International Society of the Learning Sciences

Integrating Data Literacy into Secondary School Science: An Exploratory Study of a Pilot Professional Development

Katherine Miller, Susan Yoon, Jooeun Shim, Amanda Cottone
kmmiller@gse.upenn.edu, yoonsa@upenn.edu, jshim@gse.upenn.edu, amandaco@sas.upenn.edu
University of Pennsylvania

Abstract: In our data-rich world, data literacy is becoming an increasingly important goal in education as data sets become bigger, and more complex. However, data literacy falls outside the traditional topic domain of secondary science curricula and so may be beyond the scope of the pedagogical content knowledge (PCK) of teachers who have received training in only one subject area. This paper presents an exploratory study of how secondary school science teachers learn and use the content and skills of integrating data literacy into their curriculum, within the context of a PD program for bioinformatics. The findings suggest that there are a separate set of PCK components that teachers need to master in order to successfully teach data literacy, and that successful PD must support teachers in developing them.

Keywords: data literacy, pedagogical content knowledge, professional development

Introduction

In our data-rich world, there are strong calls for greater focus on data literacy within education (Gebre, 2018; Gould et al., 2016; Rubin, 2020). Though it is interdisciplinary in nature, data literacy suffers from being nebulously defined and having no clear home within the traditional secondary school subject-based curriculum (Gebre, 2018; Lee & Wilkerson, 2018). As integration of subjects has been shown to present a challenge to teachers who have received training in only one subject area (Aslam et al., 2018), teachers may need additional support to develop knowledge and skills for teaching data literacy, also known as pedagogical content knowledge (PCK) (Neumann et al., 2019). Data literacy as a component of teacher education is a relatively new concept. As the way we interact with data through increased technology exposure, and greater computing power changes, the way we seek to teach students to interact with data also needs to change to promote a broader relationship with data (Lee & Wilkerson, 2018; Wise, 2020) and a greater focus on the context in which data is collected and used (Rubin, 2020). In this study, we sought to better understand how PD can support teachers in developing PCK for data literacy. Through analysis of interviews, surveys, and classroom observations we asked, how does a PD program for bioinformatics support teachers in learning and implementing the content and skills of integrating data literacy into their science classroom and existing curriculum?

Background

Researchers and practitioners are thinking about what it means to teach data literacy and what the pedagogical implications are for integrating data into the classroom (Gould et al., 2016). Wilkerson and Polman (2020) outline two core pedagogical commitments for teaching with data. (1) that teachers should be flexible with learning and promoting new tools and methods for working with data, and (2) that data based teaching should be grounded in full cycles of inquiry using real, consequential data. Lee and Wilkerson (2018) also support the notion that work with data should be conducted within the context of meaningful scientific investigations, and add that the interdisciplinary nature of data should be highlighted with connections between math and science classes, and that the complex and dynamic nature of data should be explicitly discussed and multivariate relationships explored. Since training for secondary school teachers is often specialized by subject, these pedagogical skills may be new to many teachers (Aslam et al., 2018) and require new PCK. A review by Neumann and colleagues (2019) offered a consensus definition of PCK as "teachers' personal knowledge that drives their planning for, implementation of, and reflection on instruction" (p. 856). In merging knowledge of pedagogy and knowledge of content, PCK is topic specific, which can hinder a teacher in integrating new content material that may fall outside their learned topic domain (Aslam et al., 2018). Data literacy currently exists in that space outside of traditional secondary school topic domains (Gebre, 2018) and as such, most science teachers are not supported in developing PCK for data literacy. Additionally, data literacy as a concept is still being defined which further complicates efforts to support teachers in integrating it into their classrooms

Data literacy is an emerging concept without a clear definition or set of core competencies that are widely accepted across different fields (Wolff et al., 2016). A recent review of literature on data literacy attempted to

synthesize a working definition of the concept as "the ability to ask and answer real-world questions from large and small data sets through an inquiry process, with consideration of ethical use of data" (Wolff et al., 2016, p. 23). However, this definition leaves space for a number of different competencies, especially as data sets become more complex, and as students become participants in data creation (Gebre, 2018; Wise, 2020). In their report on data use in secondary school, Lee and Wilkerson (2018) begin to create a framework for some of these competencies which could help in defining PCK for data use. They outline four understandings that are required to develop data literacy: a) *measurement and sampling*, b) the *characteristics of data*, c) different types of *data representation*, and d) *making inferences from data*. Rubin (2020) supports these themes, and adds a fifth understanding, which is e) *context*, the who, when, where, what, why and how of data collection and use.

Defining competencies that lead to development of data literacy is a helpful step leading to integrating data literacy into science classrooms, but teachers also need to be supported in learning these competencies and developing PCK for data. Though research has shown that PD can enact change in teachers' practice (e.g., Darling-Hammond et al., 2017), it is important to design PD in a way that is mindful of the unique challenges posed by promoting data use in science classrooms. Lee & Wilkerson (2018) lay out a number of recommendations for improving teachers' knowledge and practice when it comes to working with data in the classroom. These include: providing extended time for teachers to use the technology through a full cycle of inquiry; highlighting that though technology and data-based inquiry can be used to teach content, additional data literacy skills also need to be taught in order for content learning to be achieved; and provide case-study examples of students working with data and technology. Where we know that integrating data literacy in science education content has presented challenges, teachers will need support for developing PCK in PD opportunities. For this reason, considering high quality PD characteristics (Darling-Hammond et al., 2017) is important.

Methods

This is an early stage, exploratory study that draws on qualitative data from a small pilot group of teachers. The goal of the full project is to support teachers in implementing a problem-based learning (PBL) curricular unit that challenges students to come up with data-based solutions for addressing asthma rates in their urban district. The curriculum requires reasoning with less traditional data sources such as that collected by digital sensors, and that available in large public data sets (Lee & Wilkerson, 2018).

Designing for data literacy

The PD was designed and developed in the winter and spring of 2019 and implemented over three weeks in July 2019. Teachers attended in-person for 6 hours of learning time each day, for a total of 90 hours. Of those, about 7 hours were devoted to data literacy. Data literacy instruction utilized five primary activities: a) defining data literacy, b) learning new tools for working with data, c) working with data as a learner, d) reflecting on learning about data, and e) reflecting on teaching with data. These activities were based on design principles for high quality PD and emerging theories on developing data literacy as outlined in the above section. In defining data literacy, participants explicitly highlighted the need for data literacy skills while grounding learning in the context of their classroom. The PD engaged participants as active learners while learning a new app for data analysis. Participants worked in pairs to conduct a full cycle of data collection, analysis, and communication of results, using student-facing materials from the example PBL unit. The PD provided time for reflection both on the process and experience of learning with data, and on how to apply what they learned to strategies for teaching with data in their classroom. All data literacy instruction was grounded in a community level scientific problem (high asthma rates in urban areas) and structured around a full inquiry-cycle to address that problem. Throughout the PD, the facilitators modeled effective instructional practices and encouraged collaboration.

Participants

The pilot study consisted of six teachers from a large urban school district in the northeastern US where over 85% of students are racial minorities and 100% of students are considered economically disadvantaged. Three of the teachers taught biology and three taught environmental science. Three of the participants identified as female and the other three as male. Half of the participants identified as Black and the other half as White. They had an average of 9.8 years of experience ranging from 2 to 18 years.

Data sources and analysis

The sources for this study consisted of surveys conducted pre and post PD, external evaluator memos from focus groups during the PD, classroom observations during implementation, and interviews conducted post implementation. A content knowledge survey which included a set of questions about teachers' understanding of

data literacy and their pedagogical practices was given at the beginning and end of the PD. On the last day of the PD, participants were also asked to complete an Evaluation Survey of the PD experience with a few questions specifically aimed at the data literacy content. The external evaluator on the project conducted focus group interviews with the participating teachers three times during the PD (Day 1, 5 and 10) and wrote up notes summarizing themes with supporting quotes that were shared with the research team. Classroom observations were conducted throughout the school year by different members of the research team. At the end of the school year, semi-structured post-implementation interviews were conducted with five of the six teachers which included questions that asked teachers to reflect on their teaching and their students' learning.

The data was qualitatively mined for articulations of data literacy themes in order to study the pilot run of the PD and teachers' impressions of it and success with implementation. There were some prompts in the surveys and interviews that specifically asked about impressions of the data literacy content, and these were used to frame pre and post self-reported beliefs from the six teachers. These beliefs and experiences were then compared to observation notes from classroom implementation to determine the fidelity with which the data literacy components of the curriculum were enacted.

Findings

Three primary themes were drawn from the data analysis: a) teachers came into the PD with hesitations about working with data; b) the PD helped some but not all teachers feel prepared to teach with data; and c) teachers encountered a number of challenges in implementing integrated data literacy content.

Initially, teachers demonstrated hesitancy in working with data. The pre-surveys showed that five of the six teachers held a traditional view of what working with data means in a classroom, highlighting skills such as analyzing data and creating graphs, with only one participant linking skills to larger ideas such as solving problems. This focus on discrete steps for working with data is evidenced in this typical response from the pre-survey: "[Data literacy is] the method individuals use to collect, interpret and analyze data. This can include building data tables, using variables and/or drawing graphs." Though all six teachers indicated that they had worked with data in their classrooms previously, they came in with a range of experiences for how often data was used in their classrooms, and the level of rigor it was taught with. Throughout the first week of the PD, teachers expressed nervousness about the data literacy components of the curriculum, with one teacher on the first day saying, "The most difficult aspect will be helping students analyze data and interpreting graphs. This is something that they have rarely, if ever, done in their life." Additional notes from the focus group interview on day one support this quote, stating that "they [the teachers] mentioned students struggling to understand the X and Y axis of graphs. They rely on math teachers to teach graphing; they feel insecure in doing this themselves." This shows that they were expressing uncertainty about certain aspects of data collection and analysis.

By the end of the PD, in the post-surveys, teachers showed more confidence than they had at the beginning and midway through the workshop. When asked about the efficacy of the data literacy component of the PD, all six teachers responded positively, suggesting that the goals of the PD were met and that teachers felt they had learned something useful for implementing data literacy instruction. When asked what additional support they would need in order to implement the PBL curriculum in their classroom, most teachers said they felt prepared to teach, however some of them mentioned that they were still feeling insecure, such as the teacher that wrote in response to a prompt about foreseen challenges for implementation, "the most difficult challenge will be data analysis, and chart making. I just foresee the students having difficulties... and with my own insecurities about the topic." There were hints that more teachers felt this way than chose to admit it. In response to the question *What would you have liked to spend more time on?*, universally, teachers requested more time with data collection and analysis and "More practice analyzing and creating graphs." So, despite praising the PD for adequately preparing them to teach data literacy, most teachers desired more time to internalize the concepts.

The fidelity of the data literacy components during implementation was low. Though teachers were encouraged to modify the example curriculum to fit the needs of their students and classroom, every teacher cut out at least some of the data literacy components, with some teachers skipping entire lessons and sets of objectives. During the lessons that were observed, many of the teachers struggled with aspects of working with the data. As in the following example where the researcher noted that their "Overall impression of Tamara' data analysis class is that only a surface level of data analysis (getting average) is being taught. Tamara looks a little discouraged since she has been struggling with facilitating these lessons with students." As the teachers struggled to implement the data literacy components of the unit and cut back on the activities and lessons being presented to the students, the bigger picture of the meaning and relevance of data to real-world problems got lost. Teachers fell back on approaches to data literacy that they were comfortable with, which deprioritized focusing on the meaning and relevance of the data sets for exploring real world problems. Together, these findings showed that teachers felt

underprepared to teach data literacy in their science classrooms, and though a three week intensive PD helped them to feel more prepared, that increase in confidence did not translate into effective teaching of data literacy.

Discussion and implications

In this preliminary study, the goal was to highlight themes that would help guide the researchers in modifications to the intervention and further support for teachers in developing PCK for data literacy integration. The analysis of data sources tells a story that aligns with common pitfalls of PD (Darling-Hammond, 2017) and the limited literature on teachers engaging with data literacy concepts in their classrooms (Lee & Wilkerson, 2018). By studying how teachers learned and implemented data literacy integration, this research has highlighted the ways in which the PD failed to adequately support teachers in integrating data literacy into their science classrooms, and in the process uncovered the need to focus more on PCK development. Teachers entered the PD a little wary of the data literacy and technology components. They reported learning a lot during the PD and feeling prepared to teach, while simultaneously wishing for more time to grapple with data literacy concepts. However, once they were in the classroom, the confidence faltered, and they mostly fell back on more traditional and conservative understandings of data literacy that divorced the data engagement from the content and framing problem.

The designers and facilitators of the PD overestimated the extent to which teachers would be prepared to implement the data literacy component of the PBL curriculum. The PD was designed based on assumptions that participating teachers would have knowledge and skills about the complexity of data and tools for data analysis that would translate quickly onto the new technology and data sets. However, the challenges that teachers faced in implementing align with the two core competencies for teaching with data laid out by Wilkerson and Polman (2020) and suggest that there may be a separate set of PCK components that teachers need to master in order to successfully teach data literacy, beyond those commonly considered part of science teachers' PCK. In order to better support teachers in integrating data literacy into their science classrooms, PD needs to focus on not only the content of data literacy and its' connections to science content, but also on the separate PCK that teachers need to implement with complex data and new technologies. The findings from this preliminary study add to the growing literature (e.g., Wilkerson & Polman, 2020; Wise, 2020) that highlight the need to support teachers in rethinking the way they engage with data in their science classrooms.

References

Aslam, F., Adefila, A., & Bagiya, Y. (2018). STEM outreach activities: an approach to teachers' professional development. *Journal of Education for Teaching, 44*(1), 58–70.

Darling-Hammond, L., Hyler, M. E. and Gardner, M. (2017), *Effective Teacher Professional Development*. Learning Policy Institute.

Gebre, E. H. (2018). Young adults' understanding and use of data: Insights for fostering secondary school students' data literacy. *Canadian Journal of Science, Mathematics and Technology Education*, *18*(4), 330–341.

Gould, R., Machado, S., Ong, C., Johnson, T., Molyneux, J., Nolen, S., Tangmunarunkit, H., Trusela, L. & Zanontian, L. (2016, July). *Why Teach Data Science?* Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), Berlin, Germany.

Lee, V. R., & Wilkerson, M. (2018). Data use by middle and secondary students in the digital age: A status report and future prospects. Commissioned Paper for the National Academies of Engineering, and Medicine, Board on Science Education, *Committee on Science Investigations and Engineering*, 1–43.

Neumann, K., Kind, V., & Harms, U. (2019). Probing the amalgam: the relationship between science teachers' content, pedagogical and pedagogical content knowledge. *International Journal of Science Education*, *41*(7), 847-861.

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.

Wilkerson, M. H. & Polman, J. L. (2020). Situating data science: Exploring how relationships to data shape learning. *Journal of the Learning Sciences, 29*(1), 1–10.

Wise, A. F. (2020). Educating the next generation of data scientists for the next generation of data. *Journal of the Learning Sciences, 29*(1), 165–181.

Wolff, A., Gooch, D., Montaner, J. J. C., Rashid, U., & Kortuem, G. (2016). Creating an understanding of data literacy for a data-driven society. *The Journal of Community Informatics, 12*(3), 9–26.

Acknowledgements

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