International Society of the Learning Sciences

Coherence across Conceptual and Computational Representations of Students' Scientific Models

Nicole M. Hutchins, Vanderbilt University, nicole.m.hutchins@vanderbilt.edu
Satabdi Basu, SRI International, satabdi.basu@sri.com
Kevin W. McElhaney, Digital Promise, kmcelhaney@digitalpromise.org
Jennifer L. Chiu, University of Virginia, jlchiu@virginia.edu
Sarah J. Fick, Washington State University, s.fick@wsu.edu
Ningyu Zhang, Vanderbilt University, ningyu.zhang@vanderbilt.edu
Gautam Biswas, Vanderbilt University, gautam.biswas@vanderbilt.edu

Abstract: We articulate a framework for characterizing student learning trajectories as they progress through a scientific modeling curriculum. By maintaining coherence between modeling representations and leveraging key design principles including evidence-centered design, we develop mechanisms to evaluate student science and computational thinking (CT) proficiency as they transition from conceptual to computational modeling representations. We have analyzed pre-post assessments and learning artifacts from 99 6th grade students and present three contrasting vignettes to illustrate students' learning trajectories as they work on their modeling tasks. Our analysis indicates pathways that support the transition and identify domain-specific support needs. Our findings will inform refinements to our curriculum and scaffolding of students to further support the integrated learning of science and CT.

Introduction

The Next Generation Science Standards (NGSS; NGSS Lead States, 2013) identify modeling as one of the eight science and engineering practices that prepare students for the 21st century workforce. Modeling enables students to generate, integrate, and test theoretical ideas (Lehrer & Schauble, 2015), providing authentic experiences that deepen students' understanding of scientific phenomena. The practice of modeling is also central to the discipline of computer science (CS) (e.g., K-12 CS Framework (2016)). Integrating science and CS has compelled researchers to develop frameworks guiding their synergistic learning in K-12 classrooms (e.g., Sengupta et al., 2013; Weintrop et al., 2016). While educational policy promotes the integration of CS practices as part of science instruction, little guidance is provided on how teachers may support students in meeting these expectations across disciplines (e.g., NGSS Lead States, 2013).

Technology-enhanced environments can scaffold students in computational modeling processes that engage inquiry and problem solving (e.g., Jonassen, Strobel, & Gottdenker, 2005; Keating et al., 2002; Sengupta et al., 2013; Weintrop et al., 2016). Computational modeling experiences need to be anchored to strong underlying conceptual models of the phenomena being investigated in order to leverage the unique affordances of both conceptual and computational model representations. Together, conceptual and computational representations provide a more complete depiction of phenomena and support students in deriving linkages between model representations (Frederiksen et al., 1999). Limited research describes how students make transitions and connections among model representations and identifies the instructional supports they require.

In this paper, we focus on middle school students' learning of science and computational thinking (CT) as they engage in modeling tasks over the course of a 3-week curriculum unit integrating Earth science, engineering design, and CT (Chiu et al., 2019). Our approach centers on understanding students' ability to model the Earth science concept of water runoff by working through a sequence of representations from conceptual to computational modeling. Our study addresses the following research questions: (1) How do we characterize students' learning trajectories from conceptual to computational modeling? and (2) How do the different modeling vignettes that students generate along this trajectory relate to one another and show students' learning of relevant science and CT concepts?

Theoretical and empirical foundations

Supporting scientific modeling using conceptual and computational representations

Scientific modeling enables students to delve deeper into understanding phenomena by explaining and predicting system behavior (Schwarz & White, 2005). Examples of common conceptual models in K-12 science education include models of food webs, planetary orbits, and the water cycle. Recently, there have been efforts to introduce computational modeling methods to help students gain a better understanding of scientific phenomena (e.g., Sengupta et al., 2013; Weintrop et al., 2016). Computational models provide explicit mechanisms for constructing and visualizing phenomena, reasoning about scientific processes as a sequence of events (as opposed to creating aggregated mathematical models), and providing explicit opportunities to study step-by-step changes in model behavior (versus continuous dynamics). Moreover, computational models of scientific phenomena help to contextualize complex CT concepts (e.g., conditional logic) to support CT learning.

However, research has identified that students may have significant difficulties with developing computational models in science. For instance, students may struggle to understand the science concepts underlying their computational models and to represent system variable relationships computational expressions (Sengupta et al., 2013). These challenges may stem from students' difficulty in identifying model components, interactions among components, or the underlying scientific principles underlying the model (e.g., Forbes, Zangori, & Schwarz, 2015). Students may have trouble generalizing or abstracting the scientific processes into variables and expressions (Basu et al., 2016). These findings indicate a need to anchor the design of computational modeling tasks to conceptual modeling activities and to examine the types of support students may need to make the transition from conceptual to computational models.

For effective participation in this important STEM practice, students and teachers need support as they grapple with the dual complexities of scientific and computational modeling. There is limited research that examines how students' science and CT knowledge evolves over the course of a computational modeling unit that introduces science and CT concepts and practices. Additional research can advance understanding of the synergistic relation between science and CT and how learning in each discipline supports and/or impedes learning in the other. These insights will in turn help scaffold integrated learning of the underlying science and CT concepts. In this paper, we analyze students' learning artifacts as they develop a sequence of models of water runoff to examine how instruction can help students transition from conceptual to computational models.

Design perspectives

We integrated three design perspectives in developing an integrated science + CT computational modeling unit:

- **Evidence-centered design (ECD)** - We use ECD (Mislevy & Haertel, 2006) to analyze the integrated science + CT domain, systematically unpack the target science and computing concepts and practices, and identify connections between the science and CT views of the modeling practice. ECD helps designers link the features of instructional tasks and assessments to evidence of students' proficiency with the target knowledge and skills. In this case, we aim to promote both science and CT-oriented modeling proficiency across evolving modeling representations.
- **Maintain coherence across system representations** - In order to address novice students' difficulties in deriving links among the evolving representations, the conceptual coherence needs to be made explicit (Ainsworth, 2006; Frederiksen et al., 1999). For instance, in this study, each of the evolving modeling representations makes explicit the conservation relationship among rainfall, absorbed water, and water runoff at different levels of abstraction and generality.
- **Domain-specific modeling languages (DSMLs)** - We have created a DSML to make it easier for students to express their scientific understanding of the science concepts and relations into a set of computational constructs that model the science phenomena (e.g., water runoff) (Hutchins et al., 2020). DSMLs enable students to place greater focus on the science concepts and system variables without having to attend to the details of program construction.

Overall, we hypothesize that the three design perspectives will support students as they work through the linked modeling representations from conceptual to computational, thus reducing the difficulties they face in integrating science and CT modeling concepts.

Curriculum description

The Water Runoff Challenge (WRC) is a three-week, NGSS-aligned unit that challenges students to redesign their schoolyard using different surface materials to minimize the amount of water runoff after a storm while adhering to a series of design constraints. These include the overall cost and accessibility, while providing for different functionalities for the schoolyard (Chiu et al., 2019). The WRC targets NGSS performance expectations for upper elementary Earth science and engineering design curricula, emphasizing the movement of surface water in a system after heavy rainfall and the human impact of this runoff on the environment.

Figure 1 illustrates our designed learning trajectory and criteria from conceptual to computational model construction. Initially, students were expected to apply the matter conservation principle (science concept): Total rainfall = total absorption + total runoff, through the construction of paper-and-pencil conceptual models. Each subsequent modeling form required application of additional CT concepts to specify the model in a more general form (see Figure 1). To support this, we implemented an intermediate paper-and-pencil Rule Creation task (Figure 2(b)) to elicit an additional representation of the science phenomenon enabling students to express the relation between the science concepts: total rainfall, total absorption, absorption limit, and total runoff. Students are tasked with expressing three scenarios (i.e., when rainfall is greater, less than, and equal to the surface absorption limit) as semi-structured rules. These relations take into account the conservation laws while using conditional logic expressions to specify when different situations apply (e.g., no runoff versus a certain amount of runoff). Students then transfer their rules into a computational model using the given DSML blocks to create the model components (i.e., the three rules). Translating the rules to the computational modeling activity requires additional knowledge of variables and mathematical and relational operators. We detail each activity, below.

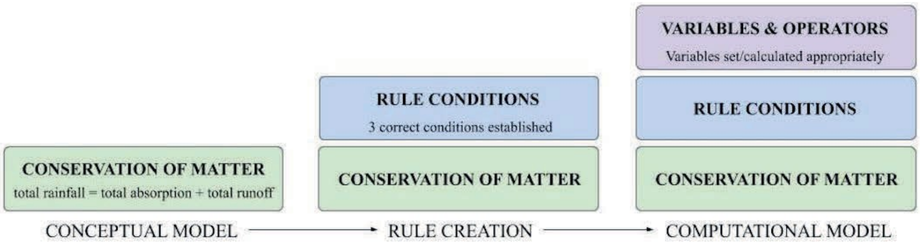
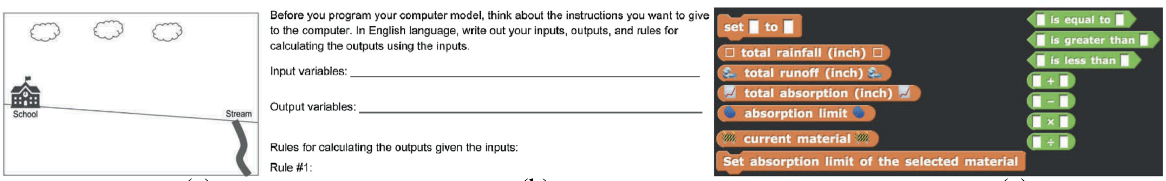


Figure 1. Learning trajectory from conceptual to computational model tasks.

In the WRC, students are introduced to the science concepts of matter conservation and the absorption characteristics of different surface materials. Students then develop pencil-and-paper conceptual models that express the amount of water runoff in terms of the total rainfall and water absorbed by the different materials. As a first step to understanding runoff, students use paper and pencil to generate an input-output model that includes rainfall, absorption, and runoff. Students can describe the conservation relation numerically, pictorially, and/or through a descriptive written response. For instance, to complete Figure 2(a), students are tasked with predicting the amount of absorption and runoff for 3 inches of rainfall and a 1-inch absorption limit of the surface material. As a second step in the model evolution process, the students then create a more precise conceptual model on paper, where they create rules to describe the three different runoff conditions.

After additional scaffolding, where students practiced writing conditional constructs in an unplugged activity (the Rule Creation task, Figure 2(b), described above), the students created their computational runoff model in the computational modeling environment using DSML constructs (Figure 2(c)) that facilitate the translation of the runoff rules into the computational model (Chiu et al., 2019). Example student models appear in the Findings section (Figure 3). The DSML blocks help students assign variables to specific values, and translate their runoff rules to "if" constructs (e.g., "if total rainfall is greater than the absorption limit", then "set total runoff to [total rainfall —absorption limit]"). Students also needed to assign the value of total rainfall and the absorption limit before the conditional block statements. After students constructed a working computational runoff model, they could study the effects of different surface materials on runoff from the schoolyard to the surrounding area.

In designing each of the model building activities, we maintained coherence across the three representations, and gradually introduced students to CT concepts and practices. This approach provides a framework for evaluating students' modeling artifacts across different representations and how these representations support students' learning trajectories.



(a) (b) (c)

Figure 2. Task examples for paper-and-pencil conceptual modeling (a) and rule creation (b), and the DSML used for the computational modeling task (c).

Methods

We conducted a three-week classroom study with 99 sixth-grade students in the U.S. using the WRC. All participating students had some prior programming experience with block-structured programming using Scratch (Maloney et al., 2004). The participating teachers were experienced science teachers and received four days of professional development before the study. Three researchers provided additional support but mostly acted as observers during the study. Students worked for 45 min per day, three days a week during their regular science classes, and 75 min, twice a week with additional personalized-learning time.

Data sources, scoring, and analysis

We examined three primary types of student data: paper-based artifacts, students' final computational models, and a pre-post assessment. All tasks evaluated for this paper were completed in class.

Students' paper-based **Conceptual Models** were coded based on whether their representation of the conservation of matter principle (see Figure 1) was *mechanistic* (and correct), *numeric* (and correct), or *developing*. In order to achieve a *mechanistic* score, students were required to show or use text (via pictorial or descriptive representation) the causal relations resulting in the division of the total rainfall into absorption of water and the amount of runoff using a correct algebraic or numerical expression. A *numerical* score indicated the representations showed all of the required elements and correct values, without describing the causal relations. A *developing* score reflected misunderstandings or errors in students' application of the conservation of matter principle. Students' numerical, pictorial, and written descriptions were coded separately, and students' understanding was represented by the highest score they received on a single part.

For the **Rule Creation** task, each rule that students developed was scored separately. For Rule Conditions (see Figure 1), students received scores for expressing the correct conditional relation between total rainfall and absorption limit (e.g., if total rainfall is greater than the absorption limit). The conservation relation was scored for a correct expression of the values for each required output: total absorption and total runoff.

Students' **Computational Models** were scored using a predefined rubric targeting the application of the conservation of matter rules, the conditional statements for the different rules, and the variable assignments (for rainfall, absorption, runoff). The maximum score possible is 15. To achieve that, students had to assign appropriate values to total rainfall and absorption limit, generate the code for the three conditional statements based on a comparison of variables (total rainfall and absorption limit) and update the absorption and runoff variables for each of the three conditions. Students were given points for generalizability only if they used expressions comprising variables and operators to express variable values (e.g., setting total runoff to "total rainfall — absorption limit" in the overflow condition) as opposed to just assigning numeric values to variables.

Students completed a paper-and-pencil **pre-post assessment** that was split into a science and engineering component and a CT component. Our science and engineering pre-post assessment aligns with a number of NGSS Performance Expectations (PEs). Students could get a maximum score of 23 points. The CT assessment tasks were aligned with the concepts and practices addressed in the modeling activities (e.g., variables, operations, conditionals, program development) and had a maximum score of 13 points. The rubrics used for coding and scoring these assessments were updated from our previous work (McElhaney et al., 2019). Two researchers received 5 hours of training on the rubrics, graded 5% of the test submissions (randomly selected) together to establish initial grading consistency, and then graded another 20% to establish inter-rater reliability (Cohen's $\kappa\kappa$ at ≥ 0.8 level on all items). All differences in the coding were discussed and resolved before the remaining 75% of test submissions were graded by a single researcher.

To answer our research questions, we summarize class learning trajectories through correlation analysis and present three contrasting student vignettes (with pseudonyms Alex, Marley, and Taylor). Where appropriate, we position the performance of these three students relative to the range of student performances we observed.

Findings

Summary of student performance on modeling artifacts and pre-post assessments

Students successfully completed the WRC as intended. Pre-post assessment scores were determined to be normally distributed and a paired t-test analysis showed significant learning gains in science and engineering ($p < 0.0001$, Cohen's $d = 0.82$) and CT ($p < 0.0001$, Cohen's $d = 0.83$). We classified 32 students' Conceptual Models as mechanistic, 59 as numerical, and 7 as developing (1 student packet missing). For the Rules Creation task's Rule Condition component, 57 students correctly described the three conditions while 35 had errors in at least one rule condition (remaining responses were either illegible or missing). For the conservation of matter component, 35 students correctly calculated both absorption and runoff for each rule with 63 students incorrectly calculating or missing elements (1 student packet missing). The main issue with the conservation of matter performance during rule creation were missing variable assignments for either runoff or total absorption in each condition (e.g., students would only describe the resulting runoff when describing what happens when total rainfall is greater than the absorption limit). The Computational Model mean score for the class (we could retrieve 62 student models) was 13.75 (stdev = 2.42), which we will use as reference during the case studies. Spearman's Rho Correlation analysis indicated a moderate but significant correlation between the pictorial representation and Rule-based conceptual model scores (r = 0.35, p = 0.0007, n = 90), but a small non-significant correlation between Rule-based model and Computational Model scores (r = 0.19, p = 0.13, n = 62), implying the relation may not be linear. We observed that 68% of the students had correctly working computational models and 92% implemented the three conditions correctly. These results suggest that students' understanding of the runoff system improved upon constructing their computational models (though some students received help from the teachers or the researchers when constructing their models). For a deeper evaluation of students' modeling processes over time, we present contrasting vignettes for three students. Each student's final computational model is shown in Figure 3.

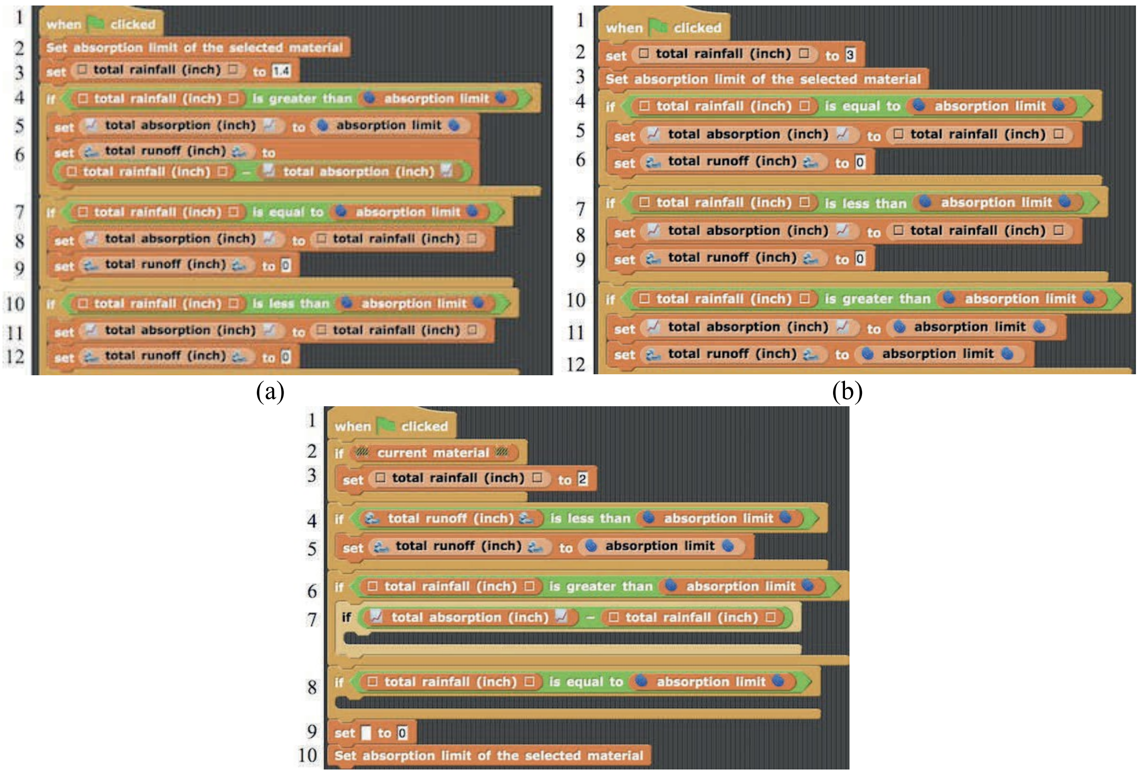


Figure 3. Final code for Alex (a), Marley (b), and Taylor (c).

Student 1: Strongly integrated science and CT

We selected Alex (name altered) for their strong performance on the pre-post assessment and curricular activities. On the pre-post assessment, Alex's score improved from 17.5 to 20 on the science and engineering assessment and from 10 to 13 and on the CT assessment.

During the conceptual modeling task, Alex correctly modeled each rule and demonstrated mechanistic understanding of conservation of matter. In response to the prompt to model water flow for 3 inches of rainfall and a 1inch absorption limit, Alex wrote: "*1 inch of that gets absorbed into the ground and since that's the absorption limit, the rest of the two inches becomes runoff*." This response indicated that Alex correctly calculated the runoff and total absorption based on the absorption limit and total rainfall and correctly described the process and mechanistic causal relations.

Alex's conservation of matter knowledge transferred to the Rule Creation task, where they could correctly define each rule (both the conditions and the output) via descriptive written responses. For example, in describing the "greater than" rule, Alex wrote: "*If the total rainfall is greater than absorption limit, set absorption to absorption limit, set total runoff to total rainfall—absorption*." This also translated to their building the correct computational model for a score of 15 out of 15. As shown in Figure 3(a), Alex correctly built each condition (lines 4, 7 and 10) and set each variable in the correct location and in a generalizable form (e.g., for the greater than condition (lines 4-6), the student calculated total runoff using the subtraction operator and used the variables total rainfall and total absorption, after setting total absorption to the absorption limit). Using our conceptual framework (Figure 1), Alex's learning trajectory is illustrated in Figure 4.

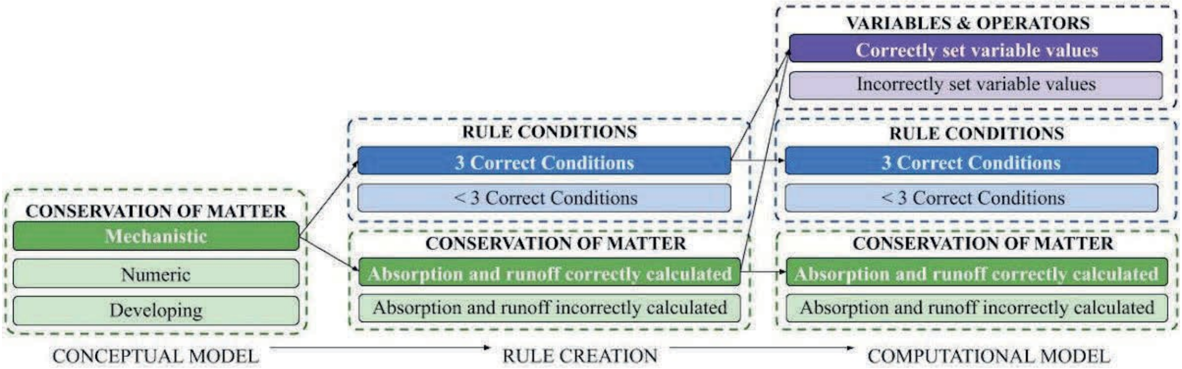


Figure 4. Alex's learning trajectory.

Student 2: Conservation of matter difficulties, high CT proficiency

Marley demonstrated pre-post CT learning gains (improving from a score of 6 to 8), but Marley's science and engineering scores decreased from pre-to-post (from 14 to 11). These results are consistent with Marley's trajectory from conceptual to computational model as described below.

Marley appeared to have difficulties implementing the conservation relation during the conceptual modeling task. For instance, when illustrating what happens when rainfall is less than the absorption limit, Marley indicated an absorption total greater than the total rainfall and stated "*it absorbs more rainfall than actual rain*" indicating a potential confusion about total absorption and absorption limit. During the Rule Creation task, Marley showed knowledge of conditional logic, providing a written response addressing each condition (e.g., "*If total rain = absorption [limit]*"); however, they did not correctly define the outputs for each rule indicating they did not fully understand the conservation law. Marley's ability to apply conditional logic translated to the computational modeling task. Their final computational model (Figure 3b) shows that Marley eventually built correct conditional blocks for two rules (lines 4-6 and 7-9), but struggled with the condition where rainfall exceeds the absorption limit (lines 10-12), thus receiving a score of 12 out of 15.

These results indicate that Marley leveraged conditional logic understanding to support the translation of the conservation of matter concepts into computational form (e.g., decomposing the code into conditional parts and then constructing blocks needed for each output). However, difficulties in the science domain may have limited their ability to correctly build and debug the model. Marley's learning trajectory is illustrated in Figure 5. While they demonstrated CT proficiency, the persistence of their struggles with the conservation law affected their ability to construct the correct computational model.

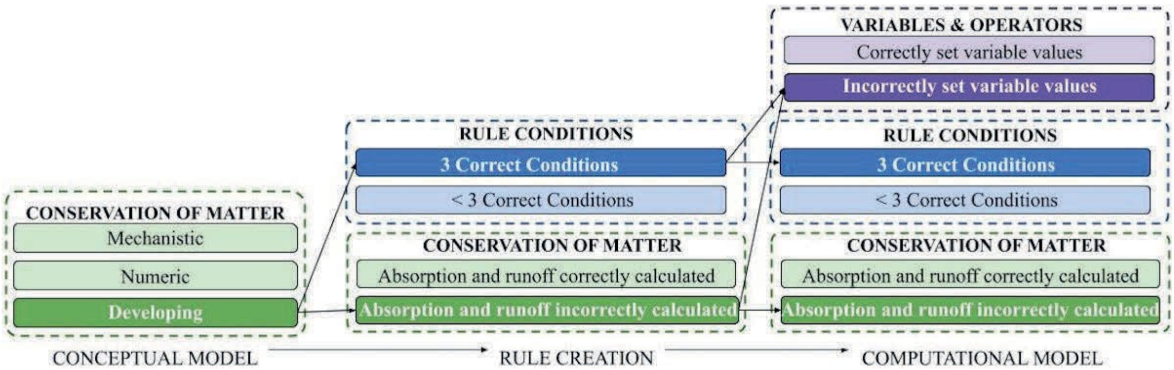


Figure 5. Marley's learning trajectory.

Student 3: Conservation of matter progress, but difficulties with CT integration

Taylor earned a relatively low score on the pre-assessment, but did achieve pre-post gains in both science and engineering (improving from 11.5 to 14.5) and CT (improving from 5 to 8). During the conceptual modeling task, Taylor demonstrated a numerical understanding of the conservation of matter rules. Taylor correctly determined the total absorption and calculated the runoff based on a prompt of a total rainfall of 3 inches and an absorption limit of 1 inch. In their description, Taylor wrote "*the cloud rained 3 inches and the ground absorbed 1 and others become runoff*." This description does indeed include all needed variables, but does not describe the mechanistic reasoning behind why or how to calculate the runoff. However, for the Rules Creation task, Taylor was able to correctly define each rule condition and successfully calculate absorption and runoff for each rule, demonstrating an improved ability to apply the central matter conservation relationship.

The computational modeling task appeared to be difficult for Taylor. Although the science task indicated their ability to correctly apply domain knowledge, Taylor had difficulties translating that knowledge to a computational form. As illustrated in Figure 3(c), Taylor tried multiple arrangements of conditional blocks, including an if-block in which the expression was set to a "*total absorption —total rainfall*" expression (line 7) inside of the if-block for the greater than condition (line 6). Taylor's computational model received a score of 5 out of 15. these code snippets indicate that Taylor understood how to create if-blocks for each rule condition (e.g., total rainfall is greater than absorption limit); however, they seemed unable to translate their domain knowledge to correctly assign variables (e.g., total runoff), debug, or correct their code.

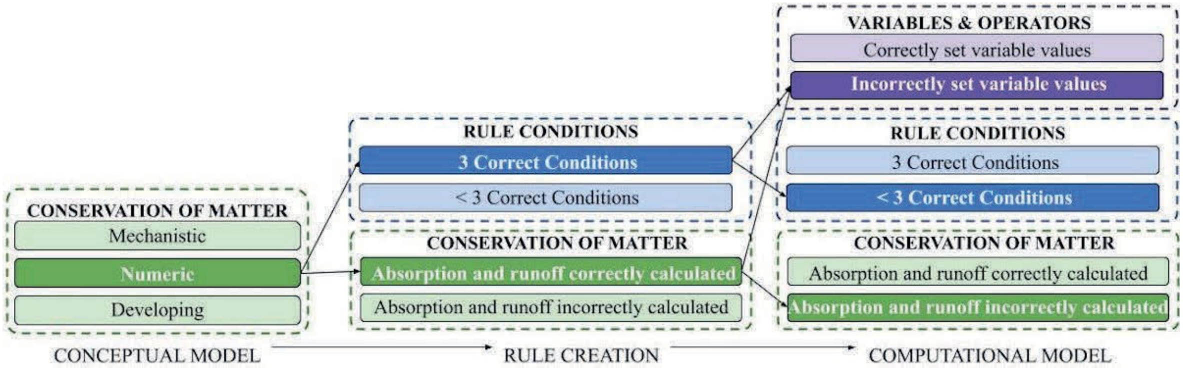


Figure 6. Taylor's learning trajectory.

Interestingly, after being shown the expert computational model and completing the remainder of the tasks, Taylor demonstrated learning gains in both the science and engineering and CT components, illustrating the potential for computational science modeling to help contextualize difficult CT concepts (e.g., conditional logic). Although they were not able to achieve a working computational model on their own, Taylor's successes in prior domain-knowledge application and usage of the model to problem-solve suggest improvements in their CT knowledge and skill. Taylor's final learning trajectory is shown in Figure 6.

Discussion and future implications

The overarching goal of our analysis is to develop linked model representations to scaffold students' computational models in science and understand how students transition across model representations to identify situations where they may have difficulties. This is especially critical in the context of integrated science and CT instruction and serves to disentangle the contributions of each domain and determine how students apply science and CT knowledge across model representations. Leveraging our conceptual framework (Figure 1), our analysis examines the impact of science domain knowledge on CT applications (and vice versa). Overall results showed that while there was a strong correlation between the conceptual model and the rule-based model, there was not a strong correlation between the rule-based model and the computational model, indicating additional personalized support may be needed for the more complex computational modeling task. For instance, although Marley was unable to correctly apply the conservation of matter rules during the conceptual modeling and rule creation tasks, their CT abilities guided them to an almost complete computational model and a high CT posttest score. Alternatively, Taylor indicated improvements in science, but demonstrated difficulties in translating that knowledge to a computational form. In these cases, using the curriculum design and student performance, we can identify the science and CT specific supports needed by the student in the computational modeling environment, and help students develop successful learning-by-modeling trajectories. In addition, these results align with other research findings that CT can serve as a vehicle for learning STEM concepts, but limitations in domain understanding may also impede computational model construction (cf. Sengupta et al., 2013). In terms of curriculum design, the majority of students indicated some proficiency in developing if-blocks in the computational model corresponding to the rules created in the Rules task. We believe this finding highlights a successful implementation of our coherence principle to support the transition from conceptual to computational modeling.

As schools move toward increased integration of computation in K-12 STEM classrooms, the learning sciences community must advance its understanding of the learning processes and support needs of students. Our analysis provides an exploratory step in identifying such cases, deepening understanding of how students integrate domain and CT knowledge for the development of multiple modeling representations. The identification of domain-specific instances of support along the learning trajectory may be supportive of more personalized (individual or group) feedback by the system or teacher. In addition, to promote student learning, our curriculum design approach and the coherence among modeling representations provide a systematic framework for evaluating applications of science and CT concepts over time. We believe this approach can inform curriculum design and scaffolding approaches that deepen understanding of how the domains are integrated, how students translate that knowledge to new representations, and where students may need additional support. Further research in applying our approach to other domains may support generalizability.

References

Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and instruction*, *16*(3), 183-198.

Basu, S., Biswas, G., Sengupta, P., Dickes, A., Kinnebrew, J. S., & Clark, D. (2016). Identifying middle school students' challenges in computational thinking-based science learning. *Research and Practice in Technology Enhanced Learning*, *11*(1), 1-35.

Chiu, J., McElhaney, K. W., Zhang, N., Biswas, G., Fried, R., Basu, S., & Alozie, N. (2019, April). A principled approach to NGSS-aligned curriculum development integrating science, engineering, and computation: a pilot study. Paper presented at the *NARST Annual International Conference*, Baltimore, MD.

Frederiksen, J.R., White, B.Y. and Gutwill, J. (1999), Dynamic mental models in learning science: The importance of constructing derivational linkages among models. *J. Res. Sci. Teach.*, *36*, 806-836.

Forbes, C. T., Zangori, L., & Schwarz, C. V. (2015). Empirical validation of integrated learning performances for hydrologic phenomena: 3rd-grade students' model-driven explanation-construction. *Journal of Research in Science Teaching*, *52*(7), 895-921.

Hutchins, N.M., Biswas, G., Zhang, N., Snyder, C., Lédeczi, Á., & Maróti, M. (2020). Domain-Specific Modeling Languages in Computer-Based Learning Environments: a Systematic Approach to Support Science Learning through Computational Modeling. *Int J Artif Intell Educ*, 30, 537–580.

Jonassen, D., Strobel, J., & Gottdenker, J. (2005). Model building for conceptual change. Interactive Learning Environments, 13(1-2), 15-37.

K–12 Computer Science Framework. (2016). Retrieved from http://www.k12cs.org.

Keating, T., Barnett, M., Barab, S. A., & Hay, K. E. (2002). The virtual solar system project: developing conceptual understanding of astronomical concepts through building three-dimensional computational models. *Journal of Science Education and Technology*, *11*(3), 261-275.

Lehrer, R., & Schauble, L. (2015). The development of scientific thinking. In R. M. Lerner, L.S. Liben, & U. Mueller (Eds.), *Handbook of child psychology and developmental science*, *2*(7), 671-714.

Maloney, J., Burd, L., Kafai, Y., Rusk, N., Silverman, B., & Resnick, M. (2004) Scratch: A sneak preview. In Proceedings of Creating, Connecting, and Collaborating through Computing, 104–109.

McElhaney, K.W., Basu, S., Wetzel, T., Boyce, J. (2019). Three-dimensional assessment of NGSS upper elementary engineering design performance expectations. Paper Presented at the *2019 NARST Annual International Conference*.

Mislevy, R. J., & Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice*, *25*(4), 6-20.

NGSS Lead States. (2013). *Next Generation Science Standards: For states, by states*. Washington, DC: National Academies Press.

Schwarz, C. V., & White, B. Y. (2005). Metamodelling knowledge: Developing students' understanding of scientific modelling. *Cognition and Instruction*, *23*(2), 165–205.

Sengupta, P., Kinnebrew, J.S., Basu, S., Biswas, G., Clark, D. (2013). Integrating Computational Thinking with K-12 Science Education Using Agent-based Computation: A Theoretical Framework. *Education and Information Technologies*, *18*(2), 351-380.

Weintrop, D., Beheshti, E., Horn, M., Orton, K., Jona, K., Trouille, L., & Wilensky, U. (2016). Defining computational thinking for mathematics and science classrooms. *Journal of Science Education and Technology*, *25*(1), 127–147.

Acknowledgments

We gratefully acknowledge the support of the National Science Foundation award DRL-1742195. We also appreciate the contributions of other research team members and the suggestions of two anonymous reviewers. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.