International Society of the Learning Sciences

Surfacing Collective Pedagogical Content Knowledge for Data Literacy: An Exploratory Study with High School Science Teachers

Katherine M. Miller, Susan Yoon, and Jooeun Shim
kmmiller@gse.upenn.edu, yoonsa@upenn.edu, jshim@gse.upenn.edu
University of Pennsylvania

Abstract: There is a need for greater focus on the critical aspects of data literacy at the high school level. However, data literacy requires explicit integration between math and science, which high school teachers often aren't experienced with. Thus, teaching data literacy may require different PCK than that typical for high school science. This study adds to underrepresented research on PCK for data literacy by engaging a group of experienced teachers in collaboratively attempting to define PCK for data literacy through discussing their experiences implementing a unit designed to support the development of data literacy in students. Three components of PCK for data literacy are surfaced and examined.

Introduction

In our data-rich world, there are strong calls for greater focus on data literacy as increasingly, 21st century jobs require some level of data literacy (e.g., Gould et al., 2016). Rubin (2020) identifies five critical aspects of knowledge for working with data: a) *data in context*, b) *variability of data*, c) *data in aggregate*, d) *data visualization*, and e) *inferences with data*. For the purpose of this study, a sixth critical aspect highlighted by Kjelvik and Schultheis (2019) has been added: f) *complexity of data sets*. These six critical aspects of data literacy serve as a foundation for what students should know about how to work with authentic data. There has been limited research on pedagogical best practices for teaching data literacy, especially through authentic, complex, messy data (Gould et al., 2016; Mike, 2020). However, Nadelson and Seifert (2017) discussed the importance of highlighting the interdisciplinary nature of data, with connections between math and science classes made explicit. This sort of content integration is not something that high school teachers are often experienced or supported in doing (Kelley & Knowles, 2016), especially when it comes to engaging students with complex authentic data (Kjelvik & Schultheis, 2019). Thus, teachers need support in identifying the pedagogical content knowledge (PCK) required to support student learning with and about data (Kelley & Knowles, 2016). Though some research has been done on PCK for STEM integration (e.g., Vossen et al., 2020), PCK for data literacy is an under-explored concept (Mike, 2020). This study builds on a larger study on integrating bioinformatics lessons into high school science classrooms using student-collected data. We work with a small group of experienced teachers to collaboratively examine their understanding of PCK for data literacy using established methods of surfacing PCK (i.e., content representations and video reflections). This exploratory study seeks to answer the question: what are some components of PCK for data literacy in high school science classrooms?

Background

High school science teachers often have not been given the time or resources to engage students in full authentic investigations which address the interdisciplinary nature of data and as such they may need support in determining the best pedagogical strategies (Gould et al., 2016; Nadelson & Seifert, 2017). Currently very little is known about what PCK for teaching with and about data looks like and what sort of experiences are needed to support it (Gould et al., 2016; Mike, 2020). PCK is defined as the "special amalgam of content and pedagogy that is uniquely the province of teachers" (Shulman, 1987, p. 8) and has been shown to be subject and even topic specific (Smith & Banilower, 2015). As such, PCK for data literacy may require knowledge that is distinct from that which science teachers already hold. The science education community has outlined a model for PCK that identifies three levels of PCK that interact with each other: collective, personal, and enacted, where collective PCK encompasses the common knowledge held by a group of teachers made up of an amalgamation of their personal knowledge and experiences with enaction (Carleson & Daehler, 2019). Additionally, PCK has been shown to be composed of distinct components (e.g., Park & Chen, 2012), the two most commonly studied of which are *knowledge of student understanding*, and *knowledge of teaching strategies*.

Studies have shown that working collaboratively allows teachers to surface deeper and more nuanced PCK (Nilsson & Loughran, 2012; Vossen et al., 2020). Research that has supported teachers in successful identification and growth of PCK often uses one or both of two methods: content representations (CoRes) and video reflections. A CoRe is a template which prompts a teacher or group of teachers to think about teaching a particular topic through big ideas with questions such as: *Why is it important for students to know this? What else*

do you know about this idea that you do not intend students to know yet? What is your knowledge of students' thinking which influences your teaching of this idea? Studies have found that enabling teachers to make their practice, and their thinking on their practice visible allowed them to then reflect effectively on that knowledge (e.g., Vossen et al., 2020). Using a CoRe as an artifact allows teachers to develop a shared language around the content knowledge and the PCK for teaching a particular topic (Nilsson & Loughran, 2012). Engaging in video reflections prompts teachers to examine and make visible their thinking while enacting a lesson. By combining CoRes and video reflections on data literacy lessons, this study hopes to support a small group of experienced science teachers in surfacing PCK for data literacy.

Methods

Context

This exploratory study exists within the larger context of an intervention designed to support teachers in implementing a STEM-integrated problem based learning unit on bioinformatics in their science classrooms. The unit uses mobile learning through portable air quality sensors connected to a phone app to engage students with active collection of real-world data to address the provided problem of asthma rates in urban areas. The intervention included a summer professional development which took place over four weeks during July 2021, conducted asynchronously on the edX testing platform called Edge. After the summer course, four participating teachers were invited to participate in a PCK Extension Workshop Program which would focus explicitly on PCK. The extension program consists of approximately 20 hours of workshops spread over fall and winter 2021 while teachers are actively engaged in implementing the bioinformatics unit. The series of workshops includes virtual synchronous meet ups with some asynchronous work to support preparation for these meetings. During the workshops, teachers collectively build CoRes on the different critical aspects of data literacy while learning about PCK as a concept and how it can be identified and measured. Additionally, teachers conduct video reflections of their implementation in which they identify examples of enacted PCK and discuss how they support student learning. As the workshop program is still in progress, this paper focuses on initial findings on how participating teachers are conceptualizing PCK for data literacy.

The four participating teachers in this study work in three different public schools in a large urban school district in the northeastern United States in which over 90% of students qualify for free or reduced lunch. The three schools represent typical schools within the district, with large populations of minority students, English as a second language learners, and students who are socioeconomically disadvantaged. The four teachers consist of three who identify as female and one who identifies as male, three who identify as white and one who identifies as South Asian. They have an average teaching experience of 12.75 years with the most experienced teaching for 22 years while the least experienced is only in her second year of teaching in a formal classroom after a number of years working in informal education. All four teachers are implementing the bioinformatics unit in a high school science classroom.

Data sources and analysis

The data sources for this study consist of three CoRes built collaboratively during the extension workshops focused on the critical aspects of *data in context*, *data visualizations*, and *inferences with data*, written reflections completed by teachers after watching two videos of themselves and their colleagues teaching data literacy lessons that focused on those three critical aspects, and transcripts of the sessions during which the CoRes were created and the video reflections discussed. Five sessions, each 90 minutes in length (or 450 minutes total) were used for this analysis. A constant comparative analysis (Glaser, 2008) was conducted on the transcripts of the sessions and the written artifacts to identify emerging themes on PCK for data literacy.

Findings

Teachers have begun to surface an initial set of components of PCK for data literacy. These include two aspects of *knowledge of students' understanding* of data as well as *knowledge of a teaching strategy* that can be applied to help engage those student understandings.

Knowledge of students' understanding of data

Two initial themes that have arisen from teachers discussions of PCK for data literacy are that students: a) are not used to questioning data and b) struggle with the ambiguity of real-world authentic data. Concerning the first theme, while high school science does often use data, students are usually asked to interpret it without being asked to interrogate it, however a primary component of data literacy is being able to place data in its context and

understand how the context might affect the results and interpretations. In order for teachers to effectively teach the context component of data literacy, it is important for them to have strong knowledge of how and why students' predisposition to data may hinder that learning. To surface and build on that component of teacher PCK, the teachers were asked to explicitly identify what they knew and understood about student thinking. In response to a prompt in the CoRe on *inference with data* about student thinking, the teachers wrote, "Students will see "expert" conclusions and assume the expert must know - so why would they need a different conclusion." In the discussion around this understanding Mary stated that, "the whole point of [students] learning data literacy is so that they are not stuck believing that three or four out of four dentists means something really magical." This is revealing knowledge about student understanding, that students are "stuck" seeing data as truth without training to question it or even the knowledge that they can. Hallie added that though students are "able to see the 'what' they don't go for the 'so what'" which is referencing students' tendency to read data, but not question it. The knowledge that students are not used to questioning data also came up while teachers were building a CoRe on *data visualization* and discussing how easily people can be manipulated by "pretty" graphs. Mary said, "[students] are mostly on the receiving end of data and that is a place they can get manipulated by beautiful fonts and fancy bright red." In using the word 'manipulated' to talk about how students interface with data visualizations, she is again revealing knowledge of how students are unused to questioning data and instead accept it as is. By revealing this knowledge, the teachers are making visible a component of PCK which will help them better address students' preconceptions.

With respect to the second theme of struggle with ambiguity, while students are usually conditioned to accept data they see externally as fact and are not experienced in questioning it, they often take an opposite approach with their own data in that rather than accepting data they collect, they assume there is a "right" answer and their data needs to match that. Again, this is important knowledge for teachers to have in order to engage their students with data literacy, specifically with the component of building inferences with data and creating data visualizations to display their findings. When completing a CoRe on *data visualization* teachers wrote that creating data visualizations can "feel wishy-washy for students who are accustomed to 'only one of the choices is correct'" and that when creating data visualizations "students can be overly cautious with taking intellectual/design risks, to the point of paralysis. Many won't start until they know they are doing it 'right'" In further discussions about this understanding that students have about working with data, Manisha brought up the phrase "integrity of data" and how when it comes to their own data students are uncomfortable with ambiguity being okay. Additionally, while building the CoRe on *inference with data* teachers discussed ways to engage this student understanding through having them collect data on tangible objects so that they can see that "data is real and personal" to combat students' need for data to be "right" as measured by an arbitrary authority. Surfacing the knowledge that students struggle with ambiguity of data allowed teachers to verbalize this component of PCK and apply it to strategies for teaching data literacy.

Knowledge of instructional strategies for teaching with and about data

While teachers were building the CoRes and discussing their implementations of the data literacy lessons, they began to identify one particular instructional strategy to work with students' understanding of data as needing to be "right" and unquestionable. The idea of data as a story, or as having a story to tell, surfaced during many of the teachers' conversations. It was discussed as a way to support students in shifting the way they interface with data by thinking about a data set as a text and using text-analysis tools to deconstruct the 'story' the data is trying to tell. One of the big ideas they collectively chose while building the CoRe for *data visualization* was "Data visualizations tell a particular story about a set of data at a particular point in time." Additionally, while discussing the CoRe for *data in context,* Will said, "[students] are looking at one set of data, but there's this whole other world of data that you're choosing to ignore to fit your story and that's what I want them to understand, is you can use data to properly to support any position that you want."

When teachers were discussing a video of Hallie teaching a data literacy lesson, the instructional strategy of unpacking data as if it's a text and using literary analysis tools to support students in gaining experience and confidence critiquing data began to take shape. Hallie explained her emergent thinking on this saying, "If we were reading a short story and we say [...] what's the background we can assume this character might have had because of the way that they speak or whatever." She then connected that to analyzing a data visualization with, "There is a story that it's telling. [...] so, using a kind of text analysis [the students] can ask questions like who were these people? Why were they asking [these questions]?" This led to a larger conversation about different ways that students could be introduced to this process with Mary suggesting, "Offer a data set and invite groups of students to use the same data to tell various stories - don't tell each group what the others were asked to do - and then have students present their visualizations to the larger group." Manisha then added a suggestion to use the sugar refinery association's involvement with the diabetes association as a way to show students competing stories and "not to tell them the other side [at the beginning] and then just drop one bomb of a perspective and make them rethink

everything that they believed before that." Both of these examples are methods of introducing students to thinking about data as telling a story. While Mary's suggestion engages with students' understanding about data being unambiguous, Manisha engages with their tendency to accept data at face value without interrogating or questioning it. The strategy of treating data like a literary story can be used to engage both student understandings addressed in the previous section.

Implications

This paper reports on the initial findings of an ongoing study that investigates the surfacing of collective PCK for data literacy in high school science. As the research wraps up, a more robust analysis of the data will be conducted in order to achieve validity on these initial findings. Additionally, extensive interviews with the participating teachers will be conducted at the end of the workshop series to explore the processes used in the workshop and their effectiveness in supporting teachers to surface and expand their PCK for data literacy. However, the initial findings show that experienced high school science teachers hold some PCK for teaching with and about data and that, in support of previous research on developing PCK (e.g., Nilsson & Loughran, 2012, Vossen et al., 2020), PCK can be surfaced through collective discussion of the critical aspects of data literacy as mediated by CoRes and video reflections. This research hopes to continue to add to the burgeoning literature on teaching for data literacy at the high school level (e.g., Mike, 2020) by working with experienced teachers to understand what PCK for data literacy looks like, how it may or may not be different from traditional PCK for science, and how teachers can be supported to make this PCK visible and accessible. This study will contribute to identifying the supports needed to improve students' knowledge of the critical aspects of data literacy through identifying PCK for teaching with and about data.

References

Carlson, J., & Daehler, K. R. (2019). The refined consensus model of pedagogical content knowledge in science education. In A. Hume, R. Cooper, & A. Borowski (Eds.), *Repositioning pedagogical content knowledge in teachers' knowledge for teaching science* (pp. 77–92). Springer, Singapore.

Glaser, B. (2008). The constant comparative method of qualitative analysis. *Grounded Theory Review, 3*(7).

Gould, R., Machado, S., Ong, C., Johnson, T., Molyneux, J., Nolen, S., Tangmunarunkit, H., Trusela, L., & Zanontian, L. (2016). Teaching data science to secondary students: The mobilize introduction to data science curriculum. IASE 2016 Roundtable Paper.

Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. *International Journal of STEM Education*, *3*(1).

Kjelvik, M. K., & Schultheis, E. H. (2019). Getting messy with authentic data: Exploring the potential of using data from scientific research to support student data literacy. *CBE Life Sciences Education*, *18*(2), 1–8.

Mike, K. (2020). Data science education: Curriculum and pedagogy. Proceedings of the 2020 *ACM Conference on International Computing Education Research*. Virtual event, New Zealand.

Nadelson, L. S., & Seifert, A. L. (2017). Integrated STEM defined: Contexts, challenges, and the future. *Journal of Educational Research*, *110*(3), 221–223.

Nilsson, P., & Loughran, J. (2012). Exploring the development of pre-service science elementary teachers' pedagogical content knowledge. *Journal of Science Teacher Education*, *23*(7), 699–721.

Park, S., & Chen, Y. C. (2012). Mapping out the integration of the components of pedagogical content knowledge (PCK): examples from high school biology classrooms. *Journal of Research in Science Teaching, 49*(7), 922–941.

Rubin, A. (2020). Learning to reason with data: How did we get here and what do we know? *Journal of the Learning Sciences, 29*(1), 154–164.

Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review, 57*, 1–22.

Smith, P. S,. & Banilower, E. R. (2015). Assessing PCK: A new application of the uncertainty principle. In A. Berry, P. Friedrichsen, & J. Loughran (Eds.), *Re-examining pedagogical content knowledge in science education* (pp. 28-42). Routledge.

Vossen, T. E., Henze, I., De Vries, M. J., & Van Driel, J. H. (2020). Finding the connection between research and design: the knowledge development of STEM teachers in a professional learning community. *International Journal of Technology and Design Education*, *30*(2), 295–320.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. #1812738