energy). But the thing is, we did make a feedback loop so we can't move it now.*" Here Zion recognizes that he has created a feedback loop structure, and that the circular nature of this structure means that Kinetic Energy no longer functions as an independent variable that he can freely manipulate. While it doesn't appear that he deliberately chose to create this feedback loop structure, he was able to quickly identify its presence in his model and its impact on model behavior, thus showing a relatively sophisticated application of ST to interpreting his

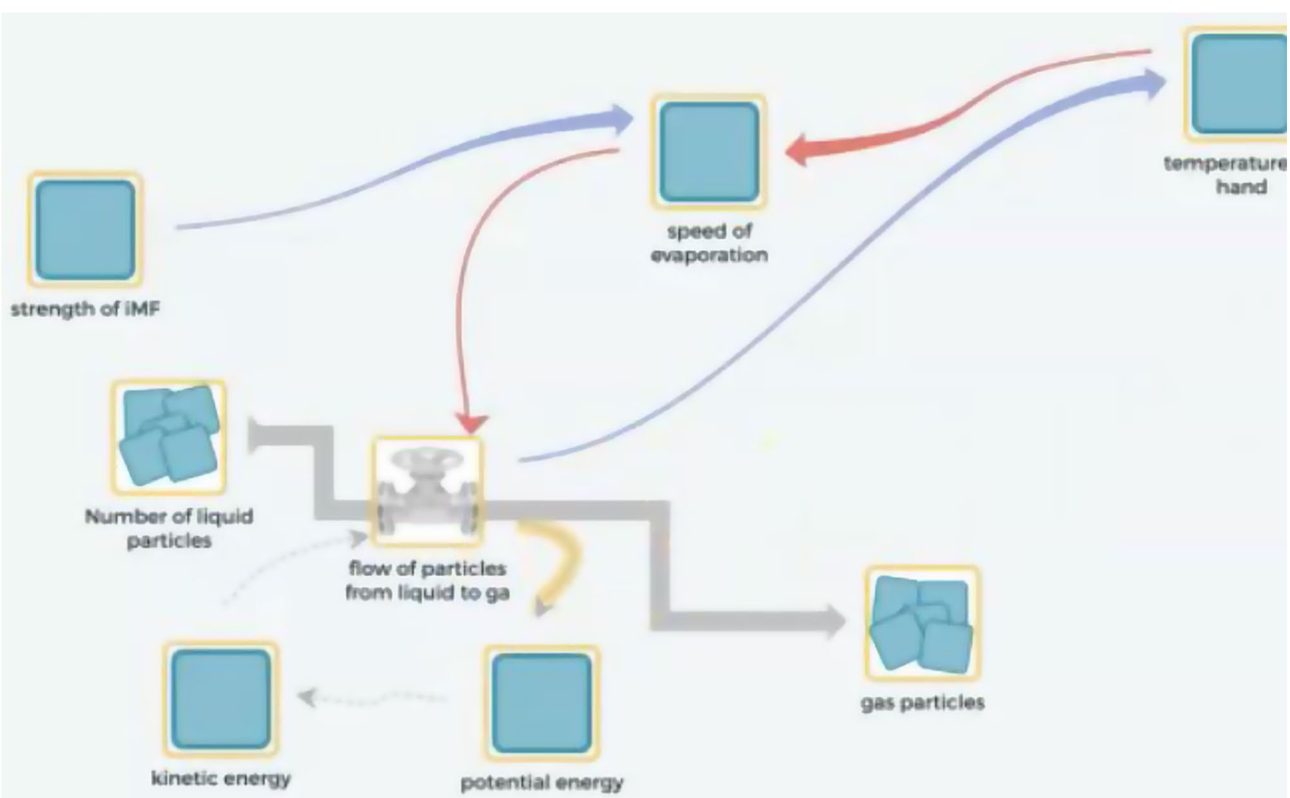

Fig. 10 Amy and Leia's model with feedback loops

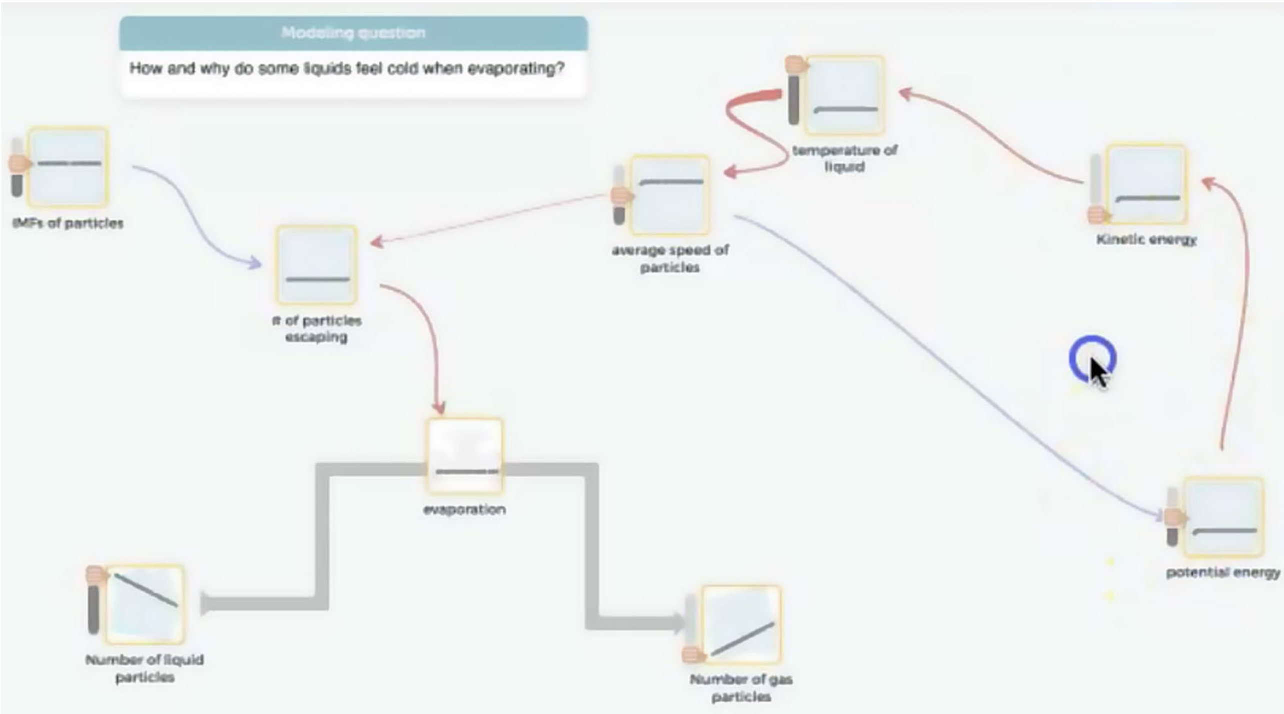

Fig. 11 Conrad and Zion's energy feedback model

model's structure. Zion subsequently removed the direct relationship between potential energy and kinetic energy, thus eliminating this feedback loop structure.

Our case study analysis of 4440 min of collective class time (880 min across five groups) provides evidence of students engaging in a broad range of ST behaviors. Additionally some behaviors, such as analyzing single causal relationships occurred far more common than other behaviors such as constructing and interpreting feedback loops and circular causal chains. Through these screencasts, we came to deeply appreciate the value of student writing and student discourse in understanding student ST. Across multiple examples, there was often a substantial incongruity between student's visual representations of evaporative cooling through their computational models and their written and spoken discourse about their models. Another key finding was the overlap between the various ST behaviors students demonstrated as they built and revised their models in real time. This overlap suggests that a natural synergy may exist between different aspects of ST and further emphasizes how the affordances of computational modeling, particularly the simulation features, support multiple aspects of ST. As such these results strongly suggest that examining student written and verbal discourse during the modeling process can provide invaluable insights into student ST that are overlooked when one only examines student models as a final product or through post-modeling interviews.

Research question 2: How do student ST behaviors change over the course of this unit?

Over the course of the evaporative cooling unit, there was a substantial shift in the ST behaviors of the students in this class (Tables 4A and 4B). These changes in student ST behaviors reflect both the nature of the curriculum for the evaporative cooling unit and changing student priorities as the unit progressed. In general we have subdivided the unit into two distinct halves: the pre-potential energy phase and the post potential energy phase based on whether or not students have been introduced to potential energy in the context of evaporative cooling. The first six lessons represent the pre-potential energy phase, where students are being exposed to new elements of evaporative cooling but have not been introduced to potential energy and thus only have enough information to include a unitary collector and flow system representing the transfer of mass from liquid to gas through evaporation. The final five lessons, or the post-potential energy phase, take place once students have learned about potential energy and can therefore include a secondary collector and flow system that represents the transfer of energy from kinetic to potential through evaporative cooling. In general the post-potential energy phase also focuses on model refinement and validation through the temperature vs. time experiment. Both linear causal analysis and model output analysis were fairly consistent across the unit as a whole. Although some student groups did take advantage of the graphing features

Table 4A Student ST behaviors over the course of the evaporative cooling unit. Student ST behaviors each day of class

Category	L1	L2	L3	L4	L5	L6	L7	L8	L9	L10	L11
EV	0	3.25	0	45.5	24.75	18.25	0	18.5	0	5.5	8.75
SC	0	9.25	0.5	59.5	20	34.5	0	38.5	0.25	20.5	21.75
LC	0	0	0	4	1.75	6.75	0	3.75	0	1.5	6.5
MO	0	3.75	0.5	51.25	13.75	14	0	18	2.5	36	35.25
CC	0	6.5	0	21	6	4.75	0	4.5	0	1.25	2
CF	0	23.75	0	25.5	8.5	11.5	0	7.25	0.75	5	5
FL	0	0	0	0	0	1.75	0	8.25	1	0	13.75

Note L1 = Lesson 1, L2 = Lesson 2…

Table 4B Student ST behaviors over the course of the evaporative cooling unit. Student ST behaviors aggregated into the "Pre-Potential Energy" Phase (First 6 lessons) and the "Post Potential Energy Phase" (Last 5 lessons)

Category	Pre potential energy (minutes)	Pre potential energy percent	Post potential energy (minutes)	Post potential energy percent
EV	91.75	3.8%	32.75	1.6%
SC	123.75	5.2%	81	4.1%
LC	12.5	0.5%	11.75	0.6%
MO	83.25	3.5%	91.75	4.6%
CC	38.25	1.6%	7.75	0.4%
CF	69.25	2.9%	18	0.9%
FL	1.75	0.1%	23	1.2%

of SageModeler after being introduced to these features during the last few lessons of the unit, the total amount of time spent analyzing model output did not change substantially. Because the students did not exhibit any substantial change in the overall amount of time analyzing model output, it appears that some students shifted from using the simulation features to analyze model output towards using the graphical features to compare model output with external data as was intended by the design of this unit (Fig. 8).

In contrast, students spent substantially more time evaluating system variables in the pre-potential energy phase of the unit (92 min; 3.8% of class time) compared to the post potential energy phase of the unit (32.75 min; 1.6% of class time). Such a large drop off suggests that student models began to crystallize around a common set of variables in the post potential energy phase. While this can partially be explained by the fact that the post potential energy phase did not encourage students to add additional variables (once Potential energy had been introduced), it also suggests students becoming more hesitant to remove or reconsider any existing variables. The more modest decline in analyzing single causal relationships (5.2% of class time to 4.1% of class time) also supports the idea of increasing hesitation to modify their models as the unit progressed. However, the 81 min (4.1% of class time) spent on single causal relationships in the post-potential energy phase still demonstrates that students were actively rearranging and modifying the relationships in their models throughout the whole unit.

In a far more dramatic change, students were far less likely to be focusing on ST behaviors associated with collector and flow systems (choosing collector variables and constructing and interpreting collector and flow structures) in the second half of the unit (1.6% and 2.9 % of class time vs. 0.4% and 0.9% of class time respectively). This data strongly reflect the finding that once students have finished trying to make changes with the collector and flow system at the heart of the model backbone, they largely avoid making any further modifications to any collector and flow structures. Such hesitation to work with collector and flow structures in the post-potential energy phase suggests a lack of confidence in their understanding of collector and flow structures and demonstrates the inherent difficulty of creating collector and flow structures. Ironically, the unit was designed to encourage students to begin making a parallel collector and flow structure showing the transformation between kinetic energy and potential energy in the post potential energy phase, as students should have been creating a collector and flow model showing how the kinetic energy of liquids transforms into the potential energy of gas. As such the hesitation of students to work with collector and flow features in the post potential energy phase largely prevented them from making these necessary improvements to their models. This is a clear example of how despite the unit being ostensibly designed to promote a particular behavioral pattern (in this case students adding a parallel collector and flow relationship), student's experiences in the pre-potential energy phase led to a contradictory outcome.

While students spent less time working with collector and flow structures, the amount of time working with feedback loops and circular structures substantially increased in the latter parts of this unit (from 0.1% to 1.2 % of class time) during the post potential energy phase. Given that many of these feedback loops were created by accident, it does suggest that students were more likely to try more complex arrangements of relationships in their models. There were also more instances of students identifying feedback loops that were present in their model, even if they did not always actively interpret their impact on model behavior. Student use of feedback language can be partially explained by the efforts of Mr. H to use the language of feedback loops towards the end of the unit in the post-potential energy phase. However, despite these efforts to include feedback loops and adjacent concepts in the post-potential energy phase of the unit, their relative scarcity suggests that additional efforts are needed to support students in understanding feedback structures in this unit.

Research question 3: What supports from the learning environment assist students with applying ST in this unit?

In addition to finding strong evidence of students using ST as they built and revised models, we also investigated how the learning environment created in this unit supported students in these endeavors. Through examining student screencasts and classroom videos, we found two broad aspects of the learning environment that impacted student use of ST in this evaporative cooling unit: student cooperation through peer groups and teacher pedagogical support.

Student cooperation through peer groups

Throughout the unit, students had many opportunities to work with other peer groups to share ideas and engage in collaborative conversations around ST. As shown in a previous example, the peer review process created an opportunity for Robyn and Phillip to use the simulation features to interpret the relationship between IMF and rate of evaporation in Zion's model (Fig. 7).

> *Phillip: Why does the IMF decrease the rate of evaporation?*
> *Robyn: Oh, fix this (moves the sliders so they are even and then continues to move the IMF slider). It doesn't really change it a lot though.*
> *Phillip: (moves temperature up and down) Temperature does though.*

This peer review session not only facilitated Robyn and Phillip in analyzing model output, but also allowed Zion to share his thoughts on the causal relationship between IMF and rate of evaporation and how this single causal relationship interacted with the relationship between temperature and rate of evaporation.

> *Robyn: (to Zion): I just was wondering why the IMF of the liquid doesn't change it much, but it might not have much to do with it.*
> *Zion: Well, I have temperature to be exponentially increasing (their model shows that as temperature increases, the rate of evaporation will increase exponentially). So, the lower down the temperature is, the less impact the IMF will have.*

This conversation showcases both Robyn and Phillip's ability to interpret model output and Zion's understanding of how two related aspects of his model (temperature and IMF) work together to impact model behavior on a common downstream variable. As such it represents how peer feedback can be a mutually beneficial process strengthening the ST prowess of both student groups.

In another example, Zion offers to assist Amy and Leia who are trying to figure out how to incorporate IMF into their model (Fig. 12). When Amy and Leia reach out to Zion, they first ask for his advice on the relationship between evaporation and the "strength of the IMF". Taking a broader approach to his critique, Zion points out multiple issues with individual relationships in this model.

> *Zion: My first issue is that evaporation rate has no impact on temperature. You have IMF that affects the speed of evaporation, but you also need temperature to affect the speed of evaporation. Also, IMF decreases the speed of evaporation not increases.*
> *Leia: So, IMF decreases the speed because the higher the IMF, the slower it is?*
> *Zion: Yes. Also to answer your first question, evaporation doesn't affect IMF, speed of evaporation would affect evaporation.*

In a fairly rapid succession, Zion suggests that Amy and Leia add in a separate temperature variable (most likely temperature of the hand) to impact the evaporation rate, change the directionality of the relationship between IMF and speed of evaporation, and change the causality of the relationship between evaporation and the causal chain at the top of the model. Such comments show Zion's ST and seem to have helped Amy and Leia make much needed changes to their model as they subsequently incorporated most of these ideas directly into their model.

One can notice the differences between the conversation Zion had with Robyn and Phillip, and the one he had with Amy and Leia. With the latter there was not a mutual process of both groups learning from each other's models and building their ST together. Instead, Amy and Leia mostly took up Zion's ideas without any comments or counterargument and thus were not given the opportunity to engage in a mutual ST discussion as had occurred between Robyn and Zion. The differences in the tone of these two discussions provide insight into the power dynamics that can occur within peer-peer classroom interactions. While Zion felt comfortable in providing a rationale for their modeling choices to Robyn and Phillip, Amy and Leia seemed to defer to Zion as an expert rather than as a mutual peer. Because

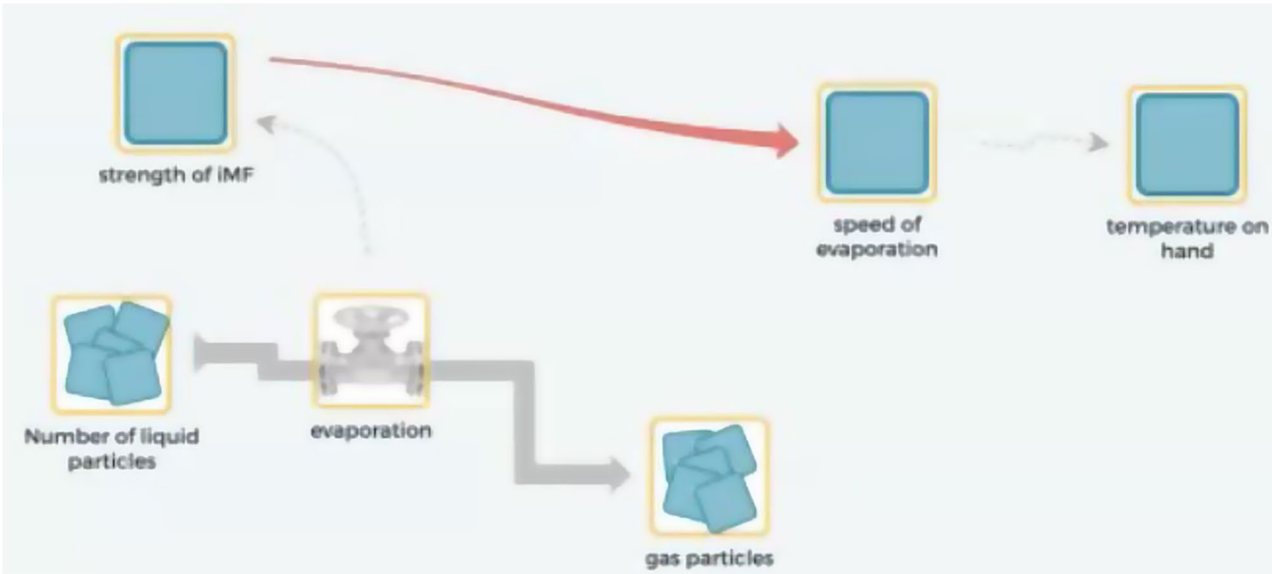

Fig. 12 Zion critiques Amy and Leia's model

peer feedback is often more productive when both parties are sharing authority, the presence of unequal power dynamics suggests that additional scaffolding for peer review conversations is needed to encourage students like Amy and Leia to take a more active role in defending their ideas. It might also be helpful to remind students like Zion to not view other students' models as a puzzle that needs to be solved but as one of many approaches to representing a phenomenon. Lastly, having a more mutualistic peer-peer interaction is likely to be more effective in helping students like Amy and Leia understand why they are making revisions to their models and thus avoid repeating the same modeling mistakes multiple times as what happened with their attempts to modify their backbone collector and flow relationship.

Teacher pedagogical supports

Perhaps one of the strongest assets available to students was Mr. H himself. Mr. H generally encouraged students to take creative liberties with their models and to make productive mistakes in this unit. However, Mr. H also frequently offered advice and guidance to students throughout the modeling process like the one addressed in a previous example (Fig. 9B), in which Mr. H reminded the students of the purpose of the core backbone thereby helping them revise the collector and flow relationship in their model. While Amy and Leia did restore their initial backbone (the transfer relationship between liquid to gas particles), a subsequent revision where they accidently deleted the transfer valve between the number of liquid and number of gas particles inadvertently led to them creating a "sink" (the "speed of conversion variable" with the faucet symbols) in their model, showing that the number of liquid and number of gas particles are removed from the system by the speed of conversion (Fig. 13). They inevitably ask Mr. H for advice.

> *Mr. H.: So, you have all of your collectors going to a sink. Aren't these (number of liquid and number of gas particles) meant to be transferred? Why are they going down?*
> *Amy: I don't know.*
> *Mr. H: Well, it seems that you have somehow turned that (speed of conversion) into a sink instead of a valve. So, all of your liquid and gas particles are being absorbed by the sink rather than transferring to each other. See how they are both decreasing. Here is how you can fix that.*

Mr. H begins by exploring Amy and Leia's model, pointing out the main structural components. He then gives Amy and Leia an opportunity to share their thoughts on the structural components of their model to see if this sink relationship was created intentionally. Upon recognizing that the students needed specific support with understanding the "sink" structure they created, Mr. H provides additional information on the behavioral impacts of the sink structure before helping them restore the initial transfer relationship. In this manner, Mr. H is giving students the information they need to improve their model and to help further their understanding of key aspects of ST. However it is important to note that while Mr. H is initially trying to help Amy and Leia figure out the source of their problem in a manner preserving their agency, his tone changes to be more direct towards the end. While this direct advice likely allowed Amy

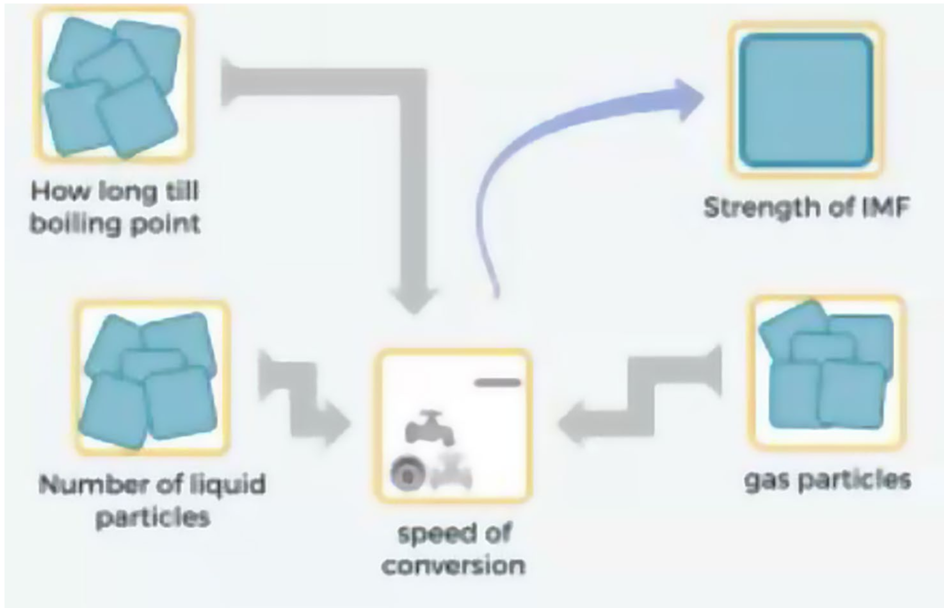

Fig. 13 Amy and Leia's "Sink" model

and Leia to more quickly improve their model, it also removed a key sensemaking opportunity and reinforced a larger pattern of these two students being "recipients of advice" reducing their agency as computational modelers and system thinkers.

In addition to providing one on one support for students needing help with creating and maintaining models with meaningful collector and flow structures, Mr. H also offered whole class instruction on key ST concepts. Soon after assisting Amy and Leia with their collector and flow relationships, Mr. H gave an informational talk on collectors and how to use them in these models.

> *There have been some issues that I have noticed creeping into your models. First off. Collectors are quantitative. They are things that you want to measure, they are things that you want to keep track of and how they flow from one part to another. That is the only time you should use a collector. When you are talking about a quantity of something flowing from one idea to another idea that has a quantity. It's appropriate for us to use collectors to track particles of liquid to particles of gas. It's not appropriate to use a collector to track temperature.*

This brief informational talk in particular seemed to help the screencast students restore the appropriate collector and flow backbone for their models, as many groups (including Amy and Leia) were still uncertain about how collectors should be used in their models prior to this brief lecture. As such it represented a key support for helping students with the ST behaviors of choosing collectors and constructing and interpreting collector and flow structures.

Discussion and conclusion

Discussion

Research question 1 reflection: the importance of in-situ written and verbal evidence for ST

This paper introduces and pilots a novel and exploratory methodological approach for evaluating students' systems thinking (ST) competencies. In contrast to prior work in the field that has focused primarily on assessing students' final model artifacts as proxies for inferring the quality and extent of their applied ST, we present findings that trace the process and progression of students' ST application. This more holistic perspective affords deeper insights into how ST becomes interwoven within students' learning progressions during the modeling process. By examining the unfolding practice of ST rather than mere end products, a more nuanced understanding of the reciprocal interactions between content knowledge acquisition and ST skill development can be garnered. This is particularly true with respect to analyzing student writing and student conversations. Student writing and student conversations throughout this study demonstrate that the presence of a specific relationship within a model does not necessarily mean that students understand the implications of said relationship or what it truly represents within their model. Indeed, such relationships could be crafted by accident, adapted from peer models, or the result of careful discussion and demonstrative of advanced ST competencies. Without having the more holistic context afforded by student written and verbal discourse, it is difficult to interpret the reasoning behind student modeling decisions.

Student conversations and verbal communication can also show the depth of student ST in ways that transcend a traditional structural analysis of student models as a finished product. For example, it is not possible to assess how students are using model output features as a means to drive ST centered conversations and model revisions from traditional pre-post test assessments or through post-modeling interviews. Likewise, the ability for students to provide rich and meaningful feedback on peer models is also a critical aspect of student ST that has not been emphasized in many earlier studies. Indeed, the richness of ST evidence collected from student writing and student verbal communication highlights the limitations of only using student models as evidence of student ST. Instead, these findings suggest that researchers should view and assess ST as a process rather than as a product. While this parallels Hmelo-Silver and colleagues' interview-based approach to assessing student ST (2007), using methods that can capture student ST behaviors during classroom activities can allow for future researchers to gain deeper insights into how students build competence with ST and to identify effective teaching strategies for supporting students in ST. This holistic approach can also help future researchers better understand the synergy between different aspects of ST during the modeling process and clarify how specific aspects of the computational modeling environment are supporting students with ST.

Research question 2 reflection: challenges with collector and flow systems and feedback loops

In addition to demonstrating the importance of written and verbal discourse in assessing student ST, this study reinforces earlier studies showing the challenges students face with understanding collector and flow systems and feedback loops (Assaraf & Orion, 2010; Cox et al., 2019; Cronin et al., 2009; Pallant & Lee, 2017). Although students were given an initial collector and flow system showing the transfer of liquid particles to gas particles as a structural backbone for their initial models (Fig. 4), student efforts to modify this backbone demonstrate a lack of understanding for collector and flow systems. It is important to reiterate that despite the unit being expressly redesigned to support students with creating a second parallel collector and flow system in the second half of the unit (post-potential energy phase), students largely avoided even making revisions to their existing collector and flow relationships in the post-potential energy phase. As such, the absence of efforts to create new collector and flow systems in the later part of the unit strongly suggests that students lacked confidence in their ability to represent change overtime in SageModeler.

As with collector and flow systems, the results of this study suggest that more support is needed for students to fully understand how to construct and interpret feedback loops. Because the curriculum introduced students to feedback loops towards the second half of the unit and Mr. H frequently reinforced the importance of feedback loops, students were able to identify feedback structures in their models. However, these same remarks also suggest that the feedback loop structure emerged through an ad-hoc process rather than being a deliberate model structure. Students identifying feedback loops that are present in their models and unpacking how these feedback loops impact local behavior is indicative of a growing understanding of feedback loops from an ST perspective. However, the absence of deliberately created feedback loops in student models provides evidence that additional support with this aspect of ST is needed.

The findings from this study showing that students have challenges with creating and interpreting collector and flow systems and feedback loops are unsurprising. Previous efforts to create hierarchical models for student ST have generally considered understanding how systems change over time (analogous to collector and flow systems in this study) and recognizing the cyclical nature of systems (analogous to feedback loops in this study) as being more difficult than choosing which variables to include in a model or setting single causal relationships between said variables (Assaraf & Orion, 2005; Monat & Gannon, 2015; Orgill et al., 2019; Stave & Hopper, 2007). In light of previous literature, these results suggest that students need additional support with building and analyzing collector and flow systems and feedback loops. As this was the first experience that these students had with building computational models with SageModeler, it is likely that they needed additional time to master the core mechanics of SageModeler, particularly as it applies to using collector and flow systems, before applying these mechanics to building a working model of evaporative cooling—a challenging concept. However, as several previous studies suggest, the time required for students to fully master the mechanics of SageModeler, particularly as they apply to dynamic time-based modeling, likely exceeds the time allotted for a single unit (Bowers et al., 2023; Eidin et al., 2020, 2023a). Instead, it appears that designing a sequence of computational modeling units where students have more time to develop mastery with different aspects of SageModeler and with the various ST behaviors discussed in this study, might be a more optimal approach.

Research question 3 reflection: the importance of discourse in supporting ST

A third major finding that has emerged from this study is the importance of discourse in supporting students with ST. Constructivist approaches to science education have long acknowledged the importance of discourse in supporting student learning (Gillies, 2008; Osborne, 2010; Premo et al., 2023). Within modeling literature, the benefits of discursive practices, such as sharing and receiving peer feedback, on improving student modeling outcomes are well established (Louca & Zacharia, 2012; Tsivitanidou et al., 2018). However, despite the strong connections established between ST and modeling, there has been little exploration as to how discourse practices benefit students in ST. These results show that when students engage in discourse, either within peer groups or between peer groups, it benefits students in making sense of evaporative cooling as a system of interconnected elements. When working within a dyad or triad, students often need to discuss why they are including specific variables or relationships within their computational model. Such discussions create an opportunity for students to unpack their evidence and reasoning for including these components in their models, thus encouraging them to engage with ST aspects of "defining a system" and "causal reasoning" on a deeper level.

By sharing and receiving feedback from other groups, students can gain insights into different ways the phenomenon can be represented as a system of interconnected elements. If a peer model has a different way of representing how a key aspect of the phenomenon changes over time, they can use this experience to support how they are representing change over time in their own model. Conversely, peer feedback can assist students in recognizing how certain structural elements of their models, such as feedback loops, impact model behavior, encouraging a deeper exploration of how their model functions as a system and supporting further model revisions.

Because discourse is an important practice for allowing students to unpack their ideas of how to represent a phenomenon as a system and for sharing these insights with their peers, it can support ST in science classrooms. It is also important to recognize that rich and meaningful discourse about scientific ideas, including ST, does not come naturally to students (Jiménez-Aleixandre et al., 2000; Lemke, 1990; McNeill & Pimentel, 2010). Quality peer discourse requires teachers to develop a classroom culture that encourages students to respectfully engage in discourse using evidence and for students to learn how to give and receive feedback in a constructive manner (Berland & Reiser, 2011; McNeill & Pimentel, 2010; Tasker & Herrenkohl, 2016). Lastly it is often necessary to consider power differentials that can occur within classrooms and be reinforced through peer discourse practices. When students, especially students of color and female students, are positioned as "listeners" or "receivers of ideas" within small group settings, they often are less likely to take an active role in science sensemaking (Patterson, 2019; Shah et al., 2020; Shah & Lewis, 2019). This is reflected in the experiences of Amy and Leia for whom discourse practices in this unit seemingly reinforced a lack of agency and independence over their model revision process.

Limitations

Although this study offers several key insights into how students use ST as they build and revise computational models, there are a number of limitations that need to be considered. As a case study that focuses on the ST behaviors of five student groups within one classroom, this study likely does not represent all possible approaches students can take towards ST within the context of computational modeling. It is also important to recognize how the magnet school nature of FHS impacted the results of this study. Because this research took place within a STEM magnet school, these students, who were in their second year at FHS, likely have more familiarity with giving and receiving feedback from their peers and using digital learning tools, such as SageModeler compared to other student populations. Therefore, it is likely that student engagement with ST would require additional support from the learning environment if this unit was implemented in a less privileged environment.

The computational modeling context of SageModeler and the curricular context of the evaporative cooling unit also shaped how students could participate in ST and how we could assess student ST in this study. While "interactions between systems" is considered to be a key aspect of systems thinking by many scholars (Bielik et al., 2023; Monat & Gannon, 2015; Verhoeff et al., 2018), the nature of evaporative cooling as a phenomenon, where the kinetic energy of liquid molecules themselves, not external energy (although some external kinetic energy is absorbed into the water from its surroundings), primarily drives the evaporation process, did not encourage students to fully consider outside forces in their computational models. Not only is "interactions between systems" not emphasized by the context of this unit, but the nature of SageModeler also limits efforts to show how the system of evaporative cooling interacts with other systems and how it represents scale. Since SageModeler has students choose from a semi-quantitative set of relationships (or make a semi-quantitative custom graph in select cases), it is difficult to show a dramatic difference in magnitude of relationships using SageModeler. This means that it is challenging to incorporate factors that have relatively small impacts on system behavior, such as humidity in the case of evaporative cooling, into student computational models, without these factors having a disproportionate impact on system behavior. As such, students were limited in their ability to model smaller scale interactions between the phenomenon of evaporative cooling and broader systems (such as the absorption of kinetic energy by the water from the broader classroom environment). Therefore, it is likely that if a different computational modeling program or a different phenomenon served as the foundation for this study that students might have been able to explore some additional aspects of ST not covered by this research. However, we must also acknowledge that all modeling programs (and indeed all models) have limitations and that more sophisticated modeling programs likely would have required additional classroom time for students to master, creating another barrier for engaging in other aspects of ST.

Conclusion and future directions

Key findings

This study investigated how a computational systems modeling unit supported students in Systems Thinking. Based on prior frameworks, coalescing in "A Framework for Computational Systems Modeling" (Arnold & Wade, 2015; Shin et al., 2022; Stave & Hopper, 2007), we developed the Dynamic Systems Thinking through Modeling Analysis Tool with the following seven indicators corresponding to different ST behaviors: evaluating system variables, analyzing single causal relationships, analyzing linear causal chains, interpreting graphic model output, choosing collector variables, discussing collector and flow structures, and discussing feedback loops and circular causal chains. Using this instrument allowed for the categorization and description of how students used ST across this unit. In particular, student written, and verbal communication and discourse provided rich insights into their ST and overall understanding of the phenomenon in ways that were difficult to capture from just examining their models as stand-alone products. Another key finding illustrated that over the course of the unit, students tended to become more reluctant to make major model revisions, particularly with regards to the collector and flow system that formed the backbone of their model. This is despite the second part of the unit being designed to encourage students to make a second collector and flow system to demonstrate the transition from kinetic to potential energy that occurs during evaporation. Conversely, students were more likely to identify and discuss feedback loops in the second half of this unit. Lastly, the results demonstrate that key pedagogical supports, such as providing opportunities for peer review and direct support to students by the teacher, were beneficial in assisting students with ST.

Implications for teachers and curriculum developers

Building on these findings, there are several implications for teachers and curriculum developers. One of the key takeaways from this study is the importance of constructing computational models as a collective endeavor. Our findings demonstrate the beneficial impact of engaging students in constructing computational models at various social interactions, including paired work, peer feedback, and whole-class plenary discussions. This approach resonates with scholarly recommendations to position modeling as a communal practice, emulating the ways in which scientists reach a consensus through dialogue and argumentation (Jordan et al., 2018; Louca & Zacharia, 2012; Tsivitanidou et al., 2018). Research also suggests that facilitating a collaborative and dialogical environment within the context of one scientific practice, such as modeling, can transfer to students' application of other scientific practices (Bierema et al., 2017). As such, we recommend that teachers who are interested in implementing computational modeling units in their classrooms encourage student collaboration and foster a cooperative learning environment where students can build on each other's ideas.

Another major teaching implication is that students need a broader timescale to develop familiarity with computational modeling tools and to build a firmer grasp of ST. Although students showed evidence of many ST behaviors across this unit and had little issue with ST behaviors such as analyzing single causal relationships and evaluating system elements, most students had difficulty with more advanced ST behaviors, particularly with discussing collector and flow structures and with discussing feedback loops and circular causal chains. Such challenges with these more advanced aspects of ST, especially with regards to creating and discussing collector and flow structures, could in part stem from a lack of understanding of how to represent such relationships within the computational modeling context of SageModeler. Therefore, it is likely that more time is needed to support students with both understanding these broader ST behaviors as well as the computational modeling environment of SageModeler. While one possible strategy would be to extend the initial introductory period students spend learning how to use SageModeler, given the length of this unit, we would suggest that future curricular developers and teachers spread out the learning of both ST and SageModeler over a number of units. This would allow for students to develop familiarity with ST and SageModeler over a broader timescale leading to a deeper understanding than is possible in a single six-week unit.

Implications for research

This study also has several implications for how the field might approach research on ST moving forward. Firstly, it demonstrates the importance of discourse as a way of assessing student ST. Because student discourse can elucidate student reasoning for setting key relationships in their models and allow us to understand if certain model structures (i.e. feedback loops) were created intentionally or through ad-hoc tinkering, it provides a depth that is lacking from ST research solely focusing on analyzing models as final products. As such, we advocate for future researchers interested in analyzing student ST to consider discourse analysis as a powerful research tool, both in the form of interviews (as shown in Hmelo-Silver et al., 2007) and through in-situ discourse as demonstrated in this study. Despite the advantages of using discourse to assess student ST, we also recognize the difficulties of scaling discourse analysis for large scale assessments. As such, we are interested in investigating how writing tasks in parallel to student models can be used to create more comprehensive assessments of student ST as part of a future project. Another possible direction for future research would be to explore the possibilities of teaching and learning ST outside of the context of computational modeling.

Supplementary information

The online version contains supplementary material available at https://doi.org/10.1186/s43031-024-00115-7.

Supplementary Material 1

Acknowledgements
We would like to acknowledge the contributions of staff from the Concord Consortium who developed and revised the SageModeler program during this research. We would also like to thank Daniel Damelin and Israel Touitou for helping us develop the Evaporative Cooling unit used in this research. We appreciate Lindsey Brennan for assisting with data collection and for her work with editing the Evaporative Cooling Unit. We also recognize Lynn Stephens and Steve Roderick for their professional feedback with regards to the Dynamic Systems Thinking through Modeling Analysis Tool. Above all, we are grateful to Mr. H for being willing to implement the Evaporative Cooling Unit in his classroom and supporting us in our research endeavors at Faraday High School.

Author contributions
Author 1 is the primary author of this manuscript. He collected most of the data from "Faraday High School", led the development of the Dynamic Systems Thinking through Modeling Analysis Tool, and analyzed the data using this research instrument. Author 1 also wrote most of this manuscript and assisted with developing the Evaporative Cooling unit used in this study. Author 2 was one of the main authors who developed the Evaporative Cooling unit. He also greatly assisted with data collection from "Faraday High School" and helped develop the Dynamic Systems Thinking through Modeling Analysis Tool. Author 2 also coauthored portions of the introduction and methods sections of this manuscript and played a major role in editing this paper in preparation for publication.

Funding
This research was funded by the National Science Foundation under Grant Nos. DRL-1842035 and DRL-1842037. Any opinions, funding, conclusions, or recommendations expressed in this manuscript are those of the authors and do not necessarily reflect the views of the National Science Foundation.

Data availability
Given the nature of our screencast data, which record student audio, we have decided to not make our screencasts available to the public to protect student anonymity. All quantitative data generated from student screencasts have been compiled into Excel spreadsheets. These data will be made available to the public upon request from the corresponding author.

Declarations

Ethical approval
This research was conducted according to the general ethical guidelines common to the field of science educational research. Prior to conducting this research, our methodologies were reviewed and approved by the appropriate institutional review board (Approval Number: E & I 19060-01). Our research protocol was designed to be minimally invasive and to be beneficial to both the student and teacher participants. Teachers, students, and parents gave informed consent and were given the opportunity to opt out of all aspects of this research project. Additionally, all participant names have been anonymized to protect student and teacher privacy. This includes the names of our three expert reviewers who have requested to remain anonymous in this publication.

Consent to participate
Our protocol for providing teachers, parents, and students with the information necessary to make decisions about participating in this research were approved by an institutional review board prior to beginning this research. All teachers, students, and parents were given the opportunity to opt out of participating in this research without penalty. For students participating in the screencast collection process, both the students and their parents signed an informed consent form agreeing to have their screen actions and classroom conversations recorded using Screencast-O-matic. Students and parents were not penalized for opting not to participate in the screencast collection process.

Consent to publish
As part of our consent forms, all students, parents, and teachers were made aware of our intent to publish our findings with the broader academic community. This agreement was made with the understanding that student data would remain anonymized, and that any identifiable data would be removed from the final product of this research. Therefore, all students and teachers received pseudonyms and any images of student faces were edited out to protect student identities.

Competing interests
The authors declare no competing interests.

Received: 14 June 2024 / Accepted: 30 October 2024
Published online: 11 November 2024

References

Abar, S., Theodoropoulos, G. K., Lemarinier, P., & O'Hare, G. M. (2017). Agent based modelling and simulation tools: A review of the state-of-the-art software. *Computer Science Review*, *24*, 13–33.

Arnold, R. D., & Wade, J. P. (2015). A definition of systems thinking: A systems approach. *Procedia Computer Science*, *44*, 669–678.

Arnold, R. D., & Wade, J. P. (2017). A complete set of systems thinking skills. *Insight*, *20*(3), 9–17.

Assaraf, O., & 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.

Assaraf, O., & Orion, N. (2010). Four case studies, six years later: Developing system thinking skills in junior high school and sustaining them over time. *Journal of Research in Science Teaching*, *47*(10), 1253–1280.

Australian Curriculum, Assessment and Reporting Authority (ACARA). (2017). Australian curriculum: F-10 curriculum: Science.

Berland, L. K., & Reiser, B. J. (2011). Classroom communities' adaptations of the practice of scientific argumentation. *Science Education*, *95*(2), 191–216.

Bielik, T., Stephens, L., Damelin, D., & Krajcik, J. S. (2020). Designing technology environments to support system modeling competence. In *Towards a*

competence-based view on models and modeling in science education (pp. 275–290). Springer International Publishing.
Bielik, T., Delen, I., Krell, M., & Assaraf, O. B. Z. (2023). Characterising the literature on the teaching and learning of system thinking and complexity in STEM education: A bibliometric analysis and research synthesis. *Journal for STEM Education Research*, 1–33.
Betley, E., Sterling, E. J., Akabas, S., Paxton, A., & Frost, L. (2021). Introduction to systems and systems thinking. *Lessons in Conservation*, *11*(1), 9–25.
Bierema, A. M. K., Schwarz, C. V., & Stoltzfus, J. R. (2017). Engaging undergraduate biology students in scientific modeling: Analysis of group interactions, sense-making, and justification. *CBE—Life Sciences Education*, *16*(4).
Bowers, J., Eidin, E., Stephens, L., & Brennan, L. (2023). Examining student testing and debugging within a computational systems modeling context. *Journal of Science Education and Technology*, 1–22.
Cabrera, D., Colosi, L., & Lobdell, C. (2008). Systems thinking. *Evaluation & Program Planning*, *31*(3), 299–310.
Campbell, T., & Oh, P. S. (2015). Engaging students in modeling as an epistemic practice of science: An introduction to the special issue of the Journal of Science Education and Technology. *Journal of Science Education and Technology*, *24*(2), 125–131.
Cox, M., Elen, J., & Steegen, A. (2019). Systems thinking in geography: Can high school students do it? *International Research in Geographical and Environmental Education*, *28*(1), 37–52.
Cronin, M. A., Gonzalez, C., & Sterman, J. D. (2009). Why don't well-educated adults understand accumulation? A challenge to researchers, educators, and citizens. *Organizational Behavior and Human Decision Processes*, *108*(1), 116–130.
Damelin, D., Krajcik, J. S., Mcintyre, C., & Bielik, T. (2017). Students making systems models an accessible approach. *Science Scope*, *40*(5), 78–83.
Danish, J., Saleh, A., Andrade, A., & Bryan, B. (2017). Observing complex systems thinking in the zone of proximal development. *Instructional Science*, *45*, 5–24.
Dominici, G. (2012). Why does systems thinking matter? *Business Systems Review*, *1*(1), 1–2. https://doi.org/10.7350/bsr.a02.2012
Driver, R. (2012). Constructivist approaches to science teaching. In *Constructivism in education* (pp. 385–400). Routledge.
Driver, R., Guesne, E., & Tiberghien, A. (1985). Some features of children's ideas and their implications for teaching. *Children's Ideas in Science*, 193–201.
Easterbrook, S. (2014, August). From computational thinking to systems thinking: A conceptual toolkit for sustainability computing. In *ICT for Sustainability 2014 (ICT4S-14)* (pp. 235–244). Atlantis Press.
Eidin, E., Bielik, T., Touitou, I., Bowers, J., McIntyre, C., & Damelin, D. (2020). *Characterizing advantages and challenges for students engaging in computational thinking and systems thinking through model construction*. ICLS.
Eidin, E., Bielik, T., Touitou, I., Bowers, J., McIntyre, C., Damelin, D., & Krajcik, J. (2023a). Thinking in terms of change over time: Opportunities and challenges of using system dynamics models. *Journal of Science Education and Technology*, 1–28.
Eidin, E., Bowers, J., Damelin, D., & Krajcik, J. (2023b). The effect of using different computational system modeling approaches on applying systems thinking. In *Frontiers in Education* (Vol. 8, pp. 1173792). Frontiers.
Fang, F. C., & Casadevall, A. (2011). Reductionistic and holistic science. *Infection and Immunity*, *79*(4), 1401–1404.
Fisher, D. M. & Systems Thinking Association. (2023, February). Systems thinking activities used in K-12 for up to two decades. In *Frontiers in Education* (Vol. 8, pp. 1059733). Frontiers Media SA.
Flood, R. L. (2010). The relationship of 'systems thinking' to action research. *Systemic Practice and Action Research*, *23*, 269–284.
Forrester, J. W. (1994). System dynamics, systems thinking, and soft OR. *System Dynamics Review*, *10*(2–3), 245–256.
Fretz, E. B., Wu, H. K., Zhang, B., Davis, E. A., Krajcik, J. S., & Soloway, E. (2002). An investigation of software scaffolds supporting modeling practices. *Research in Science Education*, *32*(4), 567–589.
Gilissen, M. G., Knippels, M. C. P., Verhoeff, R. P., & van Joolingen, W. R. (2020). Teachers' and educators' perspectives on systems thinking and its implementation in Dutch biology education. *Journal of Biological Education*, *54*(5), 485–496.
Gillies, R. M. (2008). The effects of cooperative learning on junior high school students' behaviours, discourse and learning during a science-based learning activity. *School Psychology International*, *29*(3), 328–347.
Goldstone, R. L., & Janssen, M. A. (2005). Computational models of collective behavior. *Trends in Cognitive Sciences*, *9*(9), 424–430.
Goldstone, R. L., & Wilensky, U. (2008). Promoting transfer by grounding complex systems principles. *Journal of the Learning Sciences*, *17*, 465–516.
Gotwals, A. W., & Songer, N. B. (2010). Reasoning up and down a food chain: Using an assessment framework to investigate students' middle knowledge. *Science Education*, *94*(2), 259–281.
Grapin, S. E., Llosa, L., Haas, A., & Lee, O. (2022). Affordances of computational models for English learners in science instruction: Conceptual foundation and initial inquiry. *Journal of Science Education and Technology*, *31*(1), 52–67.
Haas, A., Grapin, S. E., Wendel, D., Llosa, L., & Lee, O. (2020). How fifth-grade English learners engage in systems thinking using computational models. *Systems*, *8*(4), 47.
Hmelo-Silver, C. E., Marathe, S., & Liu, L. (2007). Fish swim, rocks sit, and lungs breathe: Expert-novice understanding of complex systems. *The Journal of the Learning Sciences*, *16*(3), 307–331.
Hmelo-Silver, C. E., Jordan, R., Eberbach, C., & Sinha, S. (2017). Systems learning with a conceptual representation: A quasi-experimental study. *Instructional Science*, *45*, 53–72.
Hopper, M., & Stave, K. A. (2008). Assessing the effectiveness of systems thinking interventions in the classroom. In *The 26th International Conference of the System Dynamics Society* (pp. 1–26). Athens, Greece.
Jiménez-Aleixandre, M. P., Bugallo Rodríguez, A., & Duschl, R. A. (2000). "Doing the lesson" or "doing science": Argument in high school genetics. *Science Education*, *84*(6), 757–792.
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.
Khajeloo, M., & Siegel, M. A. (2022). Concept map as a tool to assess and enhance students' system thinking skills. *Instructional Science*, *50*(4), 571–597.
KMK (Standing Conference of the Ministers of Education and Cultural Affairs of the Federal States in the Federal Republic of Germany). (2020). *Bildungsstandards im Fach Biologie für die Allge- meine Hochschulreife*. Wolters Kluwer.
Krajcik, J., & Shin, N. (2022). Project-based learning. In R. K. Sawyer (Ed.), *Cambridge handbook of the learning sciences* (3rd ed.). Cambridge University Press.
Lee, T. D., Gail Jones, M., & Chesnutt, K. (2019). Teaching systems thinking in the context of the water cycle. *Research in Science Education*, *49*, 137–172.
Lemke, J. L. (1990). Talking science: Language, learning, and values. Ablex Publishing Corporation, 355 Chestnut Street, Norwood, NJ 07648 (hardback: ISBN-0-89391-565-3; paperback: ISBN-0-89391-566-1).
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.
MacInnis, C. (1995). Holistic and reductionist approaches in special education: Conflicts and common ground. *McGill Journal of Education/Revue Des Sciences de L'éducation de McGill*, *30*(001).
Mambrey, S., Schreiber, N., & Schmiemann, P. (2022). Young students' reasoning about ecosystems: The role of systems thinking, knowledge, conceptions, and representation. *Research in Science Education*, *52*(1), 79–98.
McNeill, K. L., & Pimentel, D. S. (2010). Scientific discourse in three urban classrooms: The role of the teacher in engaging high school students in argumentation. *Science Education*, *94*(2), 203–229.
Meadows, D. (2008). *Thinking in systems: A primer*. Chelsea Green Publishing. White River Junction.
Mehren, R., Rempfler, A., Buchholz, J., Hartig, J., & Ulrich-Riedhammer, E. M. (2018). System competence modelling: Theoretical foundation and empirical validation of a model involving natural, social and human-environment systems. *Journal of Research in Science Teaching*, *55*(5), 685–711.
Metcalf, S. J., Krajcik, J., & Soloway, E. (2000). Model-it: A design retrospective. *Innovations in Science and Mathematics Education*, 77–115.
Michalopoulou, E., Shallcross, D. E., Atkins, E., Tierney, A., Norman, N. C., Preist, C., O'Doherty, S., Saunders, R., Birkett, A., Willmore, C., & Ninos, I. (2019). The end of simple problems: Repositioning chemistry in higher education and society using a systems thinking approach and the united nations' sustainable development goals as a framework. *Journal of Chemical Education*, *96*(12), 2825–2835.
Monat, J. P., & Gannon, T. F. (2015). What is systems thinking? A review of selected literature plus recommendations. *American Journal of Systems Science*, *4*(1), 11–26.
National Research Council (NRC). (2012). *A framework for K-12 science education: Practices, crosscutting concepts, and core ideas*. National Academies Press. NGSS Lead States.
Next generation science standards (NGSS): For states, by states. (2013). The National Academy Press.

Nguyen, H., & Santagata, R. (2021). Impact of computer modeling on learning and teaching systems thinking. *Journal of Research in Science Teaching*, *58*(5), 661–688.
Orgill, M., York, S., & MacKellar, J. (2019). Introduction to systems thinking for the chemistry education community. *Journal of Chemical Education*, *96*(12), 2720–2729.
Osborne, J. (2010). Arguing to learn in science: The role of collaborative, critical discourse. *Science*, *328*(5977), 463–466.
Pallant, A., & Lee, H. S. (2017). Teaching sustainability through system dynamics: Exploring stocks and flows embedded in dynamic computer models of an agricultural land management system. *Journal of Geoscience Education*, *65*(2), 146–157.
Patterson, A. D. (2019). Equity in groupwork: The social process of creating justice in a science classroom. *Cultural Studies of Science Education*, *14*, 361–381.
Petrosino, A. J., Lucero, M. M., & Mann, M. J. (2015). Decentralized thinking and understanding of evolution in K-12 evolution education. *Evolution: Education and Outreach*, *8*, 1–12.
Pierson, A. E., & Brady, C. E. (2020). Expanding opportunities for systems thinking, conceptual learning, and participation through embodied and computational modeling. *Systems*, *8*(4), 48.
Plate, R. (2010). Assessing individuals' understanding of nonlinear causal structures in complex systems. *System Dynamics Review*, *26*(1), 19–33.
Premo, J., Cavagnetto, A., Collins, L., Davis, W. B., & Offerdahl, E. (2023). Discourse remixed: Shifting science learning through talk. *The Journal of Experimental Education*, *91*(2), 336–357.
Raia, F. (2005). Students' understanding of complex dynamic systems. *Journal of Geoscience Education*, *53*, 297–308. https://doi.org/10.5408/1089-9995-53.3.297
Reynolds, M., Blackmore, C., Ison, R., Shah, R., & Wedlock, E. (2018). The role of systems thinking in the practice of implementing sustainable development goals. *Handbook of Sustainability Science and Research*, 677–698.
Richmond, B. (1993). Systems thinking: Critical thinking skills for the 1990s and beyond. *System Dynamics Review*, *9*(2), 113–133.
Riess, W., & Mischo, C. (2010). Promoting systems thinking through biology lessons. *International Journal of Science Education*, *32*(6), 705–725.
Roychoudhury, A., Shepardson, D. P., & Hirsch, A. S. (2017). System thinking and teaching in the context of climate system and climate change. In *Teaching and learning about climate change* (pp. 29–42). Routledge.
Shah, N., Christensen, J. A., Ortiz, N. A., Nguyen, A. K., Byun, S., Stroupe, D., & Reinholz, D. L. (2020). Racial hierarchy and masculine space: Participatory in/equity in computational physics classrooms. *Computer Science Education*, *30*(3), 254–278.
Shah, N., & Lewis, C. M. (2019). Amplifying and attenuating inequity in collaborative learning: Toward an analytical framework. *Cognition and Instruction*, *37*(4), 423–452.
Shin, N., Bowers, J., Roderick, S., McIntyre, C., Stephens, A. L., Eidin, E., Eidin, E., Krajcik, J., & Damelin, D. (2022). A framework for supporting systems thinking and computational thinking through constructing models. *Instructional Science*, 1–28.
Stave, K., & Hopper, M. (2007, July). What constitutes systems thinking? A proposed taxonomy. In *25th international conference of the system dynamics society* (Vol. 29).
Stephens, A. L., Roderick, S., Shin, N., & Damelin, D. (2023). Students do not always mean what we think they mean: A questioning strategy to elicit the reasoning behind unexpected causal patterns in student system models. *International Journal of Science and Mathematics Education*, *21*(5), 1591–1614.
Sweeney, L. B., & Sterman, J. D. (2000). Bathtub dynamics: Initial results of a systems thinking inventory. *System Dynamics Review*, *16*(4), 249–286.
Sweeney, L. B., & Sterman, J. D. (2007). Thinking about systems: Student and teacher conceptions of natural and social systems. *System Dynamics Review: The Journal of the System Dynamics Society*, *23*(2–3), 285–311.
Tasker, T. Q., & Herrenkohl, L. R. (2016). Using peer feedback to improve students' scientific inquiry. *Journal of Science Teacher Education*, *27*, 35–59.
Taylor, S., Calvo-Amodio, J., & Well, J. (2020). A method for measuring systems thinking learning. *Systems*, *8*(2), 11.
Tsivitanidou, O. E., Constantinou, C. P., Labudde, P., Rönnebeck, S., & Ropohl, M. (2018). Reciprocal peer assessment as a learning tool for secondary school students in modeling-based learning. *European Journal of Psychology of Education*, *33*, 51–73.
Verhoeff, R. P., Knippels, M. C. P., Gilissen, M. G., & Boersma, K. T. (2018, June). The theoretical nature of systems thinking. Perspectives on systems thinking in biology education. In *Frontiers in Education* (Vol. 3, pp. 40). Frontiers Media SA.
Whitehead, N. P., Scherer, W. T., & Smith, M. C. (2014). Systems thinking about systems thinking a proposal for a common language. *IEEE Systems Journal*, *9*(4), 1117–1128.
Wilensky, U., & Rand, W. (2015). *An introduction to agent-based modeling: Modeling natural, social, and engineered complex systems with NetLogo*. Mit Press.
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*, 171–209. https://doi.org/10.1207/s1532690xci2402_1
Yoon, S. A., Klopfer, E., Anderson, E., Koehler-Yom, J., Sheldon, J., Schoenfeld, I., Wendel, D., Scheintaub, H., Oztok, M., Evans, C., & Goh, S. E. (2016). Designing computer-supported complex systems curricula for the Next Generation Science Standards in high school science classrooms. *Systems*, *4*(4), 38. https://doi.org/10.3390/systems4040038
York, S., & Orgill, M. (2020). ChEMIST table: A tool for designing or modifying instruction for a systems thinking approach in chemistry education. *Journal of Chemical Education*, *97*(8), 2114–2129.
Zangori, L., Peel, A., Kinslow, A., Friedrichsen, P., & Sadler, T. D. (2017). Student development of model-based reasoning about carbon cycling and climate change in a socio-scientific issues unit. *Journal of Research in Science Teaching*, *54*(10), 1249–1273.
Zhang, B., Liu, X., & Krajcik, J. S. (2006). Expert models and modeling processes associated with a computer-modeling tool. *Science Education*, *90*(4), 579–604.
Zuckerman, O., & Resnick, M. (2005). Children's misconceptions as barriers to learning stock-and-flow modeling. In *Proceedings of the 23rd International Conference of the System Dynamics Society*.

Publisher's Note