Teachers' Intersection of Computational Thinking and Data Practices to Support Student Data Analysis during Science Investigations

Problem

The discipline of science is increasingly becoming computational (Bailey & Borwein, 2011; Foster, 2006). Bringing computational tools into the classroom gives a more realistic view of these disciplines (Augustine, 2005). 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 large complex problems (Wing, 2006). CT can be a useful supplement to instruction as it engages students in a cognitive process that students can use to become skillful in data analysis during scientific investigations (Weintrop et al., 2016). However, integrating CT instruction into established curriculum can be overwhelming for both teachers and students, unless adequate supports are provided to guide instructional efforts. 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 use CT practices to support student engagement in data analysis following a professional development program (PD) and to determine viable intersections between data practices and CT that have utility for enhancing student learning in the context of science investigations.

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. 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* is the 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. For teacher PD, we developed a crosswalk that articulates the opportunities to elaborate on data practices through computational thinking practices (Authors, *in press*), which will be shared if the proposal is accepted, but is too extensive to add to this 5-page proposal.

In this study, we gathered data about teachers' beliefs, learning processes, and lesson plans related to data practices and CT; we then observed the implementation of the lessons and collected student work samples. The study was driven by three research questions: (a) In what ways do teachers integrate CT and data practices in their lesson plans during science investigations? (b) What supports do teachers rely on when students are analyzing data? (c) Are there particular intersections of data practices and CT that have more utility for learning during science investigations?

Design

Multiple case study design (Yin, 2003) was used to examine teacher background on CT knowledge, beliefs, and experiences, the design of CT supports in lesson plans, and the outcomes of CT supports implementation in the lessons. Each lesson plan team was considered a case boundary, since implementation of the knowledge was operationalized as a lesson plan. Next, a Type-1 cross case analysis was conducted (Stake, 2006) to determine the common features and unique contexts of the CT supports in the lesson design and implementation. Twenty high school teachers of biology (n=10), chemistry (n=5), Earth science (n=2) and physics (n=4) located in the same school district in the mid-Atlantic region 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.

Measures. *Demographic information.* We gathered demographic factors such as age, gender, ethnicity, primary language, educational background, and current teaching status.

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 and after the PD.

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. A Likert scale ranging from 0 (certain cannot do at all) to 100 (highly certain can do) was used and administered before and after the PD.

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.

Observations, collection of lesson plans, and student work products. In order to determine the influence of the PD on teachers' lessons plans, artifacts of lesson planning from the previous academic year were collected. These lesson plans were used as a baseline to compare to the lesson plans produced following the PD in terms of use of data practices and CT supports. Fourteen lesson were collected after the PD and assessed for use of data practices and CT supports. When the school year begins, the research team will observe lessons and write field notes, collect student work products, and document lesson plan team revisions.

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). Two coding processes were conducted, a priori and emergent. A priori codes were the concepts taught during the CT PD such as decomposition, pattern recognition, abstraction, algorithmic thinking, and automation.

Analysis and Findings

Due to proposal space limitations, only partial results are presented. We present three case studies with preliminary cross case analysis for the purposes of this proposal. We will continue to observe lesson implementation and collect student work samples throughout the 2019-2020 academic year.

Lianne. Lianne was a high school physics teacher with a PhD in physics and 13 years of teaching experience. Before the PD, she reported that overall she was very familiar with data practices, but did not consider data practices to be important to teach and did not typically include them in her lesson plans. She did not identify the application of any data practices on the pre-test vignette. After the PD, she continued to report that she was familiar with data practices, but now found them to be very important and identified visualizing and analyzing data on the vignette. Her mid-point for self-efficacy for teaching data practices remained the same from before and after the PD at 76 out of a 100. It appears that the PD had a small positive impact on Lianne's beliefs about data practices and her knowledge about data practices.

Regarding her beliefs and knowledge about CT, Lianne reported pre-PD that she was not familiar with CT and did not consider it to be important. She did not identify any concepts about CT nor identify any CT practices on the scenario before the PD. After the PD, she had strong knowledge about decomposition and pattern recognition. Her knowledge about abstraction and automation was still developing by the end of the PD. She was able to identify the application of decomposition and automation on the scenario after the PD. However, her self-efficacy about CT decreased after the PD by eight points to a score 68. Again, the PD had a positive impact on her beliefs and knowledge about CT, particularly about decomposition.

Her lesson plans focused on Kepler's Laws and location of a satellite during an orbit. In the lesson, students began by modeling an orbit of an "earth" around the sun, and used formulas in a spreadsheet to find another planet's locations during orbit. In this investigation, she focused heavily on *data collection* and used *decomposition* to support student learning. She also used the practice of *data manipulation* and encouraged students to use *pattern recognition* to determine if the values they obtain made sense with the patterns that were found in the planet that she modeled for the class. The data and CT practices that she focused on in the lesson mirror the topics in which she had the most improvement in the PD. However, there were few data practices emphasized in the lesson as compared with lessons from other science disciplines.

Amelia. Amelia was a high school biology teacher who possesses an undergraduate degree in biology and a Master's degree in education. She had three years of teaching experience. Before the PD, she reported that she was very familiar with data practices, taught them somewhat and did not feel they were important. Amelia had some developing knowledge about visualization before the PD, but did not demonstrate knowledge of other data practices.

After the PD, she again reported she was very familiar with data practices, but now found them very important. Her knowledge improved on creating, visualizing, and collecting data and she reported her self-efficacy to increase 20 points over the PD to be 95 out of 100 at the end of the PD.

Before the PD, Amelia reported that she was very familiar with CT practices, taught them somewhat often, but did not find CT important except for abstraction and automation. When Amelia began the PD, she had strong knowledge of decomposition, and some developing knowledge on automation, but did not demonstrate other types of knowledge about CT. Additionally, she was not able to identify any CT practices on the scenario. After the PD, Amelia reported that she understood more formal knowledge about CT. She was able to identify the CT practice of decomposition on the scenario, but no other practices. Her self-efficacy of CT increased by 22 points over the PD to be 93 by the end of the PD.

Amelia's lesson plan revolved around the use of opioids so that students could gain practice with data practices. Students were to gather and visualize data from different states in the U.S. in order to make conclusions about death rates due to opioid use and compare regions. In this lesson, she focused on the data practices and corresponding CT supports found in Table 1 below. *Decomposition* and *pattern recognition* were the most frequently used CT practices to support the data practices in the lesson, and *decomposition* was clearly the concept in which Amelia had the most knowledge. Additionally, Amelia's initial value for the CT practices of *abstraction* and *automation* were seen in the practices that she used to support *manipulating* and *visualizing* data in her lesson.

Table 1. Amelia's Opioid Lesson Plan Data Practices and CT supports

Data Practice	**Supporting CT Practice**
Creating data	Decomposition and Algorithmic thinking
Collecting data	Decomposition and Pattern Recognition
Manipulating data	Decomposition, Pattern Recognition, Algorithmic thinking and Abstraction
Visualizing data	Pattern recognition, Abstraction, and Automation

Elise. Elise was an Earth science teacher with a B.S. degree in Earth science and a Master's degree in education and 16 years' experience teaching high school. Before the PD, Elise reported that she was very familiar with data practices and felt that creating and collecting data were somewhat important, but manipulating, visualizing and analyzing data were very important. Surprisingly, after the PD, she reported that overall data practices were not important. Elise did not identify any data practices on the given scenario before or after the PD, but her self-efficacy about data practices increased by 35 points to end the PD at 88 points out of 100.

She reported that she had no prior knowledge of CT before the PD and she did not identify any CT practices in the scenario. However, after the PD, she demonstrated strong knowledge on decomposition, pattern recognition, and automation and was able to identify the application of decomposition and automation in the scenario. She explained, "I was not exposed to computational thinking before this. Most components I already knew and I just had to learn a new term for the practice." Her self-efficacy of CT increased by 37 points to end the PD at 88 points out of 100.

Elise's lesson plan focused on the topic of sunspots. Students were to use data to determine if sunspots appeared in a regular pattern and then were to ask their own question about the relationship of sunspots to phenomenon on the Earth and use data to answer their question. In this series of lessons, Elise focused on the data practices and corresponding CT supports found in Table 2 below. Elise's lesson did not use the data practice of manipulating data, but did use the other data practices in her lesson. *Decomposition* was the most used of the CT practices to support data practices, which aligned with the learning she demonstrated in the PD. Elise had also demonstrated growth on the CT practices of *pattern recognition* and *automation*, which were also used to support the practices of visualizing and analyzing data.

Table 2. Elise's Sunspots Lesson Plan Data Practices and CT supports

Data Practice	Supporting CT Practice
Creating data	Decomposition
Collecting data	Abstraction
Visualizing data	Decomposition and Automation
Analyzing data	Decomposition, Pattern Finding, Abstraction and Algorithmic Thinking

Across the teachers, the CT practice of *decomposition* had the most improvement in knowledge and application. In teachers' planned lessons, *decomposition* was the most frequently used CT practice and it was typically paired with *creating data*. Similarly, *pattern recognition* and *abstraction* were categories where teachers learned more formal knowledge in the PD, and these CT supports appeared in the lesson design. *Pattern recognition* was most paired with *collecting data*, and *abstraction* was most paired with *analyzing data*. In this study, we will continue to compose case studies of all 20 teachers and link teacher learning to lesson design and ultimately to student learning about data practices through computational thinking.

Contribution to the Teaching and Learning of 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 it to support data analysis. Additionally, NARST members who conduct PD experiences may be interested in the technique of connecting learning indicators with lesson plan design, implementation, and student outcome.

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.

Authors (in press). *The Science Teacher.*

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.

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

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