High School Science Teachers' Integration of Computational Thinking into Data Practices to Support Student Investigations

Peters-Burton, E. E., Rich, P.J., Laclede, L., Stehle, S., Kitsantas, A. & Cleary, T. (2021, April 7-10). *High school science teachers' integration of computational thinking into data practices to support student investigations.* [Paper presentation]. National Association of Research on Science Teaching, Online

Bringing computational tools into the science classroom gives students a more authentic view of the discipline (Augustine, 2005; Bailey & Borwein, 2011; Foster, 2006). Computational thinking (CT) is an approach to solving problems and designing systems that requires students to think recursively, reformulate problems to see them in a different light, model relevant aspects of problems, and use abstraction and decomposition in tackling complex problems (Wing, 2006). From our perspective, CT can be a useful supplement to instruction as it involves a series of inter-related cognitive processes that students can use to adeptly engage in data analysis during scientific investigations (Weintrop et al., 2016). Often, the processes of data analysis are too prescribed or missing in high school classrooms, denying students the opportunity to engage in CT practices, such as looking for patterns, breaking down problems into smaller components, or transferring these skills to other problems (Weintrop et al., 2016). The purpose of this study was to examine how high school science teachers learn about and use CT practices to support student engagement in data practices and to determine viable intersections between data practices and CT that have utility for enhancing student learning in the context of science investigations.

Weintrop and colleagues (2016) identified five data practices that scientists engage in during investigations: (a) *creating data* – generating data from tools or observation; (b) *collecting data* - gathering and recording data; (c) *manipulating data* - sorting, filtering, cleaning, normalizing, and combining data sets; (d) *visualizing data* - communicating results with visual representation (e.g., graph, chart); and (e) *analyzing data* - extracting meaning from a data set to draw conclusions.

CT practices include decomposition, pattern recognition, abstraction, algorithm building, and automation, and are naturally linked with data analysis tactics to solve many types of problems (Shute, Sun, & Asbell-Clarke, 2017). *Decomposition* involves breaking down a complex problem into less complex sub-problems. The specific approach to problem decomposition can vary, but the purpose is the same—to reduce the main problem into manageable steps or sub-problems. *Pattern recognition* involves identifying, clustering, and modularizing of steps that repeat. The primary purpose of identifying patterns is to cluster related parts of the problem by their recurring feature(s). *Abstraction* is a process of identifying and organizing relevant information and removing unnecessary information. The general purpose of abstraction is to clarify problems and to identify generalizable solutions; essential skills for constructing models in science and engineering. *Algorithm building* is the creation of a series of precisely-defined steps or rules that leads to predictable outcomes to a problem. An algorithm is an unambiguously defined process to address an initial question. It may involve the steps to collect certain data, the steps to analyze that data, or any other defined process. The steps of the algorithm should, if built correctly, lead to a predictable solution of the problem every time, or within a known error chance. Finally, *automation* involves performing a procedure with little or no direct human interaction. This term typically refers to the use of machinery or computers to perform the automation. At this level of CT, the goal is to outsource work so that it reduces or removes the requirement for direct human action in order to achieve the desired outcome. To address this idea, a crosswalk was developed that articulates the opportunities to elaborate on data practices through computational thinking practices (Peters-Burton et al., 2020).

When conducting a science investigation in biology, chemistry, physics or earth science, students often need to obtain, organize, clean, and analyze the data in order to draw conclusions about a particular phenomenon (e.g., why tidal heights change) and to either build or test models. Merging CT and data practices has the potential to result in more effective science investigation lesson plans, potentially leading to better student learning. Students engage with these data practices to make scientific claims from the evidence found in data. The present study was driven by the following research questions: (a) What did teachers learn about CT related to data practices? (b) What CT practices were difficult for teachers to learn? (c) In what ways do teachers integrate CT and data practices in their lesson plans during science investigations? (d) What gaps were there in teachers' integration of CT into science data practices?

Design

Phenomenography was the chosen research design so that we could identify the qualitatively different ways in which people experience, conceptualize, realize and understand various aspects of a phenomena (Martin et al., 1992), in this case the teachers' learning and use of CT in science lessons involving data practices. The teachers participated in a PD consisting of the following components: (a) one-week institute targeting data practices and CT, (b) one-week institute linking self-regulated learning with CT supports during data practices, and (c) monthly meetings during the school year to reflect on implementation. The outcome of the PD resulted in 18 planned and implemented lessons that integrated CT into data practices across different science content areas. Twenty high school teachers located in the same school district in the mid-Atlantic region voluntarily participated in the study (biology (n = 9), chemistry (n = 5), Earth science (n = 2) and physics (n = 4). The district demographic makeup is 19% economically disadvantaged and 17% ELL, 48% White, 22% Asian, 18% Hispanic, 7% African American, and 6% Multiracial.

Measures

Assessment of teachers' familiarity, use, and value of data practices and CT. A questionnaire was developed to assess teachers' familiarity, frequency of use, and perceptions of importance regarding data practices and CT. The questionnaire was administered before the PD, after the PD, and after the school year implementation of the lessons.

Teacher efficacy beliefs in infusing data practices and CT into lesson plans. This scale assessed teachers' efficacy beliefs to use each component of data practices and CT. Teachers were asked to report their range of confidence for each efficacy items to account for differences in teachers experiences with students. The scale was administered before the PD, after the PD, mid-point of the school year implementation, and at the end of the school year implementation.

Assessment of teachers' knowledge and application of CT. This two-part questionnaire had one section with open-ended questions assessing teachers' conceptual understanding of CT, with the second section targeting teachers' skills in applying their CT knowledge to making suggestions to improve a lesson plan as put forth in a lesson vignette. To prevent test-retest effects, two parallel forms were generated for this assessment and were administered before and after the PD, respectively. The assessments were scored with rubrics consisting of four categories: (a) no response, (b) vague, (c) developing, and (d) proficient for each of the five focus CT practices. Individual raters scored a kappa value of .95 without discussion.

Lesson plans and student work products. Eighteen lessons were collected after the PD and assessed for use of data practices and CT supports.

Interviews on planning and implementing integrated CT into data practice lessons. All teachers were interviewed twice individually to discuss their perceptions of planning for and implementing lessons with CT integrated into data practices. The first interview focused on teacher perception of learning CT and data practices in the Professional Development (PD) and planning for lessons integrating CT into science lessons. The second interview focused on teacher implementation of the integrated lessons.

Attempts to minimize bias in this study was accomplished through the use of multiple data sources, data that represented multiple dimensions, interrater reliability, and member checks for correct representation of the results with the teachers in the PD (Maxwell, 2013).

Analysis and Findings

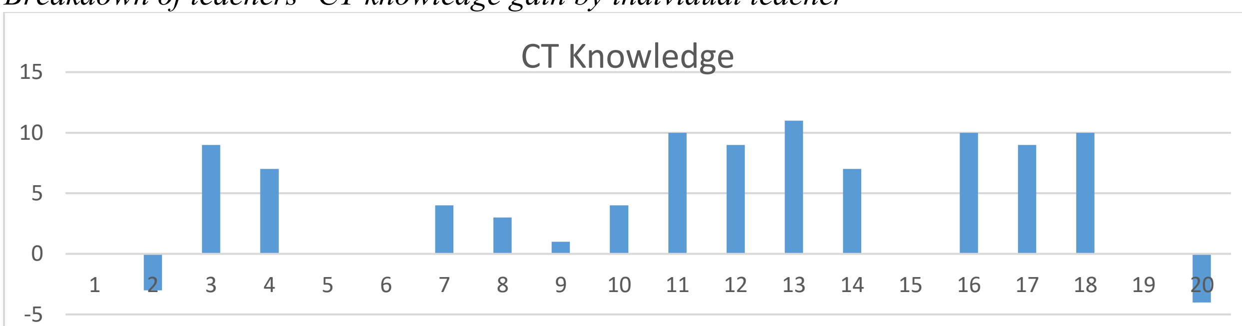
What did teachers learn about CT related to data practices?

As expected, high school science teachers were already familiar with the various data practices (Weintrop, 2006), and showed no significant growth in their familiarity, use, or value of these in the classroom. Perhaps the key take-away from high school science teachers' experience in learning about CT is also the simplest—on the whole they grew in their knowledge of and confidence of how CT might be integrated into high school science. Teachers' familiarity, value and use of CT practices was minimal prior to the PD (see Table 1). Their familiarity increased significantly following the PD ($t(17) = 4.27$, $p = .001$). Likewise, their CT self-efficacy significantly improved. The degree to which each teacher grew in this knowledge and confidence is more nuanced (see Figure 1). In the remainder of this section, we explore these nuances through the teachers' experiences in planning science lessons that integrate CT, and discuss which aspects of CT teachers found more or less difficult to integrate.

Table 1. *High School Teachers Pre and Post Knowledge of CT*
n=90 responses (18 teachers X 5 CT practices)

	No response	Vague	Developing	Proficient
Before PD	72	7	3	8
After PD	33	13	18	26

Figure 1
Breakdown of teachers' CT knowledge gain by individual teacher



What CT practices were difficult for teachers to learn?

There were five CT practices that we emphasized with these teachers: decomposition, pattern-finding, abstraction, algorithmic thinking, and automation. While most of these practices were new to teachers, decomposition was the only CT practice that nearly half of the teachers stated that they felt familiar with prior to the PD. Following the PD, they felt comfortable with both decomposition and pattern-finding and were confident that they already promoted these practices in their lessons.

However, even after PD, science teachers were less sure about their knowledge of abstraction, algorithmic thinking, and automation. This lack of confidence in that specific knowledge continued to manifest itself through their lesson plans, with decomposition and pattern-finding being the most heavily used CT practices. In regards to teacher efficacy beliefs, it is worthwhile noting that their efficacy levels for both data practices and CT not only improved following their PD but were sustained over the academic school year, indicating that the impact of the PD is likely to remain stable even when considerable time has elapsed.

In what ways do teachers integrate CT and data practices in their lesson plans during science investigations? What gaps were there in teachers' integration of CT into science data practices?

To get a better sense for how the different content areas each approached the use of CT in their lessons, we created Sankey diagram (see Figure 2). This diagram breaks down the occurrence of data practices evident in each lesson and the CT practices used with those data practices. For example, we can see that the Biology lesson on Animal Behavior relied primarily on collecting, creating, and analyzing data. In the process of creating data for this lesson, teachers mostly relied on decomposition and few other CT practices. In contrast, when collecting data for this lesson, the teachers actually engaged students in all five CT practices, emphasizing decomposition the least. This way of breaking down the science lessons is helpful in better targeting which data practices are enhanced with specific CT practices in each area of science, and which areas could use more support.

Figure 2. *Sankey Diagram of CT Integration in Science Lessons*

Although some form of automation was evident across all data practices, it is of note that this was the CT practice that science teachers relied on the least. Interviews with teachers after they had implemented their lessons supported this observation. Fully 89% (17/19) of interviewed teachers mentioned struggling with how to integrate automation more fully into their lessons. One teacher summarized the problem as "sticky" in that he believed automation was important and that it would "save us time," but that the time it would take to get his students proficient with tools to learn to automate solutions was currently too much and would take away from the lesson more than add to it. On one hand this highlights the need to teach students how to use the tools and language of automation outside of science courses (much like keyboarding, which is a skill used in many classes after they have learned it). Another teacher demonstrated that it is not just student knowledge of automation, but teacher knowledge, that is also lacking. She stated, "I have no idea of how to make a graph in Excel." This shortcoming in teacher knowledge of automation limits her ability to visualize data in efficient ways that might enable her and her students to better understand their data. Another teacher admitted that automation "makes sense" in her data analysis and that she understood it, but "applying it in our biology classes has been challenging." While all CT practices are important, automation may be the one that most science teachers struggle with.

While automation was a common challenge for all teachers, different data practices and CT use patterns emerged across the different content areas. For example, in Earth Science, there were low occurrences of creating and collecting because of the frequent use of secondary data (e.g., tides, sunspots, volcano activity). In contrast, there were high occurrences of analysis and abstraction due to the visual nature of Earth Science data. In biology, lessons showed an even distribution across all five data practices and all five CT practices. In chemistry, there were few occurrences of either data practices or CT practices in the lesson plans. Just as it may be difficult to create and collect data in earth science due to the needing to collect data across time and geography, in chemistry, the scale is microscopic; meaning that students may need to rely on data already created and collected to run their experiments. However, there may be more opportunity for using and creating algorithms, thinking abstractly and automating analyses than was evident in these lessons. Finally, similar to biology, physics integrated the most CT and data practices through the use of a Pendulums lab, which is also an experimental design lesson. We may conclude that lessons that present students the opportunity to create data also presents opportunities to apply the full range of data and CT practices.

Discussion

Teachers who had high scores on the assessments tended to put more data practices and computational thinking practices in their lesson plans. Analyzing data was used most compared to other data practices. Automation was used least compared to other CT practices. Although most content areas did not use all five data practices and all five computational thinking practices, across the lesson plans, all practices were represented. Teachers also continued to report positive efficacy beliefs in CT and data practices.

Contribution to Learning and Teaching Science

Science educators need to consider learning progressions of "teachers as learners" to optimize the offerings of a PD. As PD experiences are geared more towards science practices, teacher educators need mechanisms that will reveal the beliefs, attitudes, and cognitive processes potentially linked to improving one's lessons. Given the current science practice of CT is underdeveloped, greater knowledge about how teachers learn about CT, implement CT in their lessons, and use CT to support students in data practices are key areas to address in science education research.

Contribution to the Interests of NARST Members

NARST members who research teacher learning may be interested in the patterns we find about how teachers learn CT and how they use CT to support high school student data analysis in science. Additionally, NARST members who conduct PD experiences may be interested in the technique of connecting learning indicators with lesson plan design, implementation, and student outcomes.

References

Augustine N.R. et al. (2005). *Rising above the gathering storm: Energizing and employing America for a brighter economic future.* National Academies Press, Washington, DC.

Bailey, D. H. & Borwein, J. M. (2011). High-precision numerical integration: Progress and challenges. *Journal of Symbolic Computation, 46*, 741-754. Doi: 10.1016/j.jsc.2010.08.010

Foster, I. (2006). 2020 computing: a two-way street to science's future. *Nature, 440*(7083):419.

Martin et al. (1992). Displacement, velocity, and frames of reference: Phenomenographic studies of students' understanding and some implications for teaching and assessment. *American Journal of Physics*, *60*, 262-269.

Maxwell, J. A. (2013). *Qualitative research design: An interactive approach.* Thousand Oaks, CA: Sage.

Peters-Burton, E. E., Rich, P., Cleary, T., Burton, S., Kitsantas, A., Egan, G.*, & Ellsworth, J.* (2020). Using computational thinking for data practices in high school science. *The Science Teacher, 87(*6), 30-36.

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. *Educational Research Review*. doi:10.1016/j.edurev.2017.09.003

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*, 127-147. doi: 10.1007/s10956-015-9581-5

Wing, J. M. (2006). Computational thinking. *Communications of the ACM*, *49*(3), 33-35. doi: 10.1145/1118178.1118215